Contributions to Random Energy Models

Nabin Kumar Jana

Indian Statistical Institute
Kolkata
2007

Contributions to Random Energy Models

Nabin Kumar Jana

Thesis submitted to the Indian Statistical Institute
in partial fulfillment of the requirements
for the award of the degree of
Doctor of Philosophy.
October, 2007

Indian Statistical Institute
203, B.T. Road, Kolkata, India.

To My Parents

Acknowledgements

A simple thanks will not be enough to convey my gratitude towards my supervisor, Professor B. V. Rao for his careful guidance, constant encouragements and the time he has given me during this project.

I am grateful to Professor Rahul Roy for introducing several models in statistical physics and providing scope to visit ISI, Delhi campus.

I would like to take this opportunity to thank all my teachers in Stat-Math Unit of Kolkata, Delhi and Bangalore center for their illuminating courses and fruitful discussions.

I am thankful to Dr. T. Mukherjee, Principal of Bijoy Krishna Girls College and my departmental colleagues of this college for their constant co-operations and encouragements to complete this work.

I thank all my friends, seniors, juniors as well as the members of Stat-Math Unit for providing such a wonderful atmosphere to work on.

My thanks also goes to my friends Sourabh, Sohini, Soumenda, Dola and dada Ac' Devatmananda Avt. for their constant encouragements. Finally, I want to thank my wife who shouldered a lot of responsibilities so that I could concentrate only to my research work.

Contents

Chapter 0

Introduction

In this introductory chapter, we begin with a brief description of spin glasses in section 1. We are not physicists. The purpose of this section is to trace the history of the models. Section 2 gives a brief summary of the thesis and section 3 recalls certain known facts which will be used later in the thesis.

0.1 Origin of the problem

The models considered in this thesis have their origin in spin glass theory. Roughly, *spin glass* is a glassy state in a spin system or a disordered material exhibiting high magnetic frustration. The origin of this behavior can be either a disordered structure (such as that of a conventional, chemical glass) or a disordered magnetic doping in an otherwise regular structure. But what is a glass? Loosely speaking, it is a state of spins with local ordering (in solid state physics, this is called local 'freezing' - locally, the system looks more like an ordered solid rather than a disordered liquid) but no global ordering. Spin glass can not remain in a single lowest energy state (the ground state). Rather it has many ground states which are never explored on experimental time scales. The freezing of the spins, in spin glasses, is not a deterministic one like ferromagnetic materials. Rather they freeze in random with some memory effect.

Experiments show that the susceptibility obtained by cooling the spin glass system in the presence of a magnetic field yielded a higher value than that obtained by first cooling in zero field and then applying the magnetic field. If the spin glass is cooled below T_c (a certain critical temperature) in the absence of an external field, and then a magnetic field is applied, there is a rapid increase towards a value, called the zero-field-cooled magnetization. This value is less than the field-cooled magnetization. The following phenomenon has also been observed in the measurement of remanent magnetization (the permanent magnetization that remains after the external field is removed). We can cool in the presence of external field, remove the external field and then measure the remanent magnetization. Alternatively, first cool with out the external field, then apply the external field and measure the remanent magnetization after removing the external field. The first value is larger than the second one.

The other peculiarity of the spin glasses is its time dependence, which will be explained now, that makes it different from other magnetic systems. Above the spin glass transition temperature, T_c, the spin glass exhibits typical magnetic behavior. In other words, at temperature above T_c, if an external magnetic field is applied and the magnetization is plotted versus temperature, it follows the typical Curie law (in which magnetization is inversely proportional to temperature). This happens until T_c is reached, at which point the magnetization becomes virtually constant. This is the onset of the spin glass phase. When the external field is removed, the spin glass has a rapid decrease of magnetization to a value called the remnant magnetization, and then a slow decay as the magnetization approaches zero (or some small fraction of the original value). This decay is non-exponential and no single function can fit the curve of magnetization versus time adequately below T_c. This slow decay is particular to spin glasses. If a similar procedure was followed for a ferromagnetic substance, when the external field is removed, there would be a rapid change to a remnant value, but this value is a constant in time. For a paramagnetic material, when the external field is removed, the magnetization rapidly goes to zero. In each case, the change is very rapid and if carefully examined it is exponential decay.

Behind this strange behaviour of spin glasses, according to physicists, there are essentially two major causes. These are *quenched disorder* and *frustration*. The term "quenched disorder" refers to constrained disorder in the interactions between the spins and/or their locations but does not evolve with time. In statistical physics, a system is said to present quenched disorder when some parameters defining its behaviour are random variables which do not evolve with time, i.e., they are quenched or frozen. This is in contrast to annealed disorder, where the random variables are allowed to evolve themselves. Usually the spin orientations depend on several facts such as the interactions, external fields and thermal fluctuations. Their dynamics or thermodynamics will suggest whether to order or not. The spin glass phase is an example of spontaneous cooperative freezing (or order) of the spin orientations in the presence of the constrained disorder of the interactions or spin locations. It is thus "order in the presence of disorder". On the other hand, "frustration" refers to conflicts between interactions and the spin-ordering forces, and not all can be obeyed simultaneously. Frustration arises when pairs of spins get different ordering instructions through the various paths which link them, either directly or via intermediate spins. The relevance of frustration is that it leads to degeneracy or multiplicity of compromises forcing the system to have several ground states.

Keeping these two in mind, in 1975, S. F. Edwards and P. W. Anderson [21] produced a paper, which in the words of Sherrington [37], at one fell swoop recognized the importance of the combination of frustration and quenched disorder as fundamental ingredients, introduced a more convenient model, a new and novel method of analysis, new types of order parameters, a new mean field theory, new approximation techniques and the prediction of a new type of phase transition apparently explaining the observed susceptibility cusp. This paper was a watershed. Edwards and Anderson's new approach was beautifully minimal, fascinating and attractive but also their analysis was highly novel and sophisticated, involving radically new concepts and methods but also unusual and unproven ansätze, as well as several different approaches. In their model, two spins interact if they are neighbour to each other. The same year Sherrington and Kirkpatrick [38] proposed their model with mean field interaction. In this model all spins interact with each other. In both the cases the interaction among the spins were random and driven by Gaussian random variables. Due to rich and complicated correlation structure among the energy over the configuration space of the spins, initially the models were not easy to study analytically. To get some insight into these models, in 1980, B. Derrida [15] proposed a system without any correlation structure over the configuration space. He proposed a solvable model called Random Energy Model (REM) for spin glass theory. In REM, all the random variables are independent and identically distributed but the distribution depends on the number of particles. Like Edwards-Anderson model and SK-model, he considered these random variables to be Gaussian. But this is a toy model since the energy of the system does not depend on the configuration. Amazingly he could show that though this is a very simple model, it exhibits phase transition.

REM has no correlations at all. But the correlation structure in the Edwards-Anderson model and SK-model were very complicated. So the next idea is to study a system which exhibits correlations, but their structure is simple enough to explicitly solve the model. B. Derrida [17] proposed another model for spin glass theory in 1985, by bringing correlations through a tree structure. The tree structure comes from the configuration space. Simply put, he identifies the configuration space as the branches of a tree. This is called Generalized Random Energy Model (GREM), a generalization of the REM. Here also the driving distributions were Gaussian. In this project we will focus ourselves on REM and GREM and some related models.

0.2 Setup and Summary

For an N particle system with classical spins $+1$ or -1, a sequence of $+1$ and -1 of length N gives a configuration of the system. A typical configuration is denoted by $\sigma(N)$ or by σ when N is understood. That is, σ is a sequence of $+1$ and -1 of length N. The space of all possible configurations σ of a system is called configuration space and denoted by Σ_N or simply by 2^N since Σ_N is nothing but $\{+1,-1\}^N$. Now depending on the configuration, the system possesses some energy called Hamiltonian. For a configuration σ, it is denoted by $H_N(\sigma)$. The model is defined through the Hamiltonian. So different models have different Hamiltonian structures. In spin glass theory, the Hamiltonian is considered to be random.

When the system is cooled, it settles down at a configuration where the Hamiltonian is minimized. Hence it is very essential to get information about the configurations where the infimum of the the Hamiltonian is attained and its value. In statistical physics one analyzes this problem via the partition function of the system. The partition function, denoted by $Z_N(\beta)$, is defined as follows:

$$Z_N(\beta) = \sum_{\sigma \in \Sigma_N} e^{-\beta H_N(\sigma)}.$$

Here $\beta \geq 0$ is a parameter, represents the inverse temperature. Sometimes when the Hamiltonian depends on an external field h, we will denote the partition function as $Z_N(\beta, h)$. Now note that among all the summands in the above sum if one takes large β, only that summand will contribute where the Hamiltonian attains the minimum among all possible configurations. On the other hand, if the focus is on maximum, then instead of $-\beta$ one has to consider β in the exponent.

But the information in partition function about the minimum energy is in exponential scale. So it is customary to study the logarithm of the partition function. Also the energy of the system depends on the number of particles in the system and becomes large when N is large. To get some asymptotic result on $\log Z_N(\beta)$, one has to normalize it properly. In this case, $\frac{1}{N}$ is the correct normalization (in some sense). According to statistical physics, $-\frac{1}{\beta N} \log Z_N(\beta)$ is called the *free energy* of the system. Since one is interested in the asymptotic of the free energy, that is, in $-\frac{1}{\beta} \lim\limits_{N\to\infty} \frac{1}{N} \log Z_N(\beta)$, for mathematical purpose we can forget about the $-\frac{1}{\beta}$ term in the definition of free energy. And from now on, we will call $\lim\limits_{N\to\infty} \frac{1}{N} \log Z_N(\beta)$ as the **free energy** of the system.

In statistical physics, there is another important concept called Gibbs' distribution. This is a distribution on the configuration space. According to this, the probability of a configuration σ is proportional to $e^{-\beta H_N(\sigma)}$. In particular, if $G_N(\sigma)$ denotes the Gibbs' probability for a configuration $\sigma \in \Sigma_N$, then

$$G_N(\sigma) = \frac{e^{-\beta H_N(\sigma)}}{Z_N(\beta)}.$$

It is worth noting that, since $H_N(\sigma)$'s are random, the Gibbs' distribution is also random. Note that, Gibbs' distribution is so defined as to give maximum weight to that configuration which has minimum energy. We shall not deal with Gibbs' distributions in this thesis.

Generalized random energy model (GREM) is one model in this theory proposed by B. Derrida [17] in 1985. To describe a version of this fix an integer $n \geq 1$. For N particle system, consider a partition of N into integers $k(i, N) \geq 0$ for $1 \leq i \leq n$ so that $\sum\limits_i k(i, N) = N$. The configuration space 2^N, naturally splits into the product, $\prod 2^{k(i,N)}$ and $\sigma \in 2^N$ can be written as $\sigma_1\sigma_2 \cdots \sigma_n$ with $\sigma_i \in 2^{k(i,N)}$. An obvious n-level tree structure can be brought in the configuration space. Consider an n level tree with $2^{k(1,N)}$ many edges at the first level. These edges are denoted by σ_1, with $\sigma_1 \in 2^{k(1,N)}$. In general, below a typical edge $\sigma_1\sigma_2 \cdots \sigma_{i-1}$ of the $(i-1)$-th level there are $2^{k(i,N)}$ edges at the i-th level denoted by $\sigma_1\sigma_2 \cdots \sigma_{i-1}\sigma_i$ for $\sigma_i \in 2^{k(i,N)}$. Thus a typical branch of the tree reads like $\sigma_1\sigma_2 \cdots \sigma_n$ making a one one correspondence with 2^N, the configuration space. For each i, $1 \leq i \leq n$ and edge $\sigma_1 \cdots \sigma_i$, associate a random variables $\xi(\sigma_1 \cdots \sigma_i)$. All these random variables are i.i.d. $\mathcal{N}(0, N)$. One non random weight, $a_i > 0$ for each level is fixed. In GREM, Hamiltonian for a configuration $\sigma = \sigma_1 \cdots \sigma_n$ is defined as

$$H_N(\sigma) = \sum_{i=1}^{n} a_i \xi(\sigma_1 \sigma_2 \cdots \sigma_i). \tag{0.2.1}$$

When $n = 1$, GREM reduces to REM, another model proposed as a solvable model by B. Derrida [15] in 1980. If $a_1 = 1$ then Hamiltonians of REM are nothing but 2^N many i.i.d. $\mathcal{N}(0, N)$ random variables.

Though it was just a toy model, with correct but heuristic arguments Derrida [16] showed phase transition occurs in REM and in the low temperature the system got completely frozen. In 1986, B. Derrida and E. Gardner [18] gave the solution for the averaged free energy for GREM and in 1987, Capocaccia et al [8] gave a rigorous mathematical justification. Indeed, the convergence holds almost surely as well as in L_p for $1 \leq p < \infty$. In 1989, Galves et al [24] studied the detailed fluctuation of free energy for both the models and further analysis was carried out in 2002 for REM and other models by Bovier et al [7]. In a different direction, Dorlas and Wedagedera [20], in 2001 used the large deviation principle (LDP) [14, 44] to study the free energy for REM. In the next year, Dorlas and Dukes [19] extended this technique to GREM. Though GREM is a little complicated than REM, it is not a realistic model for spin glasses. More realistic models were proposed earlier in 1975 by Edwards and Anderson [21] (EA-model) through nearest neighbour interaction and another by Sherrington and Kirkpatrick [38] (SK-model) by mean field correction in the same year. These are the most complicated models in this theory. Though several heuristic arguments and conjectures [33] were made and several rigorous results were proved [1, 23, 25, 39, 40, 41], it was only in 2002, Guerra and Toninelli [26] showed the almost sure existence of the free energy via interpolation technique and convexity argument. A discussion of the SK-model using stochastic calculus was initiated by Comets and Neveu [12] continued in [11, 2]. For EA-model very little has been known till know. In 2003, the idea of Guerra and Toninelli has been generalized to the GREM cases by Contucci et al [13]. We thank these authors for clarifying their setup.

Note that all this analysis was done with Gaussian driving distributions. In 2004, Carmona and Hu [9] considered non-Gaussian distributions and showed that the free energy of the SK-model does not depend on the driving distribution. Rather, under some moment condition on the driving distributions the free energy of SK-model is universal (see also [10]). It should be noted that earlier already in 1983, Eisele [22] considered a class of distributions with exponentially decaying tails for the REM. He is the first to identify the relevance of LDP to study free energy for REM. He studied completely different types of phase transitions – some kind of iterated large deviation phenomena. For the analysis to go through, he assumed the existence of exponential moments of all orders for the driving distributions. The last two articles are the starting point for this thesis. Now the natural question to ask is, whether there is any universality of free energy in REM as well as in GREM? Moreover, is the existence of exponential moments of all orders necessary?

To answer the above questions our first successful attempt [28] via LDP argument was with double exponential driving distributions. In [28], we provided negative answer to the above questions. First of all, considering H_N to be i.i.d. double exponential driving distribution with parameter 1, we show that the nontrivial free energy is different from that of the Gaussian REM. Though the Hamiltonian does not depend on N, it is interesting, the system exhibits phase transition. Secondly, note that in this case Ee^{tH_N} does not exists for $t \geq 1$. Here in the first chapter, we extract the essence of the argument in [28] and state as

Theorem 0.2.1. *Let $\{\lambda_N\}$ satisfies LDP with a strictly quasi-convex rate function $\mathcal{I}(x)$. For a.e. ω, the sequence of empirical measure $\{\mu_N(\omega)\}$ of 2^N i.i.d. random variables having law λ_N satisfies LDP with rate function $\mathcal{J}$ given by,*

$$\mathcal{J}(x) = \begin{cases} \mathcal{I}(x) & \text{if } \mathcal{I}(x) \leq \log 2 \\ \infty & \text{if } \mathcal{I}(x) > \log 2. \end{cases}$$

We apply this theorem to the known Gaussian case [16,20,34], as well as to double exponential case and further to Weibull type exponentially decaying tail distributions. We also show that the energy in REM is not distribution specific rather *rate specific.* In the compact distribution section we give some partial results when there is no non-trivial rate function for the driving distributions. In the concluding section, we apply the above theorem to discrete distributions – Poisson and Binomial. There we show that even the existence of phase transition depends on the parameter of the underlying distributions. For example, if the Hamiltonian $H_N(\sigma)$ is Binomial with parameter N and p, phase transition takes place only when $p > \frac{1}{2}$.

For GREM, once again our first attempt [29] was with the double exponential driving distributions along with the LDP arguments [19]. The original formulation of GREM in the literature is slightly different from the formulation we mentioned above. In the second chapter, we start with a discussion of this reformulation. Then we bring a general tree structure in GREM and prove a basic fact which is used in the analysis of this chapter as well as for several other models considered in the next chapter. The details are in chapter 2. Briefly, we consider trees all of whose branches extend up to n-th level. Let B_{iN} be the total number of edges at the i-th level and B_N be the number of leaves of the tree. Let s_{iN}^2 be the sum of the squares of the numbers of leaves at the n-th level below each edges at the ith level. If ξ denotes a random variable having the common distribution of the $\xi(\sigma_1\sigma_2\cdots\sigma_i)$, then we have the following.

Theorem 0.2.2. *Let $\triangle = \triangle_1 \times \cdots \times \triangle_n \subset \mathbb{R}^n$. Denote $q_{iN} = P(\xi \in \triangle_i)$ for $1 \leq i \leq n$.*

a) If $\sum_{N \geq n} B_{iN} q_{1N} \cdots q_{iN} < \infty$*, for some* $i, 1 \leq i \leq n$ *then a.s. eventually,*

$$\mu_N(\triangle) = 0.$$

b) If for all $i = 1, \cdots, n$, $\sum_{N \geq n} \frac{s_{iN}^2}{B_N^2 q_{1N} \cdots q_{iN}} < \infty$*, then for any* $\epsilon > 0$ *a.s. eventually,*

$$(1-\epsilon)\mathbf{E}\mu_N(\triangle) \leq \mu_N(\triangle) \leq (1+\epsilon)\mathbf{E}\mu_N(\triangle).$$

In section 2.4, we use this result for GREM with a general family of driving distributions. For fixed $\gamma > 0$, we consider the driving distributions of $\xi(\sigma_1 \cdots \sigma_i)$ having density

$$\phi_{N,\gamma}(x) = \frac{1}{2\Gamma(\frac{1}{\gamma})} \left(\frac{\gamma}{N}\right)^{\frac{\gamma-1}{\gamma}} e^{-\frac{|x|^\gamma}{\gamma N^{\gamma-1}}} \quad -\infty < x < \infty. \tag{0.2.2}$$

Note that when $\gamma = 2$, this is the Gaussian case. We discuss this case systematically in section 2.5 and bring out the essence of this model. Here it is.

For each j, $1 \leq j \leq n$, we have a sequence of probabilities $\{\lambda_N^j, N \geq 1\}$ on $\mathbb{R}$ satisfying LDP with a good, strictly quasi convex rate function $\mathcal{I}_j$ and $\xi(\sigma_1 \cdots \sigma_i) \sim \lambda_N^j$. Define for each ω, $\mu_N(\omega)$ to be the empirical measure on $\mathbb{R}^n$, namely,

$$\mu_N(\omega) = \frac{1}{2^N} \sum_\sigma \delta \left\langle \xi(\sigma_1, \omega), \xi(\sigma_1\sigma_2, \omega), \cdots, \xi(\sigma_1 \cdots \sigma_n, \omega) \right\rangle$$

where $\delta \langle x \rangle$ denotes the point mass at $x \in \mathbb{R}^n$.

Theorem 0.2.3. *Suppose* $\frac{k(j,N)}{N} \to p_j > 0$ *for* $1 \leq j \leq n$*. Then for a.e.* ω*, the sequence* $\{\mu_N(\omega), N \geq 1\}$ *satisfies LDP with rate function* $\mathcal{J}$ *given as follows:*

Supp$(\mathcal{J}) = \{(x_1, \cdots, x_n) : \sum_{k=1}^{j} \mathcal{I}_k(x_k) \leq \sum_{k=1}^{j} p_k \log 2$ *for* $1 \leq j \leq n\}$

and

$$\mathcal{J}(x) = \begin{cases} \sum_{k=1}^{n} \mathcal{I}_k(x_k) & \text{if } x \in \textit{Supp}(\mathcal{J}) \\ \infty & \textit{otherwise.} \end{cases}$$

This result, with the help of Varadhan's Integral lemma [43,14], reduces the problem of free energy to merely calculation of certain infimum. In section 2.6, we solve this variational problem for general n and produce the explicit energy expression in the case of $\gamma > 1$ and $\gamma = 1$ by different arguments. Further, for $\gamma \geq 1$, we characterize the energy function for GREM and show that the energy function is continuous in γ. For $0 < \gamma < 1$, we only give the energy expression for $n = 2$. The beauty of the above theorem (Theorem 0.2.3) is that, it allows us to consider different distributions at different levels of the under lying tree. This we considered in [30] and here in section 2.7. Even the simple case, $n = 2$ the model exhibits a lot of interesting phenomena. For example, consider a 2 level GREM with exponential driving distribution at the first level and Gaussian in the second, and give equal weights at the two levels, that is, $a_1 = a_2$. Then even if $p_2 = 0.00001$ (very small) the system reduces to a Gaussian REM. On the other hand, if we consider a 2 level GREM with Gaussian driving distribution at the first level and exponential in the second, the system will never reduce to a Gaussian REM. Moreover, in either case, the system will never reduce to that of an exponential REM.

In the third chapter, we randomize the underlying trees. To keep the same number of furcations for all edges at given level, for fixed N, we take one Poisson random variable at each level to determine the number of furcations. We called this model as *regular Poisson GREM*. On the other hand, it is possible to keep the number of furcations depend on the edge. In other words, for each edge we can associate a Poisson random variable to determine the number of furcations for this edge. This we called *Poisson GREM*. We discussed multinomial variation also using results from [27]. These are all different methods to randomize the tree. Note that the configuration space is no longer 2^N. These models are interesting and Theorem 0.2.2 above is powerful enough to handle these models. However, in all these cases the free energy remains same as in the usual GREM. Whether there are other interesting tree structures that exhibit peculiar phenomena is not clear to us. As far as our knowledge goes, the GREM with randomized (or even nonrandomized but general) trees is not discussed in the literature.

In 2006 Bolthausen and Kistler [3] proposed a model (BK-GREM) bypassing the *ultrametricity* in the configuration space. Even in this model, they have shown that the energy of the system is again a suitable GREM energy. In section 3.4, we provide a proof via LDP arguments. Then in section 3.5, we construct $n!$ many GREMs, one corresponding to each permutation of the set $\{1, 2, \cdots, n\}$ by manipulating the weights from BK-GREM. We characterize a class of permutations so that (1) the corresponding GREM energy will be the same for all the permutations in that class, (2) this energy is the minimum over all possible $n!$ many GREMs and (3) this is the energy of the BK-GREM. Bolthausen and Kistler [3], have shown that the energy of BK-GREM is the infimum over GREM energies corresponding to all possible *chains*. Our analysis shows that instead of considering all chains, one needs to consider $n!$ many GREMs. This is still a large number. We conclude this chapter, by defining one model, called block tree GREM, where the free energy is maximum of all possible $n!$ many n level GREMs, rather than minimum as in the BK-GREM.

In the last chapter, we introduce a new model, word GREM. This brings out the crucial role played by LDP in all the earlier models. Here we start with a distribution having finite mean and λ_N denotes the law of the sample mean (of size N). Then Cramer's theorem [14] suggests that the sequence $\{\lambda_N\}$ satisfies LDP with a convex rate function given in terms of the Fenchel-Legendre transformation of the starting distribution. We consider a set of n symbols $I = \{\varsigma_1, \varsigma_2, \cdots, \varsigma_n\}$ and take S to be any collection of finite number of words formed by these n symbols. As earlier, we consider $k(i, N) \geq 0,\ 1 \leq i \leq n$ as a partition of N so that the configuration space 2^N splits as $\prod_{i=1}^{n} 2^{k(i,N)}$. For $s = \varsigma_{i_1}\varsigma_{i_2}\cdots\varsigma_{i_l} \in S$ and a configuration $\sigma = \langle\sigma^1, \cdots, \sigma^n\rangle$ where $\sigma^i \in 2^{k(i,N)}$, we denote $\sigma(s) = \langle\sigma^{i_1}, \sigma^{i_2}, \cdots, \sigma^{i_l}\rangle$. For $s \in S$, let λ^s be a probability on $\mathbb{R}$ having finite mean. Let λ_N^s denote the distribution of the mean of the first N random variables of an i.i.d. sequence with common law λ^s. For the N particle system, we have the following. For each $s \in S$ and each $\sigma \in 2^N$, we have a random variable $\xi(s, \sigma(s))$. These are independent random variables. For fixed s, they are identically distributed and the common distribution is λ_N^s. Then for a configuration $\sigma = \langle\sigma_1, \cdots, \sigma_N\rangle \in 2^N$, we define the Hamiltonian in word GREM as

$$H_N(\sigma) = Nf(\xi(\sigma)) + h\sum_{i=1}^{N}\sigma_i, \tag{0.2.3}$$

where $f : \mathbb{R}^S \to \mathbb{R}$ is a continuous function, $\xi(\sigma) = (\xi(s, \sigma(s)))_{s\in S}$ and $h \geq 0$ is the intensity of the external field.

We present a large deviation proof for the existence of the free energy for this model and apply the analysis to known [16] REM with external field.

This model includes REM, GREM and BK-GREM and may perhaps include models truly more general than these. Further, it allows external field. Moreover, different driving distributions can be used at different words in the collection.

In this project, we did not consider the analysis of Gibbs' distribution. With Gaussian driving distribution, there are several results for REM [42,4] and for GREM [5,6]. See also [35,31,32,36]. For exponential driving distribution we verified that for REM, in the high temperature regime, Gibbs's distribution converges to the uniform distribution [28] where as in the low temperature regime it converges to the Poisson-Dirichlet distribution. This is similar to that of Gaussian REM. So is it true for any other distributions considered in this thesis? Since we do not have anything substantial to say regarding this issue, we have not considered.

0.3 Large Deviation Terminology

Recall that

$$Z_N(\beta) = \sum_{\sigma} e^{-\beta H_N(\sigma)} = 2^N \mathbf{E}_\sigma e^{-\beta H_N(\sigma)}$$

where $\mathbf{E}_\sigma$ is expectation w.r.t uniform probability on 2^N space. And hence,

$$\frac{1}{N}\log Z_N(\beta) = \log 2 + \frac{1}{N}\log \mathbf{E}_\sigma e^{-N\beta\frac{H_N(\sigma)}{N}}.$$

The last term in the above equation is well known expression in the Laplace's principle. It is indeed

$$\frac{1}{N}\log\int e^{-Nf}d\mu_N,$$

where μ_N is the uniform probability on the space 2^N. The only trouble is μ_N are on different spaces. If we transport μ_N to $\mathbb{R}$ by the map $\sigma \mapsto \frac{H_N(\sigma)}{N}$, then we will arrive at exactly the Laplace type situation, where Varadhan's integral lemma comes to rescue. Since H_N depends on ω, the transported probability will be random. So the application of LDP needs careful attention.

Since there are several terminologies (using ϵ or using N etc) for large deviations, we fix our terminology now and recall some known facts. Let $\mathbb{X}$ be a Polish space.

Definition 0.3.1. A function $\mathcal{I} : \mathbb{X} \to \mathbb{R}$ is called a lower semicontinuous function if for any $a \in \mathbb{R}$, the set $\{x : \mathcal{I}(x) \leq a\}$ is a closed set. It will be called good if the set $\{x : \mathcal{I}(x) \leq a\}$ is a compact set.

The following two properties of lower semicontinuous function are worth mentioning.

Proposition 0.3.1. *Let f be a lower semicontinuous function. Then for any x,*

$$\sup_{G:\text{neighbourhood of } x} \inf_{y\in G} f(y) = f(x).$$

Proposition 0.3.2. *Let f be a good function and $\{F_n\}_n$ be a sequence of closed sets so that $F_{n+1} \subseteq F_n$ for every n and $\cap_n F_n = \{x_0\}$. Then*

$$f(x_0) = \lim_n \inf_{y\in F_n} f(y).$$

Definition 0.3.2. Let $\{\mu_N\}$ be a sequence of probabilities on $\mathbb{X}$. $\{\mu_N\}$ is said to satisfy large deviation principle with rate function $\mathcal{I}$ if

1. $\mathcal{I} : \mathbb{X} \to [0,\infty]$ is a lower semicontinuous function,

2. for any Borel set B,

$$-\inf_{x\in B^0}\mathcal{I}(x) \leq \liminf_N \frac{1}{N}\log\mu_N(B) \leq \limsup_N \frac{1}{N}\log\mu_N(B) \leq -\inf_{x\in\overline{B}}\mathcal{I}(x).$$

Further, if for $0 \leq a < \infty$, the set $\{x : \mathcal{I}(x) \leq a\}$ is a compact set, then $\mathcal{I}$ is called a good rate function.

A sufficient condition for the existence of LDP is the following:

Proposition 0.3.3. *Let $\mathbb{X}$ be a Polish space. Let $\mathcal{A}$ be an open base for $\mathbb{X}$. Let $\{\mu_N\}$ be a sequence of probabilities on $\mathbb{X}$. For each $A \in \mathcal{A}$, let $L_*(A) = -\liminf_N \frac{1}{N} \log \mu_N(A)$ and $L^*(A) = -\limsup_N \frac{1}{N} \log \mu_N(A)$. Suppose for every $x \in \mathbb{X}$,*

$$\sup_{x \in A \in \mathcal{A}} L_*(A) = \sup_{x \in A \in \mathcal{A}} L^*(A) = \mathcal{I}(x) \qquad \textit{(say).}$$

Assume moreover that either $\{\mu_N\}$ is eventually supported on a compact set or the sequence is exponentially tight, that is, given any $\alpha < \infty$, there is a compact set K such that $\limsup_N \frac{1}{N} \log \mu_N(K^c) < -\alpha$.

Then the sequence $\{\mu_N\}$ satisfies LDP with rate function $\mathcal{I}$.

The next proposition is a variation of well-known Varadhan's integral lemma, which will suggest that only we have to calculate some infimum to get the free energy limit.

Proposition 0.3.4. *Suppose the sequence of probabilities $\{\mu_N\}$ on a Polish space $\mathbb{X}$ satisfies LDP with rate function $\mathcal{I}$ and, moreover, μ_N are eventually supported on a compact set C. Then for any continuous function $f : \mathbb{X} \to \mathbb{R}$,*

$$\lim_{N \to \infty} \frac{1}{N} \log \int e^{-Nf} d\mu_N = -\inf_{x \in C} \{f(x) + \mathcal{I}(x)\}.$$

We need the following known as Cramer's Theorem.

Theorem 0.3.5. *Let $X_1, X_2, \cdots$ be i.i.d. real valued random variables with $EX_1 < \infty$ and λ_N be the law of their sample mean. Then the sequence of probabilities $\{\lambda_N\}$ satisfies LDP with a convex rate function $\mathcal{I}$ given by*

$$\mathcal{I}(x) = \sup_{\mathcal{D}} \{\theta x - \log E e^{\theta X_1}\}, \tag{0.3.1}$$

where $\mathcal{D} = \{\theta : \log E e^{\theta X_1} < \infty\} \subseteq \mathbb{R}$.

The following is also true.

Lemma 0.3.6. *Let $\mathcal{I}$ be as in the above theorem. If $\mathcal{D}_{\mathcal{I}} = \{x : \mathcal{I}(x) < \infty\}$ and $\overline{x} = EX_1 < \infty$, then*

1. $\mathcal{I}(\overline{x}) = 0$,
2. *$\mathcal{I}$ is strictly decreasing on $\{x \leq \overline{x}\} \cap \mathcal{D}_{\mathcal{I}}$,*
3. *$\mathcal{I}$ is strictly increasing on $\{\overline{x} \leq x\} \cap \mathcal{D}_{\mathcal{I}}$.*

Chapter 1

The Random Energy Model

In this chapter we discuss a toy model of spin glass theory, called Random Energy Model (REM). In the literature [16, 42], this is driven with Gaussian distributions. In [22], Eisele discussed the model with more general distributions, particularly with regularly varying distributions and the relevance of large deviation methods in this context. We study the model with other types of distributions. For instance, the driving distributions could be exponential or more generally Weibull. Or they could be compactly supported etc. Our discussion mainly relies on the idea of Dorlas and Wedagedera. In [20], they first used the large deviation techniques to get the asymptotics of the free energy.

After defining the model in the first section, we give a general large deviation result in the section 1.2 and apply the results in REM with diverse distributions in section 1.3. In section 1.4, we give partial results with compact distributions where we could not use the large deviation results. We conclude this chapter by considering the model driven by some discrete distributions.

1.1 Setup

In this model, proposed originally by B. Derrida [15], for each N, the Hamiltonian $H_N(\sigma)$ are independent over $\sigma \in \Sigma_N$. Derrida considered them to be centered Gaussian with variance N. In spite of the simplicity of this model, in [16], he showed the existence of phase transition. Using the entropy energy equation, he evaluated the limiting annealed free energy $\lim_N \frac{1}{N} E \log Z_N(\beta)$ and showed that for low temperature, that is, for β large, free energy becomes linear in β. It is known that, in fact, $\frac{1}{N} \log Z_N(\beta)$ converges a.s. [20].

1.2 Main Results

Let us consider a sequence of probabilities $(\lambda_N, N \geq 1)$ on $\mathbb{R}^n$. Assume that $\{\lambda_N\}$ satisfies large deviation principle (LDP) with a strictly quasi-convex good rate function $\mathcal{I}(x)$. An extended real valued function f, which may take the value $+\infty$ but not $-\infty$, defined on a convex set will be called a strictly quasi-convex function if for any two distinct x_1 and x_2 in $\{f(x) < \infty\}$ and for each $\theta \in (0,1)$ we have $f(\theta x_1 + (1-\theta)x_2) < \max\{f(x_1), f(x_2)\}$. For every N, let ξ_i, $1 \leq i \leq 2^N$ be i.i.d. random variables ($\mathbb{R}^n$ valued) with distribution λ_N. These random variables, of course, depend on N but to ease the notation we are suppressing their dependence on N. For every sample point ω, we define $\mu_N(\omega)$ to be the empirical measure on $\mathbb{R}^n$, namely $\mu_N(\omega) = \frac{1}{2^N}\sum \delta\langle \xi_i(\omega)\rangle$. Here $\delta\langle x\rangle$ denote the point mass at x. Now we are ready to state our first theorem.

Theorem 1.2.1. *For a.e. ω the sequence $\{\mu_N(\omega)\}$ is supported on a compact set and satisfies LDP with rate function $\mathcal{J}$ given by,*

$$\mathcal{J}(x) = \begin{cases} \mathcal{I}(x) & \textit{if } \mathcal{I}(x) \leq \log 2 \\ \infty & \textit{if } \mathcal{I}(x) > \log 2. \end{cases}$$

Proof. **Step 1:** *Let $\triangle$ be an open subset of $\mathbb{R}^n$. If $\sum 2^N \lambda_N(\triangle) < \infty$, then almost surely eventually $\mu_N(\triangle) = 0$.*

Indeed, using P for the probability on the space where the random variables are defined,

$$P(\mu_N(\triangle) > 0) = P(\xi_i \in \triangle \text{ for some } i) \leq 2^N \lambda_N(\triangle).$$

Now Borel - Cantelli completes the proof.

Step 2: *Let $\triangle$ be an open subset of $\mathbb{R}^n$. If $\sum \frac{1}{2^N \lambda_N(\triangle)} < \infty$, then for any $\epsilon > 0$, almost surely eventually*

$$(1-\epsilon)\lambda_N(\triangle) \leq \mu_N(\triangle) \leq (1+\epsilon)\lambda_N(\triangle).$$

Indeed,

$$\text{Var } \mu_N(\triangle) = \mathbf{E}\left(\frac{1}{2^N}\sum 1_\triangle(\xi)\right)^2 - \lambda_N^2(\triangle) \leq \frac{1}{2^N}\lambda_N(\triangle).$$

Now Chebyshev yields

$$P\left\{|\mu_N(\triangle) - \lambda_N(\triangle)| > \epsilon\lambda_N(\triangle)\right\} \leq \frac{1}{\epsilon^2 2^N \lambda_N(\triangle)}$$

and Borel-Cantelli completes the proof.

Since $\mathcal{I}$ is strictly quasi-convex, the set $\{\mathcal{I}(x) = \log 2\}$ does not contain any line segment, we can choose a countable open base $\mathfrak{B}$ such that for every $\triangle \in \mathfrak{B}$ either $\overline{\triangle} \cap \{\mathcal{I}(x) \leq \log 2\} = \varnothing$ or $\triangle \cap \{\mathcal{I}(x) < \log 2\} \neq \varnothing$. For instance, we could choose $\mathfrak{B}$ to be the collection all open boxes such that (i) $\mathcal{I}$ value at a corner point is different from $\log 2$; (ii) each co-ordinate of a corner point is either rational or $\pm\infty$.

Step 3: *Let $\mathcal{I}(x) > \log 2$. Then almost surely,* $\sup\limits_{x\in\triangle\in\mathfrak{B}} \{-\liminf \frac{1}{N}\log \mu_N(\triangle)\}$ *as well as* $\sup\limits_{x\in\triangle\in\mathfrak{B}} \{-\limsup \frac{1}{N}\log \mu_N(\triangle)\}$ *are ∞.*

Since $\mathcal{I}(x) > \log 2$, pick $\triangle_0 \in \mathfrak{B}$ such that $x \in \triangle_0$ and $\overline{\triangle}_0 \cap \{\mathcal{I}(x) \leq \log 2\} = \varnothing$. Then $\limsup \frac{1}{N}\log \lambda_N(\triangle_0) \leq -\inf\limits_{y\in\overline{\triangle}_0} \mathcal{I}(y) = -L < -\log 2$. Fix $\alpha > 0$ such that $-L < -\log 2 - \alpha$. Then for sufficiently large N, $\frac{1}{N}\log \lambda_N(\triangle_0) \leq -\log 2 - \alpha$, that is, $\lambda_N(\triangle_0) \leq 2^{-N}e^{-N\alpha}$. In other words, $2^N\lambda_N(\triangle_0) \leq e^{-N\alpha}$ for all large N. Thus by Step 1, a.s. eventually $\mu_N(\triangle_0) = 0$ and hence the claim.

Incidentally, this also shows the following. If K is the compact set $\{x : \mathcal{I}(x) \leq \log 2\}$ then consider a bounded open box $\triangle \in \mathfrak{B}$ such that $K \subset \triangle$. Clearly, $\overline{\triangle}^c$ is union of $2n$ many boxed from $\mathfrak{B}$. The above argument shows that μ_N is a.s. eventually zero for each of these $2n$ boxes. This shows that the sequence $\{\mu_N\}$ is a.s. eventually supported on a compact set, namely, $\overline{\triangle}$.

Step 4: *Let $\mathcal{I}(x) \leq \log 2$. Then almost surely,* $\sup\limits_{x\in\triangle\in\mathfrak{B}} \{-\liminf \frac{1}{N}\log \mu_N(\triangle)\}$ *as well as* $\sup\limits_{x\in\triangle\in\mathfrak{B}} \{-\limsup \frac{1}{N}\log \mu_N(\triangle)\}$ *are $\mathcal{I}(x)$.*

Fix $\triangle \in \mathfrak{B}$ such that $x \in \triangle$. Then $\liminf \frac{1}{N}\log \lambda_N(\triangle) \geq -\inf\limits_{y\in\triangle} \mathcal{I}(y) = -L > -\log 2$, where the last inequality is a consequence of the strict quasi-convexity of $\mathcal{I}$.

Fix $\alpha > 0$ so that $-L > -\log 2 + \alpha$. So for large N, $\frac{1}{N} \log \lambda_N(\triangle) > -\log 2 + \alpha$, that is, $\lambda_N(\triangle) > 2^{-N} e^{N\alpha}$. In other words, $2^N \lambda_N(\triangle) > e^{N\alpha}$. Now use Step 2, for any $\epsilon \in (0, 1)$, eventually

$$(1-\epsilon)\lambda_N(\triangle) \leq \mu_N(\triangle) \leq (1+\epsilon)\lambda_N(\triangle).$$

Hence by definition of LDP, we have eventually

$$-\mathcal{I}(\triangle) \leq \liminf_{N\to\infty} \frac{1}{N} \log \mu_N(\triangle) \leq \limsup_{N\to\infty} \frac{1}{N} \log \mu_N(\triangle) \leq -\mathcal{I}(\overline{\triangle}), \tag{1.2.1}$$

where as usual $\mathcal{I}(A) = \inf\limits_{x \in A} \mathcal{I}(x)$.

From the first part of the above inequality we have,

$$\sup_{\triangle \in \mathfrak{B}: x \in \triangle} \{-\liminf_{N} \frac{1}{N} \log \mu_N(\triangle)\} \leq \sup_{\triangle \in \mathfrak{B}: x \in \triangle} \mathcal{I}(\triangle) \ \leq \ \mathcal{I}(x). \tag{1.2.2}$$

Moreover, for every $\triangle \in \mathfrak{B}$ such that $x \in \triangle$ using the right side inequality of (1.2.1), we have

$$\limsup_{N} \frac{1}{N} \log \mu_N(\triangle) \leq -\mathcal{I}(\overline{\triangle}).$$

Let $\mathfrak{B}_x = \{\triangle_k \in \mathfrak{B} : k \geq 1\}$ be a subclass of $\mathfrak{B}$ so that $\overline{\triangle}_{k+1} \subset \triangle_k$ for every k and $\cap_k \triangle_k = \{x\}$. Then

$$\begin{aligned} \sup_{\triangle \in \mathfrak{B}: x \in \triangle} \{-\limsup_{N} \frac{1}{N} \log \mu_N(\triangle)\} &\geq \sup_{\triangle \in \mathfrak{B}: x \in \triangle} \mathcal{I}(\overline{\triangle}) \\ &\geq \sup_{\triangle \in \mathfrak{B}_x} \mathcal{I}(\overline{\triangle}) \\ &= \lim_{k} \mathcal{I}(\overline{\triangle}_k) \\ &= \mathcal{I}(x). \end{aligned} \tag{1.2.3}$$

The last equality follows as $\mathcal{I}$ is a good lower semicontinuous function (see Proposition 0.3.2).

From (1.2.2) and (1.2.3), it follows that

$$\sup_{\triangle\in\mathfrak{B}:x\in\triangle}\{-\liminf_N \frac{1}{N}\log\mu_N(\triangle)\} = \sup_{\triangle\in\mathfrak{B}:x\in\triangle}\{-\limsup_N \frac{1}{N}\log\mu_N(\triangle)\} = \mathcal{I}(x).$$

Now proof of Theorem 1.2.1 is completed by appealing to Proposition 0.3.3 and observing that $\{\mu_N\}$ is eventually supported on a compact set. □

Remark 1.2.1. The fact that $\mathcal{I}$ is a *good* rate function is essential in the above theorem to conclude that almost surely eventually the sequence $\{\mu_N(\omega)\}$ is supported on a compact subset of $\mathbb{R}^n$.

Remark 1.2.2. Observe that the strict quasi-convexity of the rate function in the above theorem is a technical assumption. On real line that assumption can be replaced by the assumption: $\mathcal{I}$ is strictly monotone on $\{x : \mathcal{I}(x) \in (0,\infty)\}$ or by the assumption that $\{x : \mathcal{I}(x) = \log 2\}$ is a nowhere dense set. Such a condition is needed only to ensure that there exists a countable base as mentioned in the above proof.

The implication of the above theorem in REM is amazing. To see this, let us assume that $\{\lambda_N\}$ is a sequence of probabilities on $\mathbb{R}$ and satisfies large deviation principle with a good rate function $\mathcal{I}$. Let us assume also that $\mathcal{I}$ be strictly quasi-convex or satisfies any one of the conditions in remark 1.2.2. For fixed N, let us consider 2^N i.i.d. random variables $\xi(\sigma)$, $1 \le \sigma \le 2^N$ distributed like λ_N. We can identify these 2^N many σ with the elements of $\Sigma_N = \{+1,-1\}$. Let us define the Hamiltonian for $\sigma \in \Sigma_N$ to be

$$H_N(\sigma) = N\xi(\sigma).$$

Now note that the partition function can be written as

$$Z_N(\beta) = 2^N E_\sigma e^{-\beta H_N(\sigma)},$$

where E_σ is the expectation with respect to uniform probability on the Σ_N space. Hence

$$\lim_N \frac{1}{N}\log Z_N(\beta) = \log 2 + \lim_N \frac{1}{N}\log\int e^{-N\beta x}d\mu_N(x).$$

By Theorem 1.2.1, the induced probabilities $\{\mu_N\}$ are a.s. eventually supported on a compact set. That is, for almost every ω, there is a compact set K_ω such that $\{\mu_N(\omega),\ N \ge 1\}$ are all supported on K_ω. Moreover, by previous theorem they satisfy

LDP with a good rate function a.s. Hence to find the existence of $\lim_N \frac{1}{N}\log Z_N(\beta)$, we can use Varadhan's integral lemma with any continuous function, in particular, $f(x) = \beta x$ on $\mathbb{R}$. This will lead to the following:

Theorem 1.2.2. *If the sequence of probabilities $\{\lambda_N\}$ satisfies LDP with strictly quasi-convex good rate function $\mathcal{I}$, then almost surely,*

$$\lim_{N\to\infty} \frac{1}{N}\log Z_N(\beta) = \log 2 - \inf_{\{\mathcal{I}(x)\le \log 2\}} (\beta x + \mathcal{I}(x)) . \tag{1.2.4}$$

Thus in the REM, the existence of limiting free energy is just a corollary of large deviation principle. To get the the expression of the free energy one has to solve the variational formula. Hence the calculation of asymptotics of free energy reduces to calculation of the above infimum.

Remark 1.2.3. In the literature [16,20,42], for REM, the Hamiltonian $H_N(\sigma)$ is defined as $\sqrt{N}\xi(\sigma)$, with $\xi(\sigma) \sim \mathcal{N}(0, N)$. In our case with the Gaussian driving distribution it is same as $N\xi(\sigma)$ where $\xi(\sigma) \sim \mathcal{N}(0, \frac{1}{N})$. But the large deviation technique allows us to consider H_N to be any continuous function of $\xi(\sigma)$. In other words, if $f(x)$ is a continuous function on $\mathbb{R}$ then one can define the random Hamiltonian $H_N(\sigma) = Nf(\xi(\sigma))$ where $\xi(\sigma) \sim \lambda_N$. In that case, the above theorem will reduce to

Theorem 1.2.3. *If a sequence of probabilities $\{\lambda_N\}$ on $\mathbb{R}$ satisfies LDP with strictly quasi-convex good rate function $\mathcal{I}$ and $H_N(\sigma) = Nf(\xi(\sigma))$ where $\xi(\sigma) \sim \lambda_N$ and f is a continuous function on $\mathbb{R}$, then almost surely,*

$$\lim_{N\to\infty} \frac{1}{N}\log Z_N(\beta) = \log 2 - \inf_{\{\mathcal{I}(x)\le \log 2\}} (\beta f(x) + \mathcal{I}(x)) . \tag{1.2.5}$$

Of course, the appearance of f above makes it more general, but this could be obtained from (1.2.4) by contraction principle of large deviation techniques. Different choices of functions f allows us to consider the Hamiltonian driven by other distribution which can be obtained as a function of known distributions. For instance, if we consider $f(x) = x^2$ then we can get the information of the model when its Hamiltonian is an appropriate χ^2 if λ_N as $\mathcal{N}(0, \frac{1}{N})$. Also we can consider several other functions, for which we do not know the corresponding closed form expression of the distribution of the Hamiltonian. For example, $f(x) = x\cos(1000\pi|x|^{|x|})$ etc.

Remark 1.2.4. A close look at $\inf_{\{\mathcal{I}(x)\le \log 2\}} (\beta f(x) + \mathcal{I}(x))$ suggests the following: If we consider the Hamiltonian of an N-particle system to be an odd function f of random variables $\xi(\sigma) \sim \lambda_N$ and if the sequence $\{\lambda_N\}$ satisfies LDP with a quasi-convex good rate function taking value 0 at the origin, then the contribution for the limiting free energy $\lim\limits_{N\to\infty} \frac{1}{N} \log Z_N(\beta)$ comes from only that part where the function $f(x)$ is negative. More precisely, f being an odd function $f(0) = 0$. Since $\mathcal{I}(0) = 0$, infimum in (1.2.5) is non-positive. So only points x where $f(x) \leq 0$ need to be considered while calculating the infimum. Thus, we have, almost surely,

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{\{\mathcal{I}(x)\leq \log 2\}^-} \left(\mathcal{I}(x) + \beta f(x)\right),$$

where $\{\mathcal{I}(x) \leq \log 2\}^- = \{\mathcal{I}(x) \leq \log 2\} \cap \{f(x) \leq 0\}$. For example, when $f(x) = x$ we have the following

Corollary 1.2.4. *If $f(x) = x$ and $\mathcal{I}(0) = 0$, then almost surely,*

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{\{\mathcal{I}(x)\leq \log 2\}^-} \left(\mathcal{I}(x) + \beta x\right).$$

One can see that the contribution to the free energy is only from the negative values of the random variable. This can be made precise as follows: Let $\{\lambda_N\}$ and $\{\nu_N\}$ be two sequences of probabilities satisfying LDP with a good strictly quasi-convex rate functions $\mathcal{I}_1$ and $\mathcal{I}_2$ respectively so that $\mathcal{I}_1(0) = \mathcal{I}_2(0) = 0$ and $\mathcal{I}_1(x) = \mathcal{I}_2(x)$ for $x \leq 0$. Then consideration of $\xi_N \sim \lambda_N$ or $\xi_N \sim \nu_N$ will lead to the same limiting free energy. In other words, symmetry of the random variables does not play any role in the evaluation of limiting free energy. To illustrate, if we consider the density of λ_N given by,

$$\phi_N(x) = \begin{cases} \frac{1}{2}\sqrt{\frac{N}{2\pi}} e^{-\frac{1}{2}Nx^2} & \text{for } x \geq 0 \\ \frac{N}{2} e^{Nx} & \text{for } x < 0 \end{cases}, \tag{1.2.6}$$

then from the discussion of our next section, it will be clear that this sequence $\{\lambda_N\}$

satisfies LDP with rate function

$$\mathcal{I}(x) = \begin{cases} \frac{x^2}{2} & \text{for } x \geq 0, \\ -x & \text{for } x < 0 \end{cases}$$

and hence $\{\mathcal{I}(x) \leq \log 2\} = [-\log 2, \sqrt{2 \log 2}]$. Here the distribution of λ_N is of Gaussian form in the positive part of the real line whereas on the negative part it is of exponential nature. If $\xi_N \sim \lambda_N$, then as $\inf_{\{0<x\leq\sqrt{2\log 2}\}} \left(\frac{1}{2}x^2 + \beta x\right) \geq 0$ the above corollary will reduce to

Corollary 1.2.5. *If λ_N has density ϕ_N given by (1.2.6), then almost surely,*

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{\{-\log 2\leq x\leq 0\}} (-x + \beta x) .$$

Hence the Gaussian part of the random variables does not contribute to the limiting free energy. Similarly, if we consider the density of λ_N to be

$$\phi_N(x) = \begin{cases} \frac{N}{2} e^{-Nx} & \text{for } x \geq 0 \\ \frac{1}{2}\sqrt{\frac{N}{2\pi}} e^{-\frac{1}{2}Nx^2} & \text{for } x < 0 \end{cases}, \tag{1.2.7}$$

then the rate function will be

$$\mathcal{I}(x) = \begin{cases} x & \text{for } x \geq 0 \\ \frac{1}{2}x^2 & \text{for } x < 0. \end{cases}$$

In this case Corollary 1.2.4 will reduce to

Corollary 1.2.6. *If λ_N has density ϕ_N given by (1.2.7), then almost surely,*

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{\{-\sqrt{2\log 2}\leq x\leq 0\}} \left(\frac{1}{2}x^2 + \beta x\right) .$$

Here the exponential nature of the random variable on the positive side does not play any role.

Remark 1.2.5. Suppose that the sequence of probabilities $\{\lambda_N\}$ is supported on $[0,\infty)$ and satisfies LDP with a quasi-convex rate function $\mathcal{I}$ with $\mathcal{I}(0) = 0$. For example, fix a number $\gamma > 0$ and put $\mathcal{I}(x) = x^\gamma$ for $x \geq 0$ and ∞ for $x < 0$ is such a rate function. It follows from Corollary 1.2.4 that, when $f(x) = x$, then almost surely

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{\{\mathcal{I}(x)\leq \log 2\}^-} (\mathcal{I}(x) + \beta x)\,.$$

As the sequence $\{\lambda_N\}$ is supported on non-negative real line, $\mathcal{I}(x) = \infty$ for $x < 0$. Hence $\{\mathcal{I}(x) \leq \log 2\}^- = \{\mathcal{I}(x) \leq \log 2\} \cap \{f(x) \leq 0\} = \{0\}$ and

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \log 2$$

almost surely. In this case, the system will not show phase transition.

The examples given above are rather artificial. Of course, there are natural examples of random variables $\xi(\sigma)$ whose distributions satisfy large deviation principle with a good convex rate function. In the following sections, we discuss some examples.

1.3 Distribution with exponentially decaying Tail

In this section, we consider the driving sequence of distributions (λ_N) such that for $x > 0$, $\lambda_N[-x,x]^c \sim e^{-Nx^\gamma}$ for some $\gamma > 0$.

1.3.1 Gaussian Distribution

Our first natural example is the Gaussian distribution well studied in the literature [16, 34, 20, 42]. Let λ_N be the centered Gaussian probability with variance $\frac{1}{N}$. That is, having density $\sqrt{\frac{N}{2\pi}} e^{-N\frac{x^2}{2}}$, for $-\infty < x < \infty$. It is obvious that $\lambda_N \Rightarrow 0$ as $N \to \infty$. The following is well known. It can also be obtained from Cramer's theorem 0.3.5. Since the proof is simple, we give it.

Proposition 1.3.1. *The sequence $\{\lambda_N\}$ satisfies LDP with rate function $\mathcal{I} = \frac{x^2}{2}$ on $\mathbb{R}$.*

Proof. Let $\triangle \subset \mathbb{R}$ be an open interval. Let $m = \inf_{x\in\triangle} |x|$, $M = \sup_{x\in\triangle} |x|$, and $q_N = \lambda_N(\triangle)$. With this notation, note that, we have

$$q_N \leq 2\sqrt{\frac{N}{2\pi}} \int_m^M e^{-N\frac{x^2}{2}} dx < \int_{\sqrt{N}m}^{\sqrt{N}M} e^{-\frac{x^2}{2}} dx < \int_{\sqrt{N}m}^{\infty} e^{-\frac{x^2}{2}} dx \leq \frac{1}{\sqrt{N}m} e^{-\frac{Nm^2}{2}}, \quad (1.3.1)$$

with the understanding that when $m_i = 0$, the last expression is $\frac{1}{2}$ and

$$q_N \geq \frac{1}{\sqrt{2\pi}} \int_{\sqrt{N}m}^{\sqrt{N}M} e^{-\frac{x^2}{2}} dx > \frac{1}{2} \int_{\sqrt{N}m}^{\sqrt{N}(m+\delta)} e^{-\frac{x^2}{2}} dx > \frac{\sqrt{N}\delta}{2} e^{-\frac{N}{2}(m+\delta)^2}, \quad (1.3.2)$$

for any $0 < \delta < M - m$.

From above two inequalities, we can conclude that for any open interval $\triangle$, the limit, $\lim_{N\to\infty} \frac{1}{N} \log \lambda_N(\triangle) = -\frac{m^2}{2}$. Once again, Proposition 0.3.3 completes the proof. □

Remark 1.3.1. Note that, here $\mathcal{I}(x) = \frac{x^2}{2}$ is a continuous function with compact level sets. Not only that, it is a convex function and hence quasi-convex.

As a consequence of Theorem 1.2.3, if for $\sigma \in \Sigma_N$ the random variables $\xi(\sigma) \sim \lambda_N$ and the Hamiltonian $H_N(\sigma) = Nf(\xi(\sigma))$ with any continuous function f on $\mathbb{R}$, we get the following:

Corollary 1.3.2. *If $\lambda_N \sim \mathcal{N}(0, \frac{1}{N})$, then almost surely,*

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{x^2 \leq 2\log 2} \left(\beta f(x) + \frac{x^2}{2} \right).$$

Taking $f(x) = x$, we will get the classical case where the Hamiltonian for N-particle system H_N is Gaussian with mean 0 and variance N. Note that, $\inf_{x^2\leq 2\log 2} \left(\beta x + \frac{x^2}{2}\right) = \inf_{0\leq x\leq\sqrt{2\log 2}} \left(\frac{x^2}{2} - \beta x\right)$. Let us denote the function $g(x) = \frac{x^2}{2} - \beta x$ so that $g'(x) = x - \beta$ and $g''(x) = 1 > 0$. Therefore, at $x = \beta$ the function g attains its infimum. So as long as $\beta \leq \sqrt{2\log 2}$, the $\inf_{0\leq x\leq\sqrt{2\log 2}} \left(\frac{x^2}{2} - \beta x\right)$ is attained at $x = \beta$. Moreover, as g is a decreasing function on $[0, \beta]$, for $\beta > \sqrt{2\log 2}$ the $\inf_{0\leq x\leq\sqrt{2\log 2}} \left(\frac{x^2}{2} - \beta x\right)$ is attained at $x = \sqrt{2\log 2}$. Hence we get the following

Theorem 1.3.3. *If $H_N(\sigma)$ are independent $\mathcal{N}(0,N)$, then almost surely,*

$$\begin{aligned} \lim_N \tfrac{1}{N} \log Z_N(\beta) &= \log 2 + \tfrac{\beta^2}{2} \quad \textit{if } \beta < \sqrt{2\log 2} \\ &= \beta\sqrt{2\log 2} \quad \textit{if } \beta \geq \sqrt{2\log 2}. \end{aligned}$$

As we mentioned at the beginning, this is classical.

1.3.2 Exponential Distribution

Another simple but interesting example is the exponential distribution. Let λ_N be two sided exponential probability with scale parameter $\frac{1}{N}$. That is, having density $\frac{1}{2}Ne^{-N|x|}$, for $-\infty < x < \infty$. Once again, it is obvious that $\lambda_N \Rightarrow 0$ as $N \to \infty$. Now we show the following

Proposition 1.3.4. *The sequence $\{\lambda_N\}$ satisfies LDP with rate function $\mathcal{I} = |x|$ on $\mathbb{R}$.*

Proof. Let $\triangle \subset \mathbb{R}$ be an interval. Let $m = \inf\{|x| : x \in \triangle\}$, $M = \sup\{|x| : x \in \triangle\}$, and $q_N = \lambda_N(\triangle)$. With this notation, we have

$$q_N = \frac{N}{2}\int_\triangle e^{-N|x|}dx < \int_{\sqrt{N}m}^{\infty} e^{-x}dx \geq e^{-Nm},$$

and

$$q_N \geq \int_{Nm}^{NM} e^{-x}dx > \int_{Nm}^{N(m+\delta)} e^{-x}dx > N\delta e^{-N(m+\delta)},$$

for any $0 < \delta < M - m$.

From above two inequalities, we can conclude that for any interval $\triangle$, the limit, $\lim\limits_{N\to\infty} \frac{1}{N} \log \lambda_N(\triangle) = -m$. Once again, Proposition 0.3.3 completes the proof. □

Remark 1.3.2. As in the Gaussian case, here also, note that, $\mathcal{I}(x) = |x|$ is a convex continuous function with compact level sets.

Corollary 1.3.5. *If $\xi_N(\sigma)$ are independent (over σ) two sided exponential variables with scale parameter $\frac{1}{N}$, f is a continuous function on $\mathbb{R}$ and $H_N(\sigma) = Nf(\xi(\sigma))$ then* *almost surely,*

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{|x|\le 2\log 2} (\beta f(x) + |x|) .$$

If we take $f(x) = x$, the Hamiltonian for N-particle system H_N is, of course, two sided exponential random variables with scale parameter 1, that is have density $\frac{1}{2}e^{-|x|}$ on $\mathbb{R}$. In that case, for the limiting free energy, we only need to calculate $\inf_{|x|\le 2\log 2} (|x| + \beta x)$, that is, $\inf_{0\le x\le 2\log 2} x(1-\beta)$. A simple calculation yields the following

Theorem 1.3.6. *If $H_N(\sigma)$ are independent two sided exponential random variables with scale parameter 1, then almost surely,*

$$\begin{aligned} \lim_N \tfrac{1}{N} \log Z_N(\beta) &= \log 2 \qquad \text{if } \beta < 1 \\ &= \beta \log 2 \qquad \text{if } \beta \ge 1. \end{aligned}$$

Remark 1.3.3. Interesting observation in this analysis is that, the random Hamiltonian, for N-particle system being exponential random variables with scale parameter 1, does not depend on the number of particles. Even then, the system shows a phase transition.

1.3.3 Weibull Distribution

A more general class that can be considered are the Weibull distributions. Let λ_N be the probabilities on $\mathbb{R}$ having density

$$\phi_{N,\gamma}(x) = \frac{N}{2}|x|^{\gamma-1}e^{-N\frac{|x|^\gamma}{\gamma}}, \quad -\infty < x < \infty. \tag{1.3.3}$$

This is known as Weibull distribution with shape parameter $\gamma > 0$ and scale parameter $\left(\frac{\gamma}{N}\right)^{\frac{1}{\gamma}}$. Clearly, $\phi_{N,1}$ is the usual two sided exponential density, considered in the previous subsection. We show that,

Proposition 1.3.7. *If λ_N has density $\phi_{N,\gamma}$, then $\{\lambda_N\}$ satisfy LDP with rate function $\mathcal{I} = \frac{|x|^\gamma}{\gamma}$ on $\mathbb{R}$.*

Proof. Let $\triangle \subset \mathbb{R}$ be an interval. Let $m = \inf\{|x| : x \in \triangle\}$, $M = \sup\{|x| : x \in \triangle\}$, and $q_N = \lambda_N(\triangle)$. With this notation, we have

$$q_N = \int_{\triangle} \frac{N}{2}|x|^{\gamma-1}e^{-N\frac{|x|^\gamma}{\gamma}}dx \leq \int_{N\frac{m^\gamma}{\gamma}}^{N\frac{M^\gamma}{\gamma}} e^{-x}dx < \int_{N\frac{m^\gamma}{\gamma}}^{\infty} e^{-x} = e^{-N\frac{m^\gamma}{\gamma}}dx, \tag{1.3.4}$$

and

$$q_N \geq \frac{1}{2}\int_{N\frac{m^\gamma}{\gamma}}^{N\frac{M^\gamma}{\gamma}} e^{-x}dx > \frac{1}{2}\int_{N\frac{m^\gamma}{\gamma}}^{N\frac{(m+\delta)^\gamma}{\gamma}} e^{-x}dx > \frac{\delta}{2}N(m+\theta\delta)^{\gamma-1}e^{-N\frac{(m+\delta)^\gamma}{\gamma}}, \tag{1.3.5}$$

for any $0 < \delta < M - m$ and some θ, $0 < \theta < 1$. Mean value theorem is used here.

The above two inequalities imply that for any interval $\triangle \subset \mathbb{R}$, $\lim_{N\to\infty} \frac{1}{N}\log\lambda_N(\triangle) = -\frac{m^\gamma}{\gamma}$. Thus Proposition 0.3.3 completes the proof. □

Remark 1.3.4. In this case, rate function $\mathcal{I}(x) = \frac{|x|^\gamma}{\gamma}$ is a continuous function with compact level sets. But $\mathcal{I}(x)$ is convex, only when $\gamma \geq 1$. Note that, for $0 < \gamma < 1$, $\mathcal{I}(x)$ is not convex but clearly quasi-convex and hence Theorem 1.2.2 is applicable.

Corollary 1.3.8. *If $\xi_N(\sigma)$ are independent (over σ) having density $\phi_{N,\gamma}$, f is a continuous function on $\mathbb{R}$ and $H_N(\sigma) = Nf(\xi_N(\sigma))$ then almost surely,*

$$\lim_{N\to\infty}\frac{1}{N}\log Z_N(\beta) = \log 2 - \inf_{|x|^\gamma \leq \gamma\log 2}\left(\beta f(x) + \frac{|x|^\gamma}{\gamma}\right).$$

As earlier, if we take, $f(x) = x$, the Hamiltonian for N-particle system H_N is a two sided Weibull distribution with shape parameter γ and scale parameter $\gamma^{\frac{1}{\gamma}}N^{\frac{\gamma-1}{\gamma}}$. In this case the problem of limiting free energy reduces to the calculation of $\inf_{|x|^\gamma\leq\log 2}\left(\frac{|x|^\gamma}{\gamma}+\beta x\right)$, that is, $\inf_{0\leq x\leq(\gamma\log 2)^{\frac{1}{\gamma}}}\left(\frac{x^\gamma}{\gamma}-\beta x\right)$.

For $\gamma > 1$, to calculate the above infimum, we imitate the Gaussian case. Let us denote $g(x) = \frac{x^\gamma}{\gamma} - \beta x$ so that $g'(x) = x^{\gamma-1} - \beta$ and $g''(x) = (\gamma-1)x^{\gamma-2} \geq 0$ on $[0,(\gamma\log 2)^{\frac{1}{\gamma}}]$. So g being twice differentiable convex function, the infimum of g will attain where $g' = 0$. But g' will be 0 on $[0,(\gamma\log 2)^{\frac{1}{\gamma}}]$ only when $\beta \leq (\gamma\log 2)^{\frac{1}{\gamma}}$. In that case, the infimum will occur at $x = \beta^{\frac{1}{\gamma-1}}$. When $\beta > (\gamma\log 2)^{\frac{1}{\gamma}}$, then $g' < 0$ on $[0,(\gamma\log 2)^{\frac{1}{\gamma}}]$ and hence infimum occur at $x = (\gamma\log 2)^{\frac{1}{\gamma}}$.

For $\gamma \leq 1$, the function $g(x) = \frac{x^\gamma}{\gamma} - \beta x = x^\gamma\left(\frac{1}{\gamma} - \beta x^{1-\gamma}\right)$ is a product of two functions. Here x^γ is a positive increasing function on $[0,(\gamma\log 2)^{\frac{1}{\gamma}}]$. On the other hand, $\frac{1}{\gamma} - \beta x^{1-\gamma}$ is a decreasing function taking the value $\frac{1}{\gamma}$ at 0. If this function always remains positive then clearly the minimum of g is 0 attained at $x = 0$. On the other hand, if this function takes negative value in $[0, (\gamma \log 2)^{\frac{1}{\gamma}}]$ then the infimum of g is attained at $(\gamma \log 2)^{\frac{1}{\gamma}}$. This situation occurs only when $\frac{1}{\gamma} - \beta x^{1-\gamma} = 0$ for some x in $[0, (\gamma \log 2)^{\frac{1}{\gamma}}]$. This happens only when $\beta \geq \gamma^{-\frac{1}{\gamma}} (\log 2)^{-\frac{1-\gamma}{\gamma}}$. Hence the infimum of g on $[0, (\gamma \log 2)^{\frac{1}{\gamma}}]$ is attained at $x = 0$ for $\beta < \gamma^{-\frac{1}{\gamma}} (\log 2)^{-\frac{1-\gamma}{\gamma}}$ and at $x = (\gamma \log 2)^{\frac{1}{\gamma}}$ for $\beta \geq \gamma^{-\frac{1}{\gamma}} (\log 2)^{-\frac{1-\gamma}{\gamma}}$.

We can combine the above arguments as

Theorem 1.3.9. *If $\{H_N(\sigma), \sigma \in \Sigma_N\}$ are independent having two sided Weibull distribution with shape parameter $\gamma > 0$ and scale parameter $\gamma^{\frac{1}{\gamma}} N^{\frac{\gamma-1}{\gamma}}$, then almost surely,*

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \begin{cases} \log 2 + \frac{\gamma-1}{\gamma} \beta^{\frac{\gamma}{\gamma-1}} & \text{if } \beta < (\gamma \log 2)^{\frac{1}{\gamma}}, \\ (\gamma \log 2)^{\frac{1}{\gamma}} \beta & \text{if } \beta \geq (\gamma \log 2)^{\frac{1}{\gamma}} \end{cases}$$

if $\gamma > 1$

and

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \begin{cases} \log 2 & \text{if } \beta < \gamma^{-\frac{1}{\gamma}} (\log 2)^{-\frac{1-\gamma}{\gamma}}, \\ (\gamma \log 2)^{\frac{1}{\gamma}} \beta & \text{if } \beta \geq \gamma^{-\frac{1}{\gamma}} (\log 2)^{-\frac{1-\gamma}{\gamma}} \end{cases}$$

if $\gamma \leq 1$.

Remark 1.3.5. It is easy to verify that, if λ_N has density,

$$\phi_{N,\gamma}(x) = \text{Const.} e^{-N\frac{|x|^\gamma}{\gamma}} \quad -\infty < x < \infty,$$

more precisely,

$$\phi_{N,\gamma}(x) = \frac{1}{2\Gamma(\frac{1}{\gamma})} \gamma^{\frac{\gamma-1}{\gamma}} N^{\frac{1}{\gamma}} e^{-N\frac{|x|^\gamma}{\gamma}} \quad -\infty < x < \infty$$

then $\{\lambda_N\}$ satisfies LDP with rate function $\mathcal{I}(x) = \frac{1}{\gamma}|x|^\gamma$. Note that here $\gamma = 2$ is the Gaussian distribution. Hence in REM, if we consider H_N to $\mathcal{N}(0, N)$ or two sided Weibull distribution with shape parameter $\gamma = 2$ and scale parameter $\sqrt{2N}$, they will produce the same limiting free energy. So the limiting free energy of REM is not entirely distribution specific, but it is *'rate-specific'*.

1.4 Compact Distributions

In the previous section we observed that, for the existence and evaluation of free energy, we concentrated our attention on the set $\{x : \mathcal{I}(x) \leq \log 2\}$. That is, the entire support of the random variables are not contributing to the system. To do that, we used a variant of Varadhan's lemma. In general Varadhan's lemma is applicable to the class of bounded continuous functions. In our case the functions used in the previous section are rather unbounded. As suggested by Proposition 0.3.4, if the underlying sequence of probabilities are eventually supported on a compact set, we can overcome this little technicality. Our assumption that the rate function $\mathcal{I}$ is a good rate function will ensure that the sequences of induced probabilities are almost surely eventually supported on a compact set. Since $\mathcal{I}$ is a good rate function $\{\mathcal{I}(x) \leq \alpha\}$ is a compact set for every $\alpha \in \mathbb{R}$. In particular, $\{\mathcal{I}(x) \leq \log 2\}$ is a compact set. For example, if $\{\lambda_N\}$ is a sequence of probabilities satisfying LDP with rate function $\mathcal{I}$ so that $\mathcal{I}(x) \leq \log 2$ for all $x \in \mathbb{R}$, then we may not be able to apply Varadhan's lemma to get the free energy of the system. In particular, if $\mathcal{I}(x) = 0$ for all $x \in \mathbb{R}$, we can not infer anything about the existence of the free energy of the system by large deviation techniques.

To start with, let us note that, if $\mathcal{I}$ takes two value 0 and ∞ so that $\{\mathcal{I}(x) = 0\}$ is a compact set, say C. In this case, in view of Remark 1.2.2, we can apply Theorem 1.2.3 with $f(x) = \beta x$. This will ensure the almost sure existence of the limiting free energy and is equal to $\log 2 - \inf_C \beta x = \log 2 - \beta x_0$, where $x_0 = \inf\{x : x \in C\}$. Note that, if $C \subset [0, \infty)$ then clearly the limiting free energy becomes negative for large β if $0 \notin C$ whereas if $0 \in C$ then it will be just a constant, $\log 2$. So we will not get any phase transition here.

Though we do not have a clear picture when $\mathcal{I}$ is identically 0, we have some partial results. First of all, note that $\mathcal{I}(x) = 0$ *for all* $x \in \mathbb{R}$ *iff* $\lim_{N\to\infty} \frac{1}{N} \log \lambda_N(\triangle) = 0$ *for every open subset* $\triangle$ *of* $\mathbb{R}$. This follows from definition of LDP.

Now let us consider the case, when the Hamiltonian is supported on a compact set. For each N, let λ_N be a compactly supported symmetric probability with density ϕ_N and $\{\xi_N(\sigma) : \sigma \in \Sigma_N\}$ be independent random variables having density ϕ_N. Consider the Hamiltonian

$$H_N(\sigma) = N\xi_N(\sigma).$$

Let $[-\alpha_N, \alpha_N]$ be the support of ϕ_N. Let us assume that $\alpha_N \to \alpha$ as $N \to \infty$. Here we allow the possibility that $\alpha = \infty$. For $s \geq 0$, denote $a_N(s) = P\{\xi_N(\sigma) \geq s\}$. Note that in this setup if $\{\lambda_N\}$ satisfy LDP with rate function $\mathcal{I}(x) = 0$, then $\frac{1}{N} \log a_N(s) \to 0$ as $N \to \infty$ for $0 \leq s < \alpha$.

The following theorem suggests that if the tail probability does not decay exponentially fast over N, then we can not expect any annealed phase transition.

Theorem 1.4.1. *Let* $[-\alpha_N, \alpha_N]$ *be the support of* ξ_N *and for* $s \geq 0$, *denote* $a_N(s) = P\{\xi_N(\sigma) \geq s\}$. *If* $\alpha_N \to \alpha$ *and* $\frac{1}{N}\log a_N(s) \to 0$ *as* $N \to \infty$ *for* $0 \leq s < \alpha$, *then*

$$\lim_{N\to\infty} \frac{1}{N} E \log Z_N(\beta) = \log 2 + \alpha\beta.$$

Proof. As

$$2^N e^{-\beta N \alpha_N} \leq Z_N(\beta) \leq 2^N e^{\beta N \alpha_N},$$

the proof for $\alpha = 0$ is immediate. Moreover, in this case, for every sample point

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \log 2.$$

So let $\alpha > 0$ (may be $\alpha = \infty$). Since log is concave, by Jensen's inequality

$$E \log Z_N(\beta) \leq \log E Z_N(\beta). \tag{1.4.1}$$

As H_N are bounded by $N\alpha_N$,

$$E Z_N(\beta) = 2^N E e^{\beta H_N} < 2^N e^{\beta N \alpha_N}.$$

Hence, by assumption and (1.4.1),

$$\limsup_{N\to\infty} \frac{1}{N} E \log Z_N(\beta) \leq \log 2 + \alpha\beta. \tag{1.4.2}$$

Now we show,

$$\liminf_{N\to\infty} \frac{1}{N} E \log Z_N(\beta) \geq \log 2 + \alpha\beta. \tag{1.4.3}$$

For that, with arbitrary but fixed $0 \leq s < \alpha$, let $X_N = \#\{\sigma : H_N(\sigma) \geq sN\}$. Then

$EX_N = 2^N a_N(s)$ and $EX_N^2 = 2^N(2^N-1)a_N^2(s) + 2^N a_N(s)$. Hence

$$E(X_N - EX_N)^2 = EX_N^2 - (EX_N)^2 \leq 2^N a_N(s). \tag{1.4.4}$$

If $X_N \leq 2^{N-1}a_N(s)$ then $EX_N - X_N \geq 2^{N-1}a_N(s)$ so that $(X_N - EX_N)^2 \geq 2^{2N-2}a_N^2(s)$. Let $A_N = \{X_N \leq 2^{N-1}a_N(s)\}$. So $A_N \subset \{(X_N - EX_N)^2 \geq 2^{2N-2}a_N^2(s)\}$. Hence, by Markov inequality and (1.4.4),

$$P(A_N) \leq \frac{E(X_N - EX_N)^2}{2^{2N-2}a_N^2(s)} \leq \frac{4}{2^N a_N(s)}.$$

i.e., $P(A_N^c) \geq 1 - \frac{4}{2^N a_N(s)}$. But on A_N^c,

$$Z_N(\beta) \geq X_N e^{\beta s N} \geq 2^{N-1}a_N(s)e^{\beta s N},$$

and hence

$$E\left[\log Z_N(\beta) 1_{A_N^c}\right] \geq [(N-1)\log 2 + \log a_N(s) + \beta s N]\left(1 - \frac{4}{2^N a_N(s)}\right). \tag{1.4.5}$$

Now $A_N = \{X_N = 0\} \cup \{1 \leq X_N \leq 2^{N-1}a_n(s)\}$. Since $Z_N(\beta) \geq 2^N e^{-\beta N \alpha_N}$ and $PP(X_N = 0) = (1 - a_N(s))^{2^N}$ we have

$$E\left[\log Z_N(\beta) 1_{\{X_N = 0\}}\right] \geq (N \log 2 - \beta N \alpha_N)(1 - a_N(s))^{2^N}. \tag{1.4.6}$$

On $\{1 \leq X_N \leq 2^{N-1}a_N(s)\}$, $\log Z_N(\beta) \geq \beta \max_\sigma H_N(\sigma) \geq \beta s N > 0$ and hence

$$E\left[\log Z_N(\beta) 1_{\{1 \leq X_N \leq 2^{N-1}a_N(s)\}}\right] \geq 0. \tag{1.4.7}$$

Thus from (1.4.5), (1.4.6) and (1.4.7) we have

$$\begin{aligned}\tfrac{1}{N}E\log Z_N(\beta) \;\; &\geq \;\; \left[\tfrac{N-1}{N}\log 2 + \tfrac{\log a_N(s)}{N} + \beta s\right]\left(1 - \tfrac{4}{2^N a_N(s)}\right) \\ &\quad + \left[\log 2 - \beta\alpha_N\right](1 - a_N(s))^{2^N}.\end{aligned}$$

By assumption, $\frac{1}{N}\log a_N(s) \to 0$ so that $2^N a_N(s) \to \infty$ and hence $(1-a_N(s))^{2^N} \to 0$ as $N \to \infty$. Thus, under the assumption,

$$\liminf_{N\to\infty} \frac{1}{N}E\log Z_N(\beta) \geq \log 2 + \beta s.$$

Since $0 \leq s < \alpha$ is arbitrary, we have

$$\liminf_{N\to\infty} \frac{1}{N}E\log Z_N(\beta) \geq \log 2 + \alpha\beta$$

which remain true even when $\alpha = \infty$ with the understanding that the right side of the above inequality is ∞.

This completes the proof. □

Since ξ_N has density ϕ_N with support $[-\alpha_N, \alpha_N]$ and $H_N = N\xi_N$, the support of H_N will be $[-T_N, T_N]$ where $T_N = N\alpha_N$. If we assume, φ_N be the density of H_N with support $[-T_N, T_N]$, then we can apply the above theorem with $\alpha_N = \frac{T_N}{N}$.

The following examples will illustrate the applications of the above theorem.

Example 1.4.1 (Uniform Distribution). Let $\varphi_N(x) = \frac{1}{2T_N}1_{[-T_N,T_N]}$. If $\frac{T_N}{N} \to \alpha > 0$, then $a_N(s) \to \frac{\alpha - s}{2\alpha} > 0$ for all $0 \leq s < \alpha$. So utilizing the above theorem we get,

a) if $T_N = \sqrt{N}$ then $\lim_{N\to\infty} \dfrac{1}{N}\log Z_N(\beta) = \log 2$ for every sample point,

b) if $T_N = N$ then $\lim_{N\to\infty} \dfrac{1}{N}E\log Z_N(\beta) = \log 2 + \beta$,

c) if $T_N = N^2$ then $\lim_{N\to\infty} \dfrac{1}{N}E\log Z_N(\beta) = \infty$.

Similar remarks follows for the other examples also.

Example 1.4.2. Let $\delta > 0$ be fixed and

$$\varphi_N(x) = \frac{\delta + 1}{2T_N^{\delta+1}}(T_N - |x|)^{\delta}, \qquad -T_N \leq x \leq T_N.$$

If $\frac{T_N}{N} \to \alpha > 0$, then $a_N(s) \to \frac{1}{2}\left(\frac{\alpha - s}{\alpha}\right)^{\delta+1} > 0$ for all $0 \le s < \alpha$.

Example 1.4.3. Let

$$\varphi_N(x) = \frac{1}{2T_N}\cos\frac{x}{T_N}, \qquad -\frac{\pi T_N}{2} \le x \le \frac{\pi T_N}{2}.$$

If $\frac{T_N}{N} \to \alpha > 0$, then $a_N(s) \to \frac{1}{2}\left(1 - \sin\frac{s}{\alpha}\right) > 0$ for all $0 \le s < \alpha\frac{\pi}{2}$.

Example 1.4.4. Let

$$\varphi_N(x) = \frac{N}{2(e^{NT_N} - 1)} e^{N|x|} 1_{[-T_N, T_N]}.$$

If $\frac{T_N}{N} \to \alpha > 0$, then $a_N(s) \to \frac{1}{2}$ as $N \to \infty$ for all $0 \le s < \alpha$.

In the following examples, above theorem is not applicable.

Example 1.4.5 (Truncated Double Exponential). Let

$$\varphi_N(x) = \frac{1}{2(1 - e^{-T_N})} e^{-|x|} 1_{[-T_N, T_N]}.$$

Let $\frac{T_N}{N} \to \alpha(> 0)$ as $N \to \infty$. Then $a_N(s) = \frac{e^{T_N - sN} - 1}{2(e^{T_N} - 1)}$. Hence $\frac{\log a_N(s)}{N} \to -s \ne 0$ as $N \to \infty$ for all s with $0 < s < \alpha$. Thus we can not apply the Theorem 1.4.1 any more. However, if $H_N(\sigma)$ has density $\varphi_N(x)$ and λ_N is the law of $\frac{1}{N}H_N(\sigma)$ then if $\frac{T_N}{N} \to \alpha(> 0)$, by analysis of subsection 1.3.2, we can easily see that the sequence $\{\lambda_N\}$ satisfies large deviation principle with rate function $\mathcal{I}$ given by,

$$\mathcal{I}(x) = \begin{cases} |x| & \text{for } |x| \le \alpha \\ \infty & \text{otherwise.} \end{cases}$$

Hence we can use Theorem 1.2.3 to conclude that the free energy will be same as that of exponential REM as long as $\alpha \ge \log 2$. Where as if $\alpha < \log 2$ then almost surely,

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \begin{cases} \log 2 & \text{for } 0 \le \beta \le 1 \\ \log 2 - \alpha + \beta\alpha & \text{for } \beta \ge 1. \end{cases}$$

Example 1.4.6 (Truncated Gaussian). Let

$$\varphi_N(x) = \frac{1}{C_N} e^{-\frac{x^2}{2N}} 1_{[-T_N, T_N]}.$$

Let $\frac{T_N}{N} \to \alpha(>0)$ as $N \to \infty$. Then $\frac{\log a_N(s)}{N} \to -\frac{1}{2}s^2 \neq 0$ as $N \to \infty$ for all s with $0 < \frac{1}{2}s^2 < \alpha$. Thus we can not apply the Theorem 1.4.1 once again. However, if $H_N(\sigma)$ has density $\varphi_N(x)$ and λ_N is the law of $\frac{1}{N}H_N(\sigma)$ then if $\frac{T_N}{N} \to \alpha(>0)$, by analysis of subsection 1.3.1, we can easily see that the sequence $\{\lambda_N\}$ satisfy large deviation principle with rate function $\mathcal{I}$ given by,

$$\mathcal{I}(x) = \begin{cases} \frac{1}{2}x^2 & \text{for } \frac{1}{2}x^2 \leq \alpha \\ \infty & \text{otherwise.} \end{cases}$$

Hence we can use Theorem 1.2.3 to conclude that the free energy will be same as that of Gaussian REM as long as $\alpha \geq \log 2$. Where as if $\alpha < \log 2$ then almost surely,

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \begin{cases} \log 2 + \frac{1}{2}\beta^2 & \text{for } 0 \leq \beta \leq \sqrt{2\alpha} \\ \log 2 - \alpha + \beta\sqrt{2\alpha} & \text{for } \beta \geq \sqrt{2\alpha}. \end{cases}$$

1.5 Discrete Distributions

We conclude this chapter by considering the REM driven by some discrete distributions.

1.5.1 Poisson Distribution

Let us consider the Hamiltonian for the N particle system $H_N(\sigma) \sim P(N\theta)$ where $P(N\theta)$ is the Poisson distribution with parameter $N\theta$. Let λ_N be the law $\frac{1}{N}P(N\theta)$. We also can think of λ_N as the law of the sample mean for a sample of size N from $P(\theta)$. Then by Cramer's theorem (Theorem 0.3.5), $\{\lambda_N\}$ satisfies LDP with convex good rate function $\mathcal{I}$ given by

$$\mathcal{I}(x) = \begin{cases} \theta - x + x\log\frac{x}{\theta} & \text{for } x \geq 0 \\ \infty & \text{otherwise} \end{cases}. \tag{1.5.1}$$

Hence by Theorem 1.2.3, if for $\sigma \in \Sigma_N$ the random variables $\xi(\sigma)$ is distributed like λ_N and the Hamiltonian $H_N(\sigma) = Nf(\xi(\sigma))$ with any continuous function f on $\mathbb{R}$, we have the following:

Corollary 1.5.1. *If $\lambda_N \sim \frac{1}{N}P(N\theta)$, then almost surely,*

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{\mathcal{I}(x)\leq \log 2} (\beta f(x) + \mathcal{I}(x)).$$

Notation: Note that here $\mathcal{I}$ is a convex continuous function on $[0,\infty)$ so that $\mathcal{I}(0) = \theta$; $\mathcal{I}(\theta) = 0$ and $\mathcal{I}(x) \to \infty$ as $x \to \infty$. So the set $\{x : \mathcal{I}(x) = \log 2\}$ contains only one point when $\theta < \log 2$; contains zero and one non-zero-point for $\theta = \log 2$; contains two positive points for $\theta > \log 2$. As a consequence, the set $\{\mathcal{I}(x) \leq \log 2\}$ is an interval $[x_1, x_2]$; $x_1 = 0$ in case of $\theta \leq \log 2$ where as $x_1 > 0$ in case of $\theta > \log 2$. In any case, $\theta \in (x_1, x_2)$.

Hence when $f(x) = x$ the above corollary implies that

$$\begin{aligned} \lim_N \tfrac{1}{N} \log Z_N(\beta) &= \log 2 - \inf_{[x_1,x_2]} \left\{\beta x + \theta - x + x\log\tfrac{x}{\theta}\right\} \\ &= \log 2 - \theta - \inf_{[x_1,x_2]} \left\{(\beta - 1)x + x\log\tfrac{x}{\theta}\right\}. \end{aligned}$$

To calculate the above infimum, let $g(x) = (\beta-1)x + x\log\frac{x}{\theta}$ on $[0,\infty)$. Clearly, g is a convex function. $g'(x) = \beta + \log\frac{x}{\theta}$, so that $g'(\overline{x}) = 0$ implies $\overline{x} = \theta e^{-\beta}$. Hence g attains its infimum at $\overline{x} = \theta e^{-\beta}$. We consider two cases, namely, $\theta \leq \log 2$ and $\theta > \log 2$.

$\boldsymbol{\theta \leq \log 2}$

For $\beta \geq 0$, $0 < \theta e^{-\beta} \leq \theta < x_2$. That is, the point $\overline{x} = \theta e^{-\beta}$, where g attains minimum, belongs to $\in [x_1, x_2]$ for every $\beta \geq 0$. Hence $\inf_{[x_1,x_2]} g(x) = (\beta-1)\theta e^{-\beta} - \theta e^{-\beta}\log e^{-\beta} = -\theta e^{-\beta}$

$\boldsymbol{\theta > \log 2}$

As β increases from 0 to ∞, $\theta e^{-\beta}$ decreases from θ to 0. Since $0 < x_1 < \theta$, there exists $\beta_0 > 0$ such that

$$\theta e^{-\beta_0} = x_1.$$

Clearly, for $\beta \leq \beta_0$, $\theta e^{-\beta} \in [x_1, x_2]$ so that $\inf_{[x_1,x_2]} g(x) = -\theta e^{-\beta}$. Since g attains its infimum at $\overline{x} = \theta e^{-\beta}$, g is increasing (by convexity) on $(\overline{x}, \infty)$. For $\beta > \beta_0$, $\overline{x} = \theta e^{-\beta} < \theta e^{-\beta_0} = x_1$. Thus g is increasing on $[x_1, x_2]$. As a consequence, when $\beta > \beta_0$, we have $\inf\limits_{[x_1,x_2]} g(x) = g(x_1) = \beta x_1 + \mathcal{I}(x_1) - \theta = \beta x_1 + \log 2 - \theta$.

All this leads to

Theorem 1.5.2. *Consider REM where the Hamiltonian $H_N(\sigma)$ is Poisson with parameter $N\theta$.*

a) For $\theta \leq \log 2$, almost surely,

$$\lim \frac{1}{N} \log Z_N = \log 2 - \theta + \theta e^{-\beta} \quad \textit{for } \beta \geq 0.$$

b) For $\theta > \log 2$; let x_1 be the least positive solution of $x(\log \frac{x}{\theta} - 1) = \theta - \log 2$, and $\beta_0 = \log \frac{\theta}{x_1} = \frac{\theta - \log 2}{x_1} - 1$. Then almost surely,

$$\begin{aligned} \lim \tfrac{1}{N} \log Z_N &= \log 2 - \theta + \theta e^{-\beta} && \textit{for } \beta \leq \beta_0 \\ &= \beta x_1 && \textit{for } \beta > \beta_0. \end{aligned}$$

Now if we take $f(x) = -x$, then by Corollary 1.5.1, almost surely, the limiting free energy is given by

$$\begin{aligned} \lim_N \frac{1}{N} \log Z_N(\beta) &= \log 2 - \inf_{\mathcal{I}(x) \leq \log 2} \{\mathcal{I}(x) - \beta x\} \\ &= \log 2 - \theta - \inf_{\mathcal{I}(x) \leq \log 2} \left\{ x \log \frac{x}{\theta} - (\beta + 1)x \right\}. \end{aligned}$$

To calculate the above infimum, let $g(x) = x \log \frac{x}{\theta} - (\beta + 1)x$ on $[0, \infty)$ so that $g'(x) = \log \frac{x}{\theta} - \beta$ and $g''(x) = \frac{1}{x} > 0$ on $(0, \infty)$. Hence g attains its infimum at $\underline{x} = \theta e^{\beta}$. Note that, $\underline{x} = \theta$ for $\beta = 0$ and $\underline{x} \to \infty$ as $\beta \to \infty$. So there exist $\beta_1 > 0$ such that $\mathcal{I}(\theta e^{\beta_1}) = \log 2$, that is, $\theta e^{\beta_1} = x_2$. So the infimum $\inf\limits_{\mathcal{I}(x) \leq \log 2} \left\{ x \log \frac{x}{\theta} - (\beta + 1)x \right\}$ occurs at θe^{β} for $\beta \leq \beta_1$ and at x_2 for $\theta > \beta_1$ leading to the following

Theorem 1.5.3. *In REM, if the Hamiltonian $H_N(\sigma)$ negative of Poisson with parameter $N\theta$, then almost surely*

$$\begin{aligned} \lim \tfrac{1}{N} \log Z_N &= \log 2 - \theta + \theta e^{\beta} && \textit{for } \beta \leq \beta_1 \\ &= \beta x_2 && \textit{for } \beta > \beta_1. \end{aligned}$$

1.5.2 Binomial Distribution

Let $X_N \sim B(N,p)$ where $B(N,p)$ is the Binomial distribution with parameter $p\,(0 < p < 1)$. Put $\xi_N = \frac{X_N}{N}$. Observe that when $p = 0$ or 1, then the Hamiltonian is deterministic one and uninteresting. Let λ_N be the law of ξ_N. Thus ξ_N is nothing but the proportion of heads in N tosses of a coin (with chance of heads p). By Cramer's theorem (Theorem 0.3.5), $\{\lambda_N\}$ satisfies LDP with convex good rate function $\mathcal{I}$ given by

$$\mathcal{I}(x) = \begin{cases} x\log\frac{x}{p} + (1-x)\log\frac{1-x}{1-p} & \text{for } x \in [0,1] \\ \infty & \text{otherwise.} \end{cases} \tag{1.5.2}$$

Note that here $\mathcal{I}$ is a strictly convex continuous function. Now fix a continuous function f on $\mathbb{R}$. Consider N particle system with Hamiltonian $H_N = Nf(\xi_N(\sigma))$. By Theorem 1.2.3, to calculate the limiting free energy we only have to solve the optimization problem

$$\inf_{\mathcal{I}(x)\le\log 2} \left(\beta f(x) + \mathcal{I}(x)\right).$$

Note that here $\mathcal{I}(0) = -\log(1-p)$; $\mathcal{I}(p) = 0$ and $\mathcal{I}(1) = -\log p$. So the set $\{x : \mathcal{I}(x) = \log 2\} = \{0,1\}$ when $p = \frac{1}{2}$ otherwise the set $\{x : \mathcal{I}(x) = \log 2\}$ is a singleton. Let us denote the set $\{\mathcal{I}(x) \le \log 2\}$ as $[x_1,x_2]$ where $0 < x_1 < p < x_2 = 1$ for $p > \frac{1}{2}$; $0 = x_1 < p < x_2 < 1$ for $p < \frac{1}{2}$ and $0 = x_1 < p < x_2 = 1$ for $p = \frac{1}{2}$.

When $f(x) = x$, by Theorem 1.2.3, we have almost surely,

$$\lim_N \frac{1}{N}\log Z_N(\beta) = \log 2 - \inf_{[x_1,x_2]}\left\{\beta x + x\log\frac{x}{p} + (1-x)\log\frac{1-x}{1-p}\right\}.$$

To calculate the above infimum, let $g(x) = \beta x + x\log\frac{x}{p} + (1-x)\log\frac{1-x}{1-p}$ on $[0,1]$. Clearly g is a convex function and $g'(x) = \beta + \log\frac{x(1-p)}{(1-x)p}$. So g attains its infimum at $\overline{x}(\beta)$ given by $\overline{x}(\beta) = \dfrac{p}{p+(1+p)e^{\beta}}$. We now consider two cases.

$p \le \frac{1}{2}$

As $[x_1,x_2] = [0,x_2]$ where $x_2 > p$ and $\overline{x}(\beta) \le \frac{p}{2p+1} < p$ for every $\beta \ge 0$, we have $\overline{x}(\beta) \in [x_1,x_2] = [0,x_2]$. Hence on $[x_1,x_2]$, g attains its infimum at $\overline{x}(\beta)$ and by routine algebraic manipulations, we get, $\inf_{[x_1,x_2]} g(x) = \beta\overline{x}(\beta) + \mathcal{I}(\overline{x}(\beta)) = \beta - \log(p+(1-p)e^{\beta})$.

$p > \frac{1}{2}$

Since $\overline{x}(\beta)$ decreases from $\frac{p}{2p+1} < p$ to 0 as β increases from 0 to ∞ and $0 < x_1 < p$, there exists $\beta_0 > 0$ such that

$$\overline{x}(\beta_0) = x_1.$$

Hence as $\overline{x}(\beta) \in [x_1,x_2]$ for $\beta \le \beta_0$, g attains its infimum at $\overline{x}(\beta)$ on $[x_1,x_2]$ and we get $\inf_{[x_1,x_2]} g(x) = \beta - \log(p+(1-p)e^{\beta})$. On the other hand, for $\beta > \beta_0$, g being a convex function and attains it infimum at $\overline{x} < x_1$, g is increasing for $x > \overline{x}$. Hence g attains its infimum on $[x_1, x_2]$ at x_1 leading to $\inf_{[x_1,x_2]} g(x) = g(x_1) = \beta x_1 + \mathcal{I}(x_1) = \beta x_1 + \log 2$.

All this leads to

Theorem 1.5.4. *In REM, if the Hamiltonian $H_N(\sigma)$ is Binomial with parameter N and p, then almost surely,*

a) for $p \leq \frac{1}{2}$,

$$\lim \frac{1}{N} \log Z_N = \log 2 - \beta + \log(p + (1-p)e^{\beta}) \quad \textit{for } \beta \leq 0.$$

b)for $p > \frac{1}{2}$,

$$\begin{aligned} \lim \tfrac{1}{N} \log Z_N &= \log 2 - \beta + \log(p + (1-p)e^{\beta}) && \textit{for } \beta \leq \beta_0 \\ &= -\beta x_1 && \textit{for } \beta > \beta_0. \end{aligned}$$

On the other hand, if we take $f(x) = -x$, then by Theorem 1.2.3, almost surely,

$$\begin{aligned} \lim_N \frac{1}{N} \log Z_N(\beta) &= \log 2 - \inf_{\mathcal{I}(x) \leq \log 2} \{\mathcal{I}(x) - \beta x\} \\ &= \log 2 - \inf_{[x_1,x_2]} \left\{ x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p} - \beta x \right\}, \end{aligned}$$

where we use the same notation for x_1, x_2 as in the case for $f(x) = x$.

To calculate the above infimum, let $h(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p} - \beta x$ on $[0, 1]$ so that $h'(x) = \log \frac{x(1-p)}{(1-x)p} - \beta$ and $g''(x) = \frac{1}{x(1-x)} > 0$ on $(0, 1)$. Hence g attains its infimum at $\underline{x}(\beta) = \dfrac{p}{p + (1-p)e^{-\beta}}$. Note that, $\underline{x}(0) = p$ and $\underline{x}(\beta) \to 1$ as $\beta \to \infty$. So the infimum $\inf_{[x_1,x_2]} h(x)$ is attained at $\underline{x}(\beta)$ for every $\beta > 0$ for $p \geq \frac{1}{2}$ and for $\beta \leq \beta_1$ for $p < \frac{1}{2}$, where $\beta_1 > 0$ is such that $\mathcal{I}(\underline{x}(\beta_1)) = \log 2$, that is, $\underline{x}(\beta_1) = x_2$. For $p < \frac{1}{2}$ and $\beta > \beta_1$ the infimum $\inf_{[x_1,x_2]} h(x)$ is attained at at x_2. To be more precise, we get the following:

Theorem 1.5.5. *In REM, if the Hamiltonian $H_N(\sigma)$ is negative Binomial random variable with parameter N and p, then almost surely* *a) for* $p \geq \frac{1}{2}$,

$$\lim \frac{1}{N} \log Z_N = \log 2 + \beta + \log(p + (1-p)e^{-\beta}) \quad \textit{for } \beta \geq 0.$$

b)for $p < \frac{1}{2}$,

$$\begin{aligned} \lim \tfrac{1}{N} \log Z_N \quad &= \log 2 + \beta + \log(p + (1-p)e^{-\beta}) \quad && \textit{for } \beta \leq \beta_1 \\ &= \beta x_2 && \textit{for } \beta > \beta_1. \end{aligned}$$

Chapter 2

The Generalized Random Energy Model

In the random energy model (REM) [15, 16] of Derrida, the Hamiltonians in distinct configurations are independent. The idea in generalized random energy model (GREM) is to bring an amount of dependence in the Hamiltonians through the structure of configurations. Of course, very little can be achieved by assuming an arbitrary covariance matrix. To introduce hierarchy, an n-level tree structure was suggested by Derrida [17], where the branches of the tree are in correspondence with the configuration space. In this chapter we discuss this model with some modifications. There are two essential differences from what is usually considered in the literature. First, we provide a general framework of trees. However, they will be considered in the next chapter. Second, we split the number of particles N into n groups rather than splitting the number of spins (or 'factorizing' 2 as is customary in the literature). This allows us to introduce a further randomization at the tree level, like Poisson trees and multinomial trees. These will be consider in chapter 3.

In this chapter, we specialize to the driving distributions having exponentially decaying tails. The basic inequalities lead to the large deviation principle (LDP) for the random probabilities as in the case of REM considered in the previous chapter. This leads to an explicit formula for the free energy. For the exponential GREM, the driving distribution does not depend on the number of particles. This does not make it less interesting. In fact, the Gaussian case is no more complicated than the exponential case. The present treatment clearly brings out the similarities between the two cases. There are dissimilarities too. As expected, for small values of β (inverse temperature), the energy function in the exponential case does not depend on β where as for the Gaussian it is quadratic in β. In the Gaussian case, all the weights associated with all the levels of the tree participate in the expression for free energy, where as in the exponential case it is not always so.

Even though for any finite number of particles, we have a truly n level tree, in the limit, it may collapse to a lower level tree – it may even correspond to REM (see remarks 2.6.1 and 2.6.5). This leads to the notion of reduced GREM. For such models, the energy function determines all the parameters of the model. It is also possible to characterize the energy functions. It is interesting to note that in the SK-model, subject to certain moment conditions of the underlying distribution, the energy function is universal [9], while it is not true here.

2.1 Derrida's Model

Let us first describe the model in detail. As a generalization to his REM [16], in GREM [17] Derrida introduced a tree-like structure in the energy levels. This is what we now explain. Fix a positive integer $n \geq 1$. This n will be the level of the tree. For each level $i = 1, 2, \cdots, n$ of the tree, fix number α_i so that $\alpha_i \in (1, 2)$ and $\prod_{i=1}^n \alpha_i = 2$. For fixed N, in the tree, there will be α_1^N many nodes at the first level. Below each of the first level nodes, there will be α_2^N many nodes in the 2nd level. Hence, there will be a total of $(\alpha_1\alpha_2)^N$ many nodes at the 2nd level. In general, at the ith level, there will be α_i^N many nodes below each of the $(i-1)$th level nodes giving a total $(\alpha_1\alpha_2\cdots\alpha_i)^N$ many nodes in the ith level. So at the n-th, that is, last level there will be $(\alpha_1\alpha_2\cdots\alpha_n)^N = 2^N$ many nodes (leaves). Derrida associates the configuration space Σ_N with the all possible branches from root to leaves of the above tree. Since there are 2^N many configurations, he assumes $\prod_{i=1}^n \alpha_i = 2$. To define the Hamiltonian, he associates an independent random variable to each edge of the tree. For $i = 1, 2, \cdots, n$ there are $(\alpha_1\alpha_2\cdots\alpha_i)^N$ independent Gaussian mean zero random variables $\xi_j^{(i)}$ with variance $a_i N$ associated to each of the ith level edges. Here $a_1, a_2, \cdots, a_n$ are positive numbers so that $\sum_{i=1}^n a_i = 1$. The Hamiltonian for a configuration, that is, for a branch from root to a leaf is the sum of the n random variables associated with the edges constituting the branch. So the partition function, in this model, reduces to

$$Z_N(\beta) = \sum_{i_1=1}^{\alpha_1^N} \sum_{i_2=(i_1-1)\alpha_2^N+1}^{i_1\alpha_2^N} \cdots \sum_{i_2=(i_{n-1}-1)\alpha_n^N+1}^{i_{n-1}\alpha_n^N} e^{-\beta\left(\sum_{k=1}^n \xi_{i_k}^{(k)}\right)}.$$

In the entire explanation above, we pretended that each α_i^N is an integer. But is this possible? – No. One way out is to consider $[\alpha_i^N]$. Being the number of edges, each α_i^N has to be an integer which divides 2^N because $(\alpha_1\alpha_2\cdots\alpha_n)^N = 2^N$. By the fundamental theorem of arithmetics, $\alpha_i^N = 2^{k(i,N)}$ for some positive integer $k(i, N)$. Moreover, $k(i, N)$ for $i = 1, \cdots, n$ is such that $k(1, N)+k(2, N)+\cdots+k(n, N) = N$. In other words: given any tree with 2^N leaves the construction allows only for furcations in powers of 2 at each layer. This was also noted in [13]. To eliminate the confusion regarding whether α_i^N is an integer or not, we made the natural modification to the model in the next section.

2.2 A Reformulation

We formulate GREM as follows. As above, fix an integer $n \geq 1$. Let $N \geq n$ be the number of particles, each of which can have two states/spins $+1, -1$; so that the configuration space is 2^N. Consider a partition of N, into integers $k(i, N)$ for $1 \leq i \leq n$ with each $k(i, N) \geq 1$ and $\sum_i k(i, N) = N$. The configuration space 2^N, naturally splits into product, $\prod 2^{k(i,N)}$ and $\sigma \in 2^N$ can be written as $\sigma_1 \sigma_2 \cdots \sigma_n$ with $\sigma_i \in 2^{k(i,N)}$. An obvious tree structure can be brought in the configuration space. As earlier imagine an n-level tree. There are $2^{k(1,N)}$ nodes at the first level. These will be denoted as σ_1, with $\sigma_1 \in 2^{k(1,N)}$. Below each of the first level nodes there are $2^{k(2,N)}$ nodes at the second level. The second level nodes below σ_1 of the first level will be denoted by $\sigma_1 \sigma_2$ with $\sigma_2 \in 2^{k(2,N)}$. In general, below a node $\sigma_1 \sigma_2 \cdots \sigma_{i-1}$ of the $(i-1)$-th level there are $2^{k(i,N)}$ nodes at the i-th level denoted by $\sigma_1 \sigma_2 \cdots \sigma_{i-1} \sigma_i$ for $\sigma_i \in 2^{k(i,N)}$. Thus a typical branch of the tree reads like $\sigma_1 \sigma_2 \cdots \sigma_n$. Obviously the branches are in one one correspondence with 2^N, the configuration space. At the node $\sigma_1 \cdots \sigma_i$, we place a random variables $\xi(\sigma_1 \cdots \sigma_i)$. We assume that all these random variables are i.i.d. with a symmetric distribution. We associate one weight for each level, say weight $a_i > 0$ for the i-th level. These are not random. In a configuration $\sigma = \sigma_1 \cdots \sigma_n$ the Hamiltonian is

$$H_N(\sigma) = \sum_{i=1}^{n} a_i \xi(\sigma_1 \cdots \sigma_i). \tag{2.2.1}$$

For $\beta > 0$ the partition function is

$$Z_N(\beta) = \sum_{\sigma} e^{-\beta H_N(\sigma)} = 2^N \mathbf{E}_\sigma e^{-\beta H_N(\sigma)}. \tag{2.2.2}$$

Here $\mathbf{E}_\sigma$ stands for expectation with respect to σ when the configuration space 2^N has uniform distribution. In other words, $\mathbf{E}_\sigma$ is simply the usual average over σ.

Since ξ's are random variables both H_N and Z_N are random variables. We suppress the parameter ω. As usual $\frac{1}{N} \log Z_N(\beta)$ is the free energy of the N-particle system. This is the object of study. As N changes, the common distribution of the ξ's would in general change and so in H_N.

2.3 Tree Formulation

We now reformulate the setup as a general tree structure. Though most of the trees that we consider later are *regular* –– in the sense that the number of furcations of a node depend only on its level, and not on the particular node –– the present formulation is general. It allows randomization of the tree, which we do consider later in the next chapter.

Let $n \geq 1$ be fixed integer as earlier. For each $N \geq n$, let T_N be a tree of height n with each branch extending up to the n-th level. σ_1 denotes a typical node at the first level and in general below a node $\sigma_1\sigma_2\cdots\sigma_{i-1}$ of the $(i-1)$-th level, $\sigma_1\sigma_2\cdots\sigma_{i-1}\sigma_i$ is a typical node at the i-th level. We shall now define some useful quantities associated with the tree. Let Σ_N be the set of all branches $\sigma_1\sigma_2\cdots\sigma_n$ of the tree T_N. Let B_{iN} denote the number of nodes at the i-th level. In particular, B_{nN} is the total number of branches of the tree, which will simply be denoted by B_N. For a node $\sigma_1\sigma_2\cdots\sigma_i$ of the i-th level, let $e(\sigma_1\sigma_2\cdots\sigma_i)$ denote the number of nodes at the n-th level below the node $\sigma_1\sigma_2\cdots\sigma_i$. Equivalently, $e(\sigma_1\sigma_2\cdots\sigma_i)$ is the total number of branches extending $\sigma_1\sigma_2\cdots\sigma_i$. Clearly, $\sum\limits_{\sigma_1,\cdots,\sigma_i} e(\sigma_1\cdots\sigma_i) = B_N$ for any i. Let $s^2_{iN} = \sum\limits_{\sigma_1,\cdots,\sigma_i} e^2(\sigma_1\cdots\sigma_i)$.

Assume that $\xi(\sigma_1\cdots\sigma_i)$ is a symmetric random variable associated with node $\sigma_1\sigma_2\cdots\sigma_i$. We assume that these random variables are i.i.d. Strictly speaking we should be using superscript N for the nodes, random variables etc. But for ease in reading we suppress the superscript. This should be borne in mind. We do assume that all our random variables are defined on one probability space. Consider the map $\Sigma_N \to \mathbb{R}^n$ defined by

$$\sigma \mapsto \xi_\sigma = (\xi(\sigma_1), \xi(\sigma_1\sigma_2), \cdots, \xi(\sigma_1\cdots\sigma_n)).$$

Let μ_N be the induced probability on $\mathbb{R}^n$ when Σ_N has uniform distribution, that is, each $\sigma \in \Sigma_N$ has probability $\frac{1}{B_N}$. In other words, for any Borel set $A \subset \mathbb{R}^n$,

$$\mu_N(A) = \frac{1}{B_N}\#\{\sigma : \xi_\sigma \in A\}.$$

In particular, if A is a box, say $\triangle = \triangle_1 \times \cdots \times \triangle_n$, with each $\triangle_i \subseteq \mathbb{R}$, then

$$\mu_N(\triangle) = \frac{1}{B_N} \sum_{<\sigma_1\cdots\sigma_n>} \prod_{i=1}^{n} \mathbf{1}_{\triangle_i}(\xi(\sigma_1\sigma_2\cdots\sigma_i)).$$

Denote $q_{iN} = P(\xi \in \triangle_i)$ for $1 \leq i \leq n$. Since all the $\xi(\sigma_1\cdots\sigma_i)$ (for fixed N) are i.i.d., we did not use suffix for ξ in defining q_{iN}. However since the common distribution will in general change with N, q_{iN} would in general depend on N. Then

$$\mathbf{E}\mu_N(\triangle) = q_{1N}q_{2N}\cdots q_{nN}. \tag{2.3.1}$$

Here now is the basic result.

Theorem 2.3.1. *Let $\triangle = \triangle_1 \times \cdots \times \triangle_n \subset \mathbb{R}^n$. Denote $q_{iN} = P(\xi \in \triangle_i)$ for $1 \leq i \leq n$.*

a) If $\sum_{N\geq n} B_{iN}q_{1N}\cdots q_{iN} < \infty$*, for some* $i, 1 \leq i \leq n$ *then a.s. eventually,*

$$\mu_N(\triangle) = 0.$$

b) If for all $i = 1, \cdots, n$, $\sum_{N\geq n} \frac{s_{iN}^2}{B_N^2 q_{1N}\cdots q_{iN}} < \infty$*, then for any* $\epsilon > 0$ *a.s. eventually,*

$$(1-\epsilon)\mathbf{E}\mu_N(\triangle) \leq \mu_N(\triangle) \leq (1+\epsilon)\mathbf{E}\mu_N(\triangle).$$

In proving the first part of the theorem we will use the idea of Dorlas and Dukes [19], where as for the last part, we follow Capocaccia *et al* [8].

Proof. a) Let j_0 be such that $\sum_{N\geq 1} B_{j_0 N}q_{1N}\cdots q_{j_0 N} < \infty$. Then

$$\begin{aligned}
\mu_N(\triangle) &= \frac{1}{B_N}\sum_{\sigma_1\cdots\sigma_n}\prod_{i=1}^{n}\mathbf{1}_{\triangle_i}(\xi(\sigma_1\cdots\sigma_i)) \\
&\leq \frac{1}{B_N}\sum_{\sigma_1\cdots\sigma_{j_0}}\prod_{i=1}^{j_0}\mathbf{1}_{\triangle_i}(\xi(\sigma_1\cdots\sigma_i))e(\sigma_1\cdots\sigma_{j_0}) \\
&= G_N, \text{(say)}.
\end{aligned}$$

Let A_N be the event $\{G_N = 0\}$. Observe that

$$A_N^c = \left\{\sum_{\sigma_1\cdots\sigma_{j_0}}\prod_{i=1}^{j_0}\mathbf{1}_{\triangle_i}(\xi(\sigma_1\cdots\sigma_i)) \geq 1\right\}.$$

Now by Chebyshev's inequality,

$$\mathbf{P}(A_N^c) \leq \mathbf{E}\sum_{\sigma_1\cdots\sigma_{j_0}}\prod_{i=1}^{j_0}\mathbf{1}_{\triangle_i}(\xi(\sigma_1\cdots\sigma_i)) = B_{j_0 N}q_{1N}\cdots q_{j_0 N}.$$

Thus by assumption and Borel-Cantelli, A_N will occur a.s. eventually. i.e. $G_N = 0$ and hence $\mu_N(\triangle) = 0$.

b) We first get an estimate for the variance of $\mu_N(\triangle)$.

$$\begin{aligned}
& var(\mu_N(\triangle)) \\
= \ & \mathbf{E}(\mu_N(\triangle))^2 - (\mathbf{E}\mu_N(\triangle))^2 \\
= \ & \frac{1}{B_N^2} \sum_{\substack{\sigma_1\cdots\sigma_n \\ \tau_1\cdots\tau_n}} \left[\mathbf{E} \prod_{i=1}^{n} \mathbf{1}_{\triangle_i}(\xi(\sigma_1\cdots\sigma_i))\mathbf{1}_{\triangle_i}(\xi(\tau_1\cdots\tau_i)) - q_{1N}^2 \cdots q_{nN}^2 \right] \\
\leq \ & \frac{1}{B_N^2} \sum_{j=1}^{n} \sum_{\sigma_1\cdots\sigma_j} \sum_{\substack{\sigma_{j+1}\cdots\sigma_n \\ \tau_{j+1}\cdots\tau_n \\ \sigma_{j+1}\neq\tau_{j+1}}} \mathbf{E} \prod_{i=1}^{j} \mathbf{1}_{\triangle_i}(\xi(\sigma_1\cdots\sigma_i)) \prod_{i=j+1}^{n} \mathbf{1}_{\triangle_i}(\xi(\sigma_1\cdots\sigma_i))\mathbf{1}_{\triangle_i}(\xi(\tau_1\cdots\tau_i)) \\
\leq \ & \frac{1}{B_N^2} \sum_{j=1}^{n} q_{1N}\cdots q_{jN} q_{(j+1)N}^2 \cdots q_{nN}^2 \sum_{\sigma_1\cdots\sigma_j} e^2(\sigma_1\cdots\sigma_j) \\
= \ & \frac{1}{B_N^2} \sum_{j=1}^{n} q_{1N}\cdots q_{jN} q_{(j+1)N}^2 \cdots q_{nN}^2 s_{jN}^2
\end{aligned}$$

Hence for any $\epsilon > 0$, by Chebyshev's inequality and (2.3.1)

$$\mathbf{P}(|\mu_N(\triangle) - \mathbf{E}\mu_N(\triangle)| > \epsilon \mathbf{E}\mu_N(\triangle)) < \frac{1}{\epsilon^2 B_N^2} \sum_{j=1}^{n} \frac{s_{jN}^2}{q_{1N}\cdots q_{jN}}.$$

But, in view of the assumption, the sum over N of the right side is finite. So by Borel-Cantelli lemma, a.s. eventually,

$$(1-\epsilon)\mathbf{E}\mu_N(\triangle) \leq \mu_N(\triangle) \leq (1+\epsilon)\mathbf{E}\mu_N(\triangle).$$

□

For GREM type regular trees the condition above will simplify as follows. This result is in [19] though not explicitly stated.

Corollary 2.3.2. *Let $k(i,N)$, $1 \leq i \leq n$ be positive integers with $\sum_i k(i,N) = N$. Suppose that the tree has $2^{k(i,N)}$ nodes of the i-th level below each node of the (i-1)-th level.*

a) If $\sum_{N\geq n} 2^{k(1,N)+\cdots+k(i,N)} q_{1N}\cdots q_{iN} < \infty$, for **some** *$i, 1 \leq i \leq n$, then a.s. eventually, $\mu_N(\triangle) = 0$.*

b) If $\sum_{N\geq n} 2^{-(k(1,N)+\cdots+k(i,N))} q_{1N}^{-1}\cdots q_{iN}^{-1} < \infty$, for **each** *$i = 1,\cdots,n$, then for any*

$\epsilon > 0$, *a.s. eventually,*

$$(1-\epsilon)q_{1N}\cdots q_{nN} \le \mu_N(\triangle) \le (1+\epsilon)q_{1N}\cdots q_{nN}.$$

2.4 Exponentially Decaying Driving Distributions

We fix a number $\gamma > 0$. In this section we consider an n level GREM where for the N particle system the random variables $\xi(\sigma_1\cdots\sigma_i)$ are i.i.d. having probability density

$$\phi_{N,\gamma}(x) = \text{Const.}e^{-\frac{|x|^\gamma}{\gamma N^{\gamma-1}}} \qquad -\infty < x < \infty,$$

More precisely,

$$\phi_{N,\gamma}(x) = \frac{1}{2\Gamma(\frac{1}{\gamma})}\left(\frac{\gamma}{N}\right)^{\frac{\gamma-1}{\gamma}} e^{-\frac{|x|^\gamma}{\gamma N^{\gamma-1}}} \qquad -\infty < x < \infty. \tag{2.4.1}$$

Note that $\phi_{N,1}$ is independent of N and is two sided exponential density with parameter 1. On the other hand, $\phi_{N,2}$ is Gaussian density with mean 0 and variance N. Of course, γ can be larger than 2 as well.

If we define the map $\Sigma_N = \prod_i 2^{k(i,N)} \to \mathbb{R}^n$ by

$$\sigma \mapsto \left(\frac{\xi(\sigma_1,\omega)}{N}, \frac{\xi(\sigma_1\sigma_2,\omega)}{N}, \cdots, \frac{\xi(\sigma_1\cdots\sigma_n,\omega)}{N}\right)$$

and transport the uniform probability of Σ to $\mathbb{R}^n$, we get a probability $\mu_N(\omega)$ on $\mathbb{R}^n$. In evaluating the free energy, we will be applying Varadhan's lemma (Proposition 0.3.4). This explains the factor $\frac{1}{N}$ in the above map, which was not present in the general framework of Theorem 2.3.1.

Let $\triangle = \triangle_1 \times \cdots \times \triangle_n$ be a non-empty open rectangle of $\mathbb{R}^n$. For such $\triangle$ and $1 \le i \le n$ define $m_i = \inf\limits_{x\in\triangle_i} |x|$ and $M_i = \sup\limits_{x\in\triangle_i} |x|$. Clearly, $m_i < \infty$ for all i. Observe that in case $m_i > 0$ then $\triangle_i \subseteq (-M_i, -m_i) \cup (m_i, M_i)$ and in case $m_i = 0$ then $\triangle_i \subseteq (-M_i, M_i)$. In any case $\triangle_i \subseteq (-M_i, -m_i] \cup [m_i, M_i)$ for each i. Let $\tilde{m} = (m_1, \cdots, m_n)$. Also define $q_{iN} = P(\frac{\xi}{N} \in \triangle_i)$, for $1 \le i \le n$.

First let us assume that $\gamma \ge 1$. Let $J \subset \mathbb{R}$ be an interval. Denote $m = \inf\limits_{x\in J} |x|$ and $M = \sup\limits_{x\in J} |x|$. Denote $q_N = P(\frac{\xi}{N} \in J)$. With these notations, we have the following two observations:

$$
\begin{aligned}
q_N = P\left(\frac{\xi}{N} \in J\right) &\leq \frac{1}{\Gamma(\frac{1}{\gamma})}\left(\frac{\gamma}{N}\right)^{\frac{\gamma-1}{\gamma}} \int_{Nm}^{NM} e^{-\frac{x^\gamma}{\gamma N^{\gamma-1}}} dx \\
&< \frac{1}{\Gamma(\frac{1}{\gamma})} \int_{N\frac{m^\gamma}{\gamma}}^{\infty} x^{-\frac{\gamma-1}{\gamma}} e^{-x} dx \\
&\leq \frac{\gamma^{\frac{\gamma-1}{\gamma}}}{\Gamma(\frac{1}{\gamma})(Nm^\gamma)^{\frac{\gamma-1}{\gamma}}} \int_{N\frac{m^\gamma}{\gamma}}^{\infty} e^{-x} dx \\
&= \frac{\gamma^{\frac{\gamma-1}{\gamma}}}{\Gamma(\frac{1}{\gamma})(Nm^\gamma)^{\frac{\gamma-1}{\gamma}}} e^{-N\frac{m^\gamma}{\gamma}},
\end{aligned} \tag{2.4.2}
$$

with the understanding that when $m = 0$, the last expression is 1 and

$$
\begin{aligned}
q_N = P\left(\frac{\xi}{N} \in J\right) &\geq \frac{1}{2\Gamma(\frac{1}{\gamma})}\left(\frac{\gamma}{N}\right)^{\frac{\gamma-1}{\gamma}} \int_{Nm}^{NM} e^{-\frac{x^\gamma}{\gamma N^{\gamma-1}}} dx \\
&> \frac{1}{2\Gamma(\frac{1}{\gamma})} \int_{N\frac{m^\gamma}{\gamma}}^{N\frac{m^\gamma}{\gamma}+\delta} x^{-\frac{\gamma-1}{\gamma}} e^{-x} dx \\
&> \frac{\delta}{2\Gamma(\frac{1}{\gamma})(N\frac{m^\gamma}{\gamma}+\delta)^{\frac{\gamma-1}{\gamma}}} e^{-(N\frac{m^\gamma}{\gamma}+\delta)},
\end{aligned} \tag{2.4.3}
$$

for any $0 < \delta < \frac{1}{\gamma}(M^\gamma - m^\gamma)$.

Now let $\gamma < 1$. J, m, M as above except that J is now assumed to be bounded interval of $\mathbb{R}$ so that $0 \leq m < M < \infty$. With q_N as earlier, we have

$$
\begin{aligned}
q_N = P\left(\frac{\xi}{N} \in J\right) &\leq \frac{1}{\Gamma(\frac{1}{\gamma})}\left(\frac{\gamma}{N}\right)^{\frac{\gamma-1}{\gamma}} \int_{Nm}^{NM} e^{-\frac{x^\gamma}{\gamma N^{\gamma-1}}} dx \\
&= \frac{1}{\Gamma(\frac{1}{\gamma})} \int_{N\frac{m^\gamma}{\gamma}}^{N\frac{M^\gamma}{\gamma}} x^{-\frac{\gamma-1}{\gamma}} e^{-x} dx \\
&\leq \frac{\gamma^{\frac{\gamma-1}{\gamma}}}{\Gamma(\frac{1}{\gamma})(NM^\gamma)^{\frac{\gamma-1}{\gamma}}} \int_{N\frac{m^\gamma}{\gamma}}^{\infty} e^{-x} dx \\
&= \frac{\gamma^{\frac{\gamma-1}{\gamma}}}{\Gamma(\frac{1}{\gamma})(NM^\gamma)^{\frac{\gamma-1}{\gamma}}} e^{-N\frac{m^\gamma}{\gamma}},
\end{aligned} \tag{2.4.4}
$$

with the understanding that when $m = 0$, the last expression is 1. The difference between (2.4.4) and (2.4.2) is just that in the penultimate inequality the lower bound of the integral appeared in (2.4.2) where as in (2.4.4), the upper bound of the integral appeared.

$$\begin{aligned} q_N = P\left(\frac{\xi}{N} \in J\right) &\geq \frac{1}{2\Gamma(\frac{1}{\gamma})}\left(\frac{\gamma}{N}\right)^{\frac{\gamma-1}{\gamma}} \int_{Nm}^{NM} e^{-\frac{x^\gamma}{\gamma N^{\gamma-1}}} dx \\ &> \frac{1}{2\Gamma(\frac{1}{\gamma})} \int_{N\frac{m^\gamma}{\gamma}}^{N\frac{m^\gamma}{\gamma}+\delta} x^{-\frac{\gamma-1}{\gamma}} e^{-x} dx \\ &> \frac{\delta \gamma^{\frac{\gamma-1}{\gamma}}}{2\Gamma(\frac{1}{\gamma})(Nm^\gamma)^{\frac{\gamma-1}{\gamma}}} e^{-(N\frac{m^\gamma}{\gamma}+\delta)}, \end{aligned} \tag{2.4.5}$$

for any $0 < \delta < \frac{1}{\gamma}(M^\gamma - m^\gamma)$. The difference between (2.4.5) and (2.4.3) is just that in the penultimate inequality the upper limit of the integral appeared in (2.4.3) where as in (2.4.5), the upper limit of the integral to bound e^{-x} and lower limit to bound $x^{-\frac{\gamma-1}{\gamma}}$ is used. Moreover, when $m = 0$ the lower bound for q_N can be given by $\frac{1}{2\Gamma(\frac{1}{\gamma})} \int_{\delta_1}^{\delta_2} x^{-\frac{\gamma-1}{\gamma}} e^{-x} dx$ for $0 < \delta_1 < \delta_2 < \frac{M^\gamma}{\gamma}$. As earlier, this bound does not depend on N.

From now on we assume that $\frac{k(i,N)}{N} \to p_i$ for $1 \leq i \leq n$ with $p_1 > 0$. Clearly, $\sum p_i = 1$. Let

$$\Psi = \{\tilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} \frac{|x_i|^\gamma}{\gamma} \leq \sum_{i=1}^{k} p_i \log 2,\ 1 \leq k \leq n\}. \tag{2.4.6}$$

Proposition 2.4.1. *$\mu_N \Rightarrow \delta_0$ a.s. as $N \to \infty$.*

Proof. For any $\epsilon > 0$, define $\triangle(\epsilon) = [-\epsilon, \epsilon] \times \cdots \times [-\epsilon, \epsilon] \subseteq \mathbb{R}^n$. By Markov inequality,

$$\mathbf{P}(\mu_N(\triangle^c(\epsilon)) > \epsilon) < \frac{1}{\epsilon}\mathbf{E}\mu_N(\triangle^c(\epsilon)) < \frac{n}{\epsilon}\mathbf{P}(|\xi| > \epsilon N) < \frac{2n}{\epsilon}\mathbf{P}(\xi > \epsilon N) < \frac{2n}{\epsilon} C_N e^{-N\frac{\epsilon^\gamma}{\gamma}},$$

where C_N can be obtained from (2.4.2) for $\gamma \geq 1$ and from (2.4.4) for $0 < \gamma < 1$. Since $\frac{1}{N}\log C_N \to 0$ as $N \to \infty$, the proposition follows from the Borel-Cantelli lemma. □

Proposition 2.4.2. *If $\bar{\triangle} \cap \Psi = \phi$, then a.s. eventually $\mu_N(\triangle) = 0$. Moreover, the sequence $\{\mu_N\}$ is supported on a compact set.*

Proof. $\bar{\triangle} \cap \Psi = \phi$ implies $\tilde{m} \notin \Psi$. This is seen as follows. By definition of m_i, either m_i or $-m_i$ is in $\bar{\triangle}_i$. Thus for each i, there is an $\epsilon_i = \pm 1$ such that $\epsilon_i m_i \in \bar{\triangle}_i$. Thus the vector $(\epsilon_1 m_1, \cdots, \epsilon_n m_n) \in \bar{\triangle}$ and hence $\notin \Psi$. By the symmetry of Ψ, $\tilde{m} \notin \Psi$ as well. As a consequence, for some j, $1 \le j \le n$,

$$\sum_{i=1}^{j} \frac{m_i^\gamma}{\gamma} > \sum_{i=1}^{j} p_i \log 2. \tag{2.4.7}$$

For $\gamma \ge 1$ using (2.4.2) and for $0 < \gamma < 1$ using (2.4.4) we can say that $q_{iN} < C_{iN} e^{-N\frac{m_i^\gamma}{\gamma}}$ where $\frac{1}{N}\log C_{iN} \to 0$ as $N \to \infty$ for $1 \le i \le j$. Hence as a consequence of (2.4.7) and the fact $\frac{k(i,N)}{N} \to p_i$, we have

$$\sum_{N \ge 1} 2^{k(1,N)+\cdots+k(j,N)} q_{1N} \cdots q_{jN} < \sum_{N \ge 1} e^{-N \sum\limits_{i=1}^{j} \left(\frac{m_i^\gamma}{\gamma} - \frac{k(i,N)}{N} \log 2 - \frac{1}{N} \log C_{iN} \right)} < \infty.$$

Thus by Corollary 2.3.2, a.s. eventually $\mu_N(\triangle) = 0$.

To see the last statement of the Proposition, fix any $\delta > 0$. Let J be the compact set $[-\log 2 - \delta, +\log 2 + \delta]^n$. Since the complement of this set is union of 2^n open rectangles of $\mathbb{R}^n$, each of whose closures are disjoint with Ψ, the earlier part implies that eventually $\mu_N(J) = 1$. □

Proposition 2.4.3. *If $(\bar{\triangle} \cap \Psi)^0 \neq \phi$, then for any $\epsilon > 0$ a.s. eventually*

$$(1-\epsilon) q_{1N} \cdots q_{nN} \le \mu_N(\triangle) \le (1+\epsilon) q_{1N} \cdots q_{nN}.$$

Proof. The assumption $(\bar{\triangle} \cap \Psi)^0 \neq \phi$ implies $\tilde{m} \in \Psi^0$. Indeed, since $(\bar{\triangle} \cap \Psi)^0 \neq \phi$, pick $(x_1, \cdots, x_n) \in (\bar{\triangle} \cap \Psi)^0$. By symmetry of Ψ, $(|x_1|, \cdots, |x_n|) \in \Psi^0$ as well, and now $0 \le m_i \le |x_i|$ for all i yields $(m_1, \cdots, m_n) \in \Psi^0$.

We are going to show that the hypothesis of Corollary 2.3.2(b) holds. Fix i, $1 \le i \le n$. Using (2.4.3) for $\gamma \ge 1$ and using (2.4.5) for $\gamma < 1$, we can say that $q_{jN} > C_{jN} e^{-N\frac{m_j^\gamma}{\gamma} + \delta}$ for $1 \le j \le n$ with sufficiently small $\delta > 0$. Thus

$$2^{-(k(1,N)+\cdots+k(i,N))} q_{1N}^{-1} \cdots q_{iN}^{-1} < e^{-N\left[\sum\limits_{j=1}^{i} \left(\frac{k(j,N)}{N} \log 2 - \frac{m_j^\gamma}{\gamma} + \frac{1}{N} \log C_{iN} \right) - i\delta \right]}.$$

Since $\tilde{m}$ is an interior point of Ψ, there is an $\alpha > 0$ such that $\sum_{j=1}^{i} p_j \log 2 - \sum_{j=1}^{i} \frac{m_j^\gamma}{\gamma} > \alpha$. Now use the fact that $\frac{k(j,N)}{N} \to p_j$ and $\frac{1}{N}\log C_{jN} \to 0$ as $N \to \infty$ to deduce that eventually $\sum_{j=1}^{i}\left(\frac{k(j,N)}{N}\log 2 - \frac{m_j^\gamma}{\gamma} + \frac{1}{N}\log C_{iN}\right) > \alpha$. Making $\delta > 0$ smaller, if necessary, assume that eventually $\sum_{j=1}^{i}\left(\frac{k(j,N)}{N}\log 2 - \frac{m_j^\gamma}{\gamma} + \frac{1}{N}\log C_{iN}\right) - i\delta > \alpha$. Hence, eventually $e^{-N\left[\sum_{j=1}^{i}\left(\frac{k(j,N)}{N}\log 2 - \frac{m_j^\gamma}{\gamma} + \frac{1}{N}\log C_{iN}\right) - i\delta\right]} < e^{-N\alpha}$. As a consequence,

$$\sum_{N\geq 1} 2^{-(k(1,N)+\cdots+k(i,N))} q_{1N}^{-1}\cdots q_{iN}^{-1} < \infty.$$

Hence by Corollary 2.3.2, the proposition follows. □

Remark 2.4.1. $(\bar{\triangle}\cap\Psi)^0 \neq \phi$ *implies in particular, that* $p_1 > 0$. *In fact,* $\Psi^0 \neq \phi$ *iff* $p_1 > 0$.

Now, we have the following,

Proposition 2.4.4. *For a.e. sample point* ω,

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\log\mu_N(\triangle) &= -\sum_{i=1}^{n}\frac{m_i^\gamma}{\gamma} \quad \text{if } (\bar{\triangle}\cap\Psi)^0 \neq \phi\\ &= -\infty \quad \text{if } \bar{\triangle}\cap\Psi = \phi.\end{aligned}$$

Proof. When $\bar{\triangle}\cap\Psi = \phi$, the result is immediate from Proposition 2.4.2.

Assume that $(\bar{\triangle}\cap\Psi)^0 \neq \phi$. Fix any ϵ, $0 < \epsilon < 1$. Let $\gamma \geq 1$. By (2.4.2), $\frac{1}{N}\log q_{iN} < \frac{1}{N}\log C_{iN} - \frac{m_i^\gamma}{\gamma}$ where $\frac{1}{N}\log C_{iN} \to 0$ as $N \to \infty$. Hence $\limsup \frac{1}{N}\log q_{iN} \leq \frac{m_i^\gamma}{\gamma}$. Similarly, by using (2.4.3) we get $\liminf \frac{1}{N}\log q_{iN} \geq \frac{m_i^\gamma}{\gamma}$. Thus $\lim \frac{1}{N}\log q_{iN}$ exists and equals to $\frac{m_i^\gamma}{\gamma}$ for each i. The same holds even if $0 < \gamma < 1$, where we need to use (2.4.4) and (2.4.5). Then by proposition 2.4.3 we have a.s eventually, a.s. eventually

$$(1-\epsilon)q_{1N}\cdots q_{nN} \leq \mu_N(\triangle) \leq (1+\epsilon)q_{1N}\cdots q_{nN}.$$

So by taking logarithms and using $\lim_{N\to\infty}\frac{1}{N}\log q_{iN} = -\frac{m_i^\gamma}{\gamma}$, for each i we get the proposition. □

Let us consider the map $I : \mathbb{R}^n \to \mathbb{R}$, defined as follows,

$$\begin{aligned} I(\widetilde{x}) &= \tfrac{1}{\gamma}\sum_{i=1}^{n} |x_i|^\gamma \quad \text{if } \widetilde{x} \in \Psi \\ &= \infty \qquad\qquad\quad \text{otherwise.} \end{aligned} \tag{2.4.8}$$

Theorem 2.4.5. *Almost surely, the sequence $\{\mu_N\}$ satisfies LDP with the rate function I.*

Proof. Let $\mathcal{A}$ be the collection of all rectangles $\triangle = \triangle_1 \times \cdots \times \triangle_n \subseteq \mathbb{R}^n$ such that each $\triangle_i$ is a bounded interval with rational endpoints and either $\bar{\triangle} \cap \Psi = \phi$ or $(\bar{\triangle} \cap \Psi)^0 \neq \phi$.

It is easy to check that $\mathcal{A}$ forms a base for the usual topology of $\mathbb{R}^n$. For $\triangle \in \mathcal{A}$, by Proposition 2.4.4, the limit, $-\lim\limits_{N\to\infty} \dfrac{1}{N} \log \mu_N(\triangle)$ exists almost surely. Denote this limit by $L_\triangle$. Since $\mathcal{A}$ is a countable family, out side a null set, these limits are well defined for all $\triangle \in \mathcal{A}$.

In view of Proposition 0.3.3, to complete the proof, we show that for $\tilde{x} \in \mathbb{R}^n$,

$$I(\tilde{x}) = \sup_{\tilde{x} \in \triangle \in \mathcal{A}} L_\triangle. \tag{2.4.9}$$

If $\tilde{x} \notin \Psi$, clearly $\sup\limits_{\tilde{x} \in \triangle \in \mathcal{A}} L_\triangle = \infty = I(\tilde{x})$.

Now consider, $\tilde{x} = (x_1, \cdots, x_n) \in \Psi$. Suppose $\tilde{x} \in \triangle \in \mathcal{A}$. If $\triangle = \triangle_1 \times \cdots \times \triangle_n$ with $m_i = \inf\limits_{y \in \triangle_i} |y|$, then $m_i \leq |x_i|$ and hence $\frac{m_i^\gamma}{\gamma} \leq \frac{|x_i|^\gamma}{\gamma}$. Therefore, by Proposition 2.4.4, $L_\triangle = \sum\limits_{i=1}^{n} \frac{m_i^\gamma}{\gamma} \leq \sum\limits_{i=1}^{n} \frac{|x_i|^\gamma}{\gamma}$. Thus

$$\sup_{\tilde{x} \in \triangle \in \mathcal{A}} L_\triangle \leq I(\tilde{x}). \tag{2.4.10}$$

On the other hand, consider $\epsilon > 0$ so that $\epsilon < |x_i|$ for any i with $x_i \neq 0$. Let $\triangle$ be the box with sides $\triangle_i = (x_i - \epsilon, x_i + \epsilon)$. By choice of ϵ, $m_i = \inf\limits_{y \in \triangle_i} |y|$ equals $|x_i \pm \epsilon|$ depending on the sign of x_i. Of course, if $x_i = 0$ then $m_i = 0$. Thus for the $\triangle$ so constructed, we have, $L_\triangle = \sum\limits_{\{i : x_i \neq 0\}} \frac{|x_i \pm \epsilon|^\gamma}{\gamma}$. This being true for all sufficiently small ϵ, we conclude that

$$\sup_{\tilde{x}\in\triangle\in\mathcal{A}} L_\triangle \geq \sum_{i=1}^{n} \frac{|x_i|^\gamma}{\gamma} = I(\tilde{x}) \tag{2.4.11}$$

(2.4.10) and (2.4.11) complete the proof of (2.4.9) thus completing the proof of the theorem. □

We shall now proceed towards an expression for the free energy. Denoting $f(\tilde{x}) = \sum\limits_{i=1}^{n} \beta a_i x_i$,

$$\begin{aligned}\lim_N \tfrac{1}{N}\log Z_N(\beta) &= \log 2 + \lim_N \tfrac{1}{N}\log \mathbf{E}_N e^{-Nf} \\ &= \log 2 - \inf_{\tilde{x}\in\Psi}\{\sum_{i=1}^{n} \beta a_i x_i + \sum_{i=1}^{n} \tfrac{|x_i|^\gamma}{\gamma}\}.\end{aligned}$$

by Proposition 0.3.4. This last infimum equals $\inf\limits_{\tilde{x}\in\Psi} \sum\limits_{i=1}^{n}\left(\frac{|x_i|^\gamma}{\gamma} - \beta a_i x_i\right)$. Since $\beta > 0$, $a_i > 0$ it is easy to see that the above infimum is attained when all the x_i are negative. In other words, by symmetry of Ψ, the infimum is attained at a point $-\tilde{x}$ for some $\tilde{x} \in \Psi^+ = \Psi \cap \{\tilde{x} : x_i \geq 0 \text{ for } 1 \leq i \leq n\}$. Thus

$$\lim_N \frac{1}{N}\log Z_N(\beta) = \log 2 - \inf_{\tilde{x}\in\Psi^+} \sum_{i=1}^{n}\left(\frac{x_i^\gamma}{\gamma} - \beta a_i x_i\right).$$

In this way, for the above mentioned class of driving distributions, the free energy exists almost surely and is a constant. Not only that, finding an explicit formula for the free energy reduces to calculating the above infimum.

Remark 2.4.2. It is also worth noting that the LDP holds good even when the driving distributions at various levels are different. To be more specific, let us fix n numbers $\gamma_1, \cdots, \gamma_n$; each greater than zero and consider an n level GREM where the driving distribution at the i-th level is ϕ_{N,γ_i}. More precisely, for any node $\sigma_1 \cdots \sigma_i$ at the i-th level $\xi(\sigma_1 \cdots \sigma_i)$ has density ϕ_{N,γ_i}. Of course, all the random variables are independent. Define as earlier, the map $\Sigma_N \to \mathbb{R}^n$ by

$$\sigma \mapsto \left(\frac{\xi(\sigma_1)}{N}, \frac{\xi(\sigma_1\sigma_2)}{N}, \cdots, \frac{\xi(\sigma_1\cdots\sigma_n)}{N}\right).$$

Let μ_N be the induced probability on $\mathbb{R}^n$ when Σ_N is equipped with uniform probability. The same arguments as above, with $q_{iN} = P\left(\frac{\xi(\sigma_1\cdots\sigma_i)}{N} \in \triangle_i\right)$, will show that almost surely, the sequence of probabilities $\{\mu_N, N \geq n\}$ on $\mathbb{R}^n$ satisfies LDP. In this case, with rate function I will be given by

$$\begin{aligned} I(\widetilde{x}) &= \sum_{i=1}^{n} \frac{|x_i|^{\gamma_i}}{\gamma_i} \quad \text{if } \widetilde{x} \in \Psi \\ &= \infty \quad \text{otherwise,} \end{aligned} \tag{2.4.12}$$

where

$$\Psi = \{\widetilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} \frac{|x_i|^{\gamma_i}}{\gamma_i} \leq \sum_{i=1}^{k} p_i \log 2, \;\; 1 \leq k \leq n\}, \tag{2.4.13}$$

with $p_i = \lim\limits_{N\to\infty} \frac{k(i,N)}{N}$. Let, as earlier, Ψ^+ be the part of Ψ in the positive orthant of $\mathbb{R}^n$. As a consequence of all this, we have the following:

Theorem 2.4.6. *If the driving distribution has density ϕ_{N,γ_i} at the i-th level, we have almost surely,*

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{\widetilde{x}\in\Psi^+} \left\{ \sum_{i=1}^{n} \left(\frac{x_i^{\gamma_i}}{\gamma_i} - \beta a_i x_i \right) \right\}.$$

2.5 Inside Out

A close observation of the above discussion reveals the following cute idea. Though the identification, at first glance, will look like very simple, its implication in GREM will be understood through the rest of this chapter.

Let for each j, $1 \leq j \leq n$, we have a sequence of probabilities $\{\lambda_N^j, N \geq n\}$ on $\mathbb{R}$ which obey LDP with a strictly quasi-convex continuous good rate function $\mathcal{I}_j$. That is, $\mathcal{I}_j$ has compact level sets and for any two distinct points x and y in $\{0 < \mathcal{I}_j < \infty\}$ we have $\mathcal{I}_j(\theta x + (1-\theta)y) < \max\{\mathcal{I}_j(x), \mathcal{I}_j(y)\}$ for any θ with $0 < \theta < 1$. For the sake of simplicity, we will also assume that $\mathcal{I}_j(0) = 0$. The assumption of strict quasi-convexity is purely a technical assumption and this can be replaced by similar other conditions also. For example, one can replace this by requiring that the set $\{x : \mathcal{I}_j(x) = \alpha\}$ is a nowhere dense set for every $\alpha > 0$. We mentioned this condition in Remark 1.2.2, but there we demanded this only for $\alpha = \log 2$. Now, let us denote $\{+1,-1\}^N$ by Σ_N. For each N, let $k(1,N), \ldots, k(n,N)$ be non-negative integers adding to N and put $\Sigma_{jN} = \{+1,-1\}^{k(j,N)}$. Clearly, $\Sigma_N = \Sigma_{1N} \times \Sigma_{2N} \times \cdots \times \Sigma_{nN}$ and we express $\sigma \in \Sigma_N$ as $\sigma_1\sigma_2\cdots\sigma_n$ with $\sigma_i \in \Sigma_{iN}$, in an obvious way. Suppose for fixed N, we have a bunch of independent random variables as follows: $\{\xi(\sigma_1) : \sigma_1 \in \Sigma_{1N}\}$ having distributions λ_N^1, $\{\xi(\sigma_1\sigma_2) : \sigma_2 \in \Sigma_{2N}, \sigma_1 \in \Sigma_{1N}\}$ having distributions λ_N^2 and in general $\{\xi(\sigma_1\sigma_2\cdots\sigma_{j-1}\sigma_j) : \sigma_j \in \Sigma_{jN}, \cdots, \sigma_1 \in \Sigma_{1N}\}$ having distribution λ_N^j.

Define for each ω, $\mu_N(\omega)$ to be the empirical measure on $\mathbb{R}^n$, namely,

$$\mu_N(\omega) = \frac{1}{2^N} \sum_{\sigma} \delta \left\langle \xi(\sigma_1, \omega), \xi(\sigma_1\sigma_2, \omega), \cdots, \xi(\sigma_1 \cdots \sigma_n, \omega) \right\rangle$$

where $\delta \langle x \rangle$ denotes the point mass at $x \in \mathbb{R}^n$.

Theorem 2.5.1. *Suppose* $\frac{k(j,N)}{N} \to p_j > 0$ *for* $1 \leq j \leq n$. *Then for a.e.* ω, *the sequence* $\{\mu_N(\omega), N \geq n\}$ *satisfies LDP with rate function* $\mathcal{J}$ *given as follows:*

Supp$(\mathcal{J}) = \{(x_1, \cdots, x_n) : \sum_{k=1}^{j} \mathcal{I}_k(x_k) \leq \sum_{k=1}^{j} p_k \log 2$ *for* $1 \leq j \leq n\}$

and

$$\begin{aligned} \mathcal{J}(x) &= \sum_{k=1}^{n} \mathcal{I}_k(x_k) \quad \text{if } x \in \textit{Supp}(\mathcal{J}) \\ &= \infty \qquad\qquad \text{otherwise.} \end{aligned}$$

Proof. In what follows $\triangle$ denotes a box in $\mathbb{R}^n$ with sides $\triangle_j$; $1 \leq j \leq n$ where each $\triangle_j$ is an interval. The proof consists of the following steps. The steps are executed one by one as in Propositions 2.4.2 to 2.4.4, so will not be repeated here.

Step 1: *If* $\bar{\triangle} \cap$ *Supp*$(\mathcal{J}) = \phi$, *then a.s. eventually* $\mu_N(\triangle) = 0$.

Step 2: *If* $(\bar{\triangle} \cap$ *Supp*$(\mathcal{J}))^0 \neq \phi$, *then for any* $\epsilon > 0$ *a.s. eventually*

$$(1-\epsilon) \prod_{i=1}^{n} \lambda_N^i(\triangle_i) \leq \mu_N(\triangle) \leq (1+\epsilon) \prod_{i=1}^{n} \lambda_N^i(\triangle_i).$$

Step 3: *For a.e. sample point* ω,

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \log \mu_N(\triangle) &= -\sum_{i=1}^{n} \mathcal{I}_i(\triangle_i) \quad \text{if } (\bar{\triangle} \cap \textit{Supp}(\mathcal{J}))^0 \neq \phi \\ &= -\infty \qquad\qquad \text{if } \bar{\triangle} \cap \textit{Supp}(\mathcal{J}) = \phi, \end{aligned}$$

where $\mathcal{I}_i(\triangle_i) = \inf\{\mathcal{I}(x) : x \in \triangle_i\}$.

To conclude the proof we use the idea of Theorem 2.4.5. □

We note that continuity of the rate functions $\mathcal{I}_j$ is not necessary, but then one needs to go through lim sup and lim inf of $\frac{1}{N} \log \mu_N(\triangle)$ an in Theorem 1.2.1, instead of limits which we used above.

The implications of the above theorem for GREM [17] are clear. For fixed N, and $\sigma \in \Sigma_N$ one defines the Hamiltonian

$$H_N(\sigma) = N \sum_{i=1}^{n} a_i \xi(\sigma_1 \cdots \sigma_i).$$

Here $a_i,\ 1 \leq i \leq n$ are positive numbers called weights. In the Gaussian case, it is customary to take $\sum a_i^2 = 1$, though it is not a mathematical necessity. As earlier, $Z_N(\beta) = \sum_\sigma e^{-\beta H_N(\sigma)}$. Special choices of λ_N^i lead to all the known models considered. Centered Gaussian were consider in [17, 8, 19, 29]. More general distributions as well as the cases when some p_j are zero were considered in [29]. Moreover one could take different distributions for different values of j, see §2.7 for some interesting consequences. Thus the main problem of GREM is reduced to a variational problem. Note that, if $n = 1$, GREM reduces to REM.

2.6 The Variational Problem

In this section, we derive explicit formulae for the free energy. We return back to the driving distribution given by (2.4.1), namely, having density

$$\phi_{N,\gamma}(x) = \frac{1}{2\Gamma(\frac{1}{\gamma})} \left(\frac{\gamma}{N}\right)^{\frac{\gamma-1}{\gamma}} e^{-\frac{|x|^\gamma}{\gamma N^{\gamma-1}}} \quad -\infty < x < \infty. \tag{2.6.1}$$

We now consider the model with same driving distributions at different levels. In this setup, we need to calculate the infimum

$$\inf_{\tilde{x} \in \Psi^+} \sum_{i=1}^{n} \left(\frac{x_i^\gamma}{\gamma} - \beta a_i x_i \right) \tag{2.6.2}$$

in order to get an explicit formula for the limiting free energy. Note that, putting $\gamma_i = \gamma$ got all i in (2.4.13) we get

$$\Psi = \{\tilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} |x_i|^\gamma \leq \sum_{i=1}^{k} \gamma p_i \log 2, \quad 1 \leq k \leq n\} \tag{2.6.3}$$

and as usual Ψ^+ is the part of Ψ in the positive orthant of $\mathbb{R}^n$.

2.6.1 $\gamma > 1$

First let us assume $\gamma > 1$. To evaluate the infimum let us put, for $1 \le j \le k \le n$,

$$B(j,k) = \left(\frac{(p_j + \cdots + p_k)\gamma \log 2}{a_j^{\frac{\gamma}{\gamma-1}} + \cdots + a_k^{\frac{\gamma}{\gamma-1}}} \right)^{\frac{\gamma-1}{\gamma}}. \tag{2.6.4}$$

Set $r_0 = 0$ and for $l \ge 0$ (integer),

$$\beta_{l+1} = \min_{k > r_l} B(r_l + 1, k) \quad r_{l+1} = \max\{i > r_l : B(r_l + 1, i) = \beta_{l+1}\}. \tag{2.6.5}$$

Clearly, for some K with $1 \le K \le n$, we have $r_K = n$. Put $\beta_0 = 0$ and $\beta_{K+1} = \infty$, so that $0 = \beta_0 < \beta_1 < \beta_2 \cdots < \beta_K < \beta_{K+1} = \infty$.

Fix $j \le K$ and let $\beta \in (\beta_j, \beta_{j+1}]$. Define $\widetilde{\overline{x}} \in \Psi^+$ as follows:

$$\begin{aligned} \overline{x}_i &= (\beta_l a_i)^{\frac{1}{\gamma-1}} \quad \text{if } i \in \{r_{l-1}+1, \cdots, r_l\} \text{ for some } l, 1 \le l \le j \\ &= (\beta a_i)^{\frac{1}{\gamma-1}} \quad \text{if } i \ge r_j + 1. \end{aligned} \tag{2.6.6}$$

Claim: $\inf_{\widetilde{x} \in \Psi^+} \sum_{i=1}^{n} \left(\frac{x_i^\gamma}{\gamma} - \beta a_i x_i \right)$ occurs at $\widetilde{\overline{x}}$.

In order to prove the claim, fix any $\widetilde{x} \in \Psi^+$. For $k \le j$ (recall that $j \le K$ was fixed above), first note that, by Holder's inequality,

$$\sum_{i=1}^{r_k} x_i \overline{x}_i^{\gamma-1} \le \left(\sum_{i=1}^{r_k} x_i^\gamma \right)^{\frac{1}{\gamma}} \left(\sum_{i=1}^{r_k} \overline{x}_i^\gamma \right)^{\frac{\gamma-1}{\gamma}} \le \sum_{i=1}^{r_k} \overline{x}_i^\gamma.$$

where the last inequality follows from the facts $\widetilde{x} \in \Psi^+$ and $\sum_{i=1}^{r_k} \overline{x}_i^\gamma = \sum_{i=1}^{r_k} \gamma p_i \log 2$ so that $\sum_{i=1}^{r_k} x_i^\gamma \le \sum_{i=1}^{r_k} \gamma p_i \log 2 = \sum_{i=1}^{r_k} \overline{x}_i^\gamma$.

Hence, $\sum_{i=1}^{r_k} \overline{x}_i^{\gamma-1}(\overline{x}_i - x_i) \ge 0$.

Since $\beta > \beta_j$, we have $(\frac{\beta}{\beta_l} - 1) > 0$ for $1 \le l \le j$. Moreover since β_l are increasing with l, these number $(\frac{\beta}{\beta_l} - 1)$ are decreasing. It follows that,

$$\sum_{l=1}^{j} \left(\frac{\beta}{\beta_l} - 1 \right) \sum_{i=r_{l-1}+1}^{r_l} \overline{x}_i^{\gamma-1}(\overline{x}_i - x_i) \ge 0$$

In other words, using the definition of $\overline{x}_i$,

$$\sum_{i=1}^{r_j} \beta a_i(\overline{x}_i - x_i) \geq \sum_{i=1}^{r_j} \overline{x}_i^{\gamma-1}(\overline{x}_i - x_i). \tag{2.6.7}$$

Now,

$$\begin{aligned}
&\sum_{i=1}^{r_j}\left(\frac{x_i^\gamma}{\gamma} - \beta a_i x_i\right) - \sum_{i=1}^{r_j}\left(\frac{\overline{x}_i^\gamma}{\gamma} - \beta a_i \overline{x}_i\right)\\
=&\sum_{i=1}^{r_j}\left(\frac{x_i^\gamma}{\gamma} + \beta a_i(\overline{x}_i - x_i) - \frac{\overline{x}_i^\gamma}{\gamma}\right)\\
\geq&\sum_{i=1}^{r_j}\left(\frac{x_i^\gamma}{\gamma} + \overline{x}_i^{\gamma-1}(\overline{x}_i - x_i) - \frac{\overline{x}_i^\gamma}{\gamma}\right) \qquad \text{by (2.6.7)}\\
=&\sum_{i=1}^{r_j}\left(\frac{x_i^\gamma}{\gamma} + \frac{\gamma-1}{\gamma}\overline{x}_i^\gamma - x_i\overline{x}_i^{\gamma-1}\right)\\
\geq&0.
\end{aligned} \tag{2.6.8}$$

where in the last inequality we used $x_i\overline{x}_i^{\gamma-1} \leq \frac{1}{\gamma}x_i^\gamma + \frac{\gamma-1}{\gamma}\overline{x}_i^\gamma$.

On the other hand, utilizing the definition of $\overline{x}_i$ and the inequality $\beta a_i x_i \leq \frac{x_i^\gamma}{\gamma} + \frac{\gamma-1}{\gamma}(\beta a_i)^{\frac{\gamma}{\gamma-1}}$ we have,

$$\begin{aligned}
&\sum_{i=r_j+1}^{n}\left(\frac{x_i^\gamma}{\gamma} - \beta a_i x_i\right) - \sum_{i=r_j+1}^{n}\left(\frac{\overline{x}_i^\gamma}{\gamma} - \beta a_i \overline{x}_i\right)\\
=&\sum_{i=r_j+1}^{n}\left(\frac{x_i^\gamma}{\gamma} + \frac{\gamma-1}{\gamma}(\beta a_i)^{\frac{\gamma}{\gamma-1}} - \beta a_i x_i\right)\\
\geq&0.
\end{aligned} \tag{2.6.9}$$

Clearly, (2.6.8) and (2.6.9) complete proof of the claim. This argument is in fact a generalization of Dorlas & Dukes [19], Capocaccia et. al. [8].

All this leads to the following explicit formula for the free energy.

Theorem 2.6.1. *For GREM with driving distribution having density $\phi_{N,\gamma}$ as defined* *in (2.6.1), almost surely,*

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \sum_{i=r_j+1}^{n} p_i \log 2 + \tfrac{\gamma-1}{\gamma} \sum_{i=r_j+1}^{n} (\beta a_i)^{\frac{\gamma}{\gamma-1}} + \beta \sum_{l=1}^{j} \beta_l^{\frac{1}{\gamma-1}} \sum_{i=r_{l-1}+1}^{r_l} a_i^{\frac{\gamma}{\gamma-1}}$$

$$\text{if} \quad \beta_j < \beta \leq \beta_{j+1},\ 0 \leq j \leq K-1$$

$$= \beta \sum_{l=1}^{K} \beta_l^{\frac{1}{\gamma-1}} \sum_{i=r_{l-1}+1}^{r_l} a_i^{\frac{\gamma}{\gamma-1}} \qquad \text{if} \quad \beta > \beta_K$$

Observe that for $\gamma = 2$, that is when the driving distribution is Normal, with proper identification of parameters this is essentially the same formula as in [8, 19]. In defining the β_i, Capocaccia *et. al.* use a variant in [8] (§3.2). In defining r_i, Dorlas and Dukes [19] consider the least index, where as recall that, we define r_{l+1} as $\max\{i > r_l : B(r_l+1, i) = \beta_{l+1}\}$. This makes no difference because *'nothing happens'* in between these two indices. This follows from the fact that if $a_i > 0$, $b_i > 0$ for $i = 1, 2, 3$ and $\frac{b_1}{a_1} = \frac{b_1+b_2+b_3}{a_1+a_2+a_3} < \frac{b_1+b_2}{a_1+a_2}$, then this will imply that $\frac{b_1}{a_1} = \frac{b_2+b_3}{a_2+a_3} < \frac{b_2}{a_2}$. So defining r_{l+1} as $\min\{i > r_l : B(r_l+1, i) = \beta_{l+1}\}$, when $\{i > r_l : B(r_l+1, i) = \beta_{l+1}\}$ is not a singleton set, β_{l+2} will be same as β_{l+1}. And this will continue until the maximum index of the set $\{i > r_l : B(r_l+1, i) = \beta_{l+1}\}$ is attained.

Moreover, the weights a_i in Dorlas and Dukes [19] are incorporated in the density, there was no need to assume $\sum a_i = 1$, their parameter J can be incorporated in the weights. In fact, there is one benefit of putting the weights in the density. The large deviation technique will easily allow us to consider variable weights a_{iN} depending on N at the i-th level instead of a constant weights a_i. For instance, let $a_{iN} > 0$ for all $1 \leq i \leq n$ and $N \geq 1$ be the weights of the i-th level for the N particle system with $a_{iN} \to a_i$ as $N \to \infty$. When the weights a_i did not depend on N, they were not brought in the large deviation argument. The free energy was

$$\log 2 - \inf_{\tilde{x} \in \Psi^+} \sum_{i=1}^{n} \left(\frac{x_i^2}{2} - \beta a_i x_i \right), \tag{2.6.10}$$

where

$$\Psi^+ = \{\widetilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} \frac{1}{2} x_i^2 \leq \sum_{i=1}^{k} p_i \log 2\} \cap \{\widetilde{x} \in \mathbb{R}^n : x_i \geq 0, 1 \leq i \leq n\}.$$

If we consider variable weights a_{iN} as above then they must enter in the large deviation arguments. If $\xi(i, N) \sim \mathcal{N}(0, \frac{1}{N})$, then it is not hard to show that the distribution of $a_{iN}\xi(i, N)$ satisfies LDP with rate function $\frac{x^2}{2a_i^2}$. Accordingly, we get the limiting free energy as

$$\log 2 - \inf_{\tilde{x} \in \Upsilon^+} \sum_{i=1}^{n} \left(\frac{x_i^2}{2a_i^2} - \beta x_i \right), \tag{2.6.11}$$

where

$$\Upsilon^+ = \{\widetilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} \frac{1}{2a_i^2} x_i^2 \leq \sum_{i=1}^{k} p_i \log 2\} \cap \{\widetilde{x} \in \mathbb{R}^n : x_i \geq 0, 1 \leq i \leq n\}.$$

Though we seem to have two different optimization problems in (2.6.10) and (2.6.11), they produce the same result as the former can be transform to the later using affine transforms $y_i = \frac{x_i}{a_i}$. So this will lead to the expected result that the limiting free energy of the Gaussian GREM is continuous with respect to its weights. Not only in the Gaussian case this can be made precise in all the models ($\gamma > 1$) discussed in this subsection by the same way and for other models with some extra efforts.

In the Gaussian case, that is when $\gamma = 2$, two simple cases are worth mentioning. The numbers β_j mentioned below are same as the above, in these particular cases.

Corollary 2.6.2. *(Gaussian Case)*

i) Let $0 < \frac{p_1}{a_1^2} < \frac{p_2}{a_2^2} < \cdots < \frac{p_n}{a_n^2}$. *Put* $\beta_j = \frac{\sqrt{2p_j \log 2}}{a_j}$ *for* $j = 1, \cdots, n$. *Then a.s.*

$$\begin{aligned}
\lim_N \tfrac{1}{N} \log Z_N(\beta) &= \log 2 + \tfrac{\beta^2}{2} \sum_1^n a_i^2 && \text{if } \beta < \beta_1, \\
&= \sum_{j+1}^{n} p_i \log 2 + \sum_1^j \beta a_i \sqrt{2p_i \log 2} + \tfrac{\beta^2}{2} \sum_{j+1}^{n} a_i^2 \\
& && \text{if } \beta_j \leq \beta < \beta_{j+1} \text{ for } 1 \leq j < n, \\
&= \beta \sum_1^n a_i \sqrt{2p_i \log 2} && \text{if } \beta \geq \beta_n.
\end{aligned}$$

ii) Let $\frac{p_1}{a_1^2} = \frac{p_2}{a_2^2} = \cdots = \frac{p_n}{a_n^2} > 0$. *Then a.s.*

$$\begin{aligned}
\lim_N \tfrac{1}{N} \log Z_N(\beta) &= \log 2 + \tfrac{\beta^2}{2} \sum_1^n a_i^2 && \text{if } \beta < \sqrt{\tfrac{2 \log 2}{\sum a_i^2}} \\
&= \beta \sqrt{2 \log 2 \sum a_i^2} && \text{if } \beta \geq \sqrt{\tfrac{2 \log 2}{\sum a_i^2}}.
\end{aligned}$$

Remark 2.6.1. We say that an n level GREM with some particular driving distribution is in *reduced form*, if the limiting free energy of the model can not be obtained from any k level GREM with same driving distribution where $k < n$.

For a Gaussian n-level GREM, as the above analysis shows, if it can not be obtained as a k-level GREM then β_is are defined for $1 \leq i \leq n$. On the other hand if it can be obtained as the energy function of a k-level Gaussian GREM for some $k < n$, then the β_is of the construction are only for $1 \leq i \leq k$.

If a GREM is in reduced form, according to this definition, we do not know whether its energy function can be obtain as that of a k-level GREM for some $k < n$ with, of course, different driving distributions. Along with the setup of the model in this subsection, we are lucky enough to get the explicit expression of the limiting free energy. Moreover we know the explicit expression of the β_is where the expression of the free energy are changing. We observed in this case that there may be at most n many β_is. But we do not know, whether this the intrinsic property of the model or there are some driving distributions so that for an n-level GREM, we can get more than n many β_is.

Remark 2.6.2. It can be shown that the energy function determines the parameters of the model for every $\gamma > 1$ and one could characterize functions those arise as energy functions for GREM. As observed in in the above Corollary, an n level GREM may reduce to a k level GREM for some $k < n$ or even to a REM. In such a case, some weights a_i occur in groups and get added up. Of course, in such a case when the model is not in reduced form, clearly it is not possible to recover the weights from the formula for energy. But it is interesting to note that when the GREM is in reduced form, we can recover the parameters from the energy function. To make the statement precise first of all note that, in this set up, GREM is in reduced from if and only if all the p_i, a_i are non zero and $\frac{p_1^{\frac{\gamma-1}{\gamma}}}{a_1} < \frac{p_2^{\frac{\gamma-1}{\gamma}}}{a_2} < \cdots < \frac{p_n^{\frac{\gamma-1}{\gamma}}}{a_n}$, let us assume this to be the case. This is similar to that of the Gaussian case. Note that, in this case $\beta_i = \frac{(\gamma p_i \log 2)^{\frac{\gamma-1}{\gamma}}}{a_i}$ for $1 \leq i \leq n$. From Theorem 2.6.1, it follows that the limiting free energy $\mathcal{E}(\beta)$ is a continuous function with $\mathcal{E}(0) = \log 2$. It has a continuous derivative $\mathcal{E}'(\beta)$ with

$$\mathcal{E}'(\beta) = \begin{cases} 0 & \text{if } \beta = 0 \\ \beta^{\frac{1}{\gamma-1}} \sum\limits_{i=k+1}^{n} a_i^{\frac{\gamma}{\gamma-1}} + \sum\limits_{i=1}^{k} \beta_i^{\frac{1}{\gamma-1}} a_i^{\frac{\gamma}{\gamma-1}} & \text{if } \beta_k < \beta \leq \beta_{k+1},\ 0 \leq k \leq n-1 \\ \sum\limits_{i=1}^{n} \beta_i^{\frac{1}{\gamma-1}} a_i^{\frac{\gamma}{\gamma-1}} & \text{if } \beta > \beta_n. \end{cases}$$

Further,

$$\mathcal{E}''(\beta) = \begin{cases} \frac{1}{\gamma-1}\beta^{\frac{2-\gamma}{\gamma-1}} \sum\limits_{1}^{n} a_i^{\frac{\gamma}{\gamma-1}}, & \text{for } 0 < \beta < \beta_1 \\ \frac{1}{\gamma-1}\beta^{\frac{2-\gamma}{\gamma-1}} \sum\limits_{k+1}^{n} a_i^{\frac{\gamma}{\gamma-1}}, & \text{for } \beta_k < \beta < \beta_{k+1}, \text{ for } 1 \leq k < n \\ 0, & \text{for } \beta > \beta_n. \end{cases}$$

The energy function can be characterized in this case. To start with, observe that the above energy function has the following properties:

i) $\mathcal{E}(0) = \log 2$ and $\mathcal{E}'(0) = 0$,

ii) $\mathcal{E}$ is a continuously differentiable function,

iii) denote $x_k = \frac{(\gamma p_k \log 2)^{\frac{\gamma-1}{\gamma}}}{a_k}$; $c_k = \frac{1}{\gamma-1}\beta^{\frac{2-\gamma}{\gamma-1}} \sum\limits_{k}^{n} a_i^{\frac{\gamma}{\gamma-1}}$ and $\theta = \frac{2-\gamma}{\gamma-1}$ then $0 \equiv x_0 < x_1 < \cdots < x_n < x_{n+1} \equiv \infty$; $c_1 > c_2 > \cdots > c_n > c_{n+1} \equiv 0$ and $\theta > -1$ with $\mathcal{E}''(\beta) = (1+\theta)c_i\beta^\theta$ in (x_{i-1}, x_i) for $1 \leq i \leq n+1$.

Conversely, let f be a function on $[0, \infty)$ such that

i)$f(0) = \log 2$ and $f'(0) = 0$,

ii) f has continuous first derivative,

iii) there are finitely many points $0 < x_1 < \cdots < x_n$ and $c_1 > \cdots > c_n > c_{n+1} = 0$ so that the left and right derivatives of f' are unequal at x_i for $1 \leq i \leq n$ and $f''(x) = x^\theta c_i$ in (x_{i-1}, x_i) for $1 \leq i \leq n+1$ with $x_0 = 0$ and $x_{n+1} = \infty$. Then f is the energy function for γ-GREM with driving distribution having parameter $\gamma = \frac{\theta+2}{\theta+1}$,

$p_i = \frac{\theta+1}{(\theta+2)\log 2}(x_i(c_i - c_{i+1}))^{\theta+2}$ and $a_i = (c_i - c_{i+1})^{\frac{1}{\theta+2}}$ for $1 \le i \le n$ if

$$\sum_{i=1}^{n} x_i^{\theta+2}(c_i - c_{i+1}) = \frac{\theta+2}{\theta+1}\log 2. \tag{2.6.12}$$

In particular for a Gaussian GREM, it is in reduced from if and only if all the p_i, a_i are non zero and $\frac{p_1}{a_1^2} < \frac{p_2}{a_2^2} < \cdots < \frac{p_n}{a_n^2}$. When that is the case, from Theorem 2.6.1 with $\gamma = 2$, it follows that the limiting free energy $\mathcal{E}(\beta)$ is piecewise quadratic continuous function with $\mathcal{E}(0) = \log 2$. It has a continuous derivative $\mathcal{E}'(\beta)$ with $\mathcal{E}'(0) = 0$ and

$$\mathcal{E}''(\beta) = \begin{cases} \sum_{1}^{n} a_i^2, & \text{in } (0, \frac{\sqrt{2p_1 \log 2}}{a_1}), \\ \sum_{k+1}^{n} a_i^2, & \text{in } (\frac{\sqrt{2p_k \log 2}}{a_k}, \frac{\sqrt{2p_{k+1}\log 2}}{a_{k+1}}), \text{ for } 1 \le k \le n-1, \\ 0, & \text{if } \beta > \frac{\sqrt{2p_n \log 2}}{a_n}. \end{cases}$$

Moreover, if f is a C^1 function on $[0,\infty)$ with $f(0) = \log 2$ and $f'(0) = 0$ so that there are finitely many points $0 < x_1 < \cdots < x_n$ where the left and right derivatives of f' are unequal and f'' is a positive constant, say, c_i in (x_{i-1}, x_i) with $x_0 = 0$ and $x_{n+1} = \infty$. Then f is the energy function for some Gaussian GREM iff

$$c_1 > \cdots > c_n > c_{n+1} = 0 \quad \text{and} \quad \sum_{i=1}^{n} x_i^2(c_i - c_{i+1}) = 2\log 2. \tag{2.6.13}$$

2.6.2 $\gamma = 1$

Now, let us assume $\gamma = 1$. Note that $\gamma = 1$ represents the two sided exponential distribution with mean 0 and parameter 1. In this case, we can not use the above argument directly as the ratios $B(j,k)$ defined in (2.6.4), the constants a_i appear with exponent $\frac{\gamma}{\gamma-1}$. However, to get the expression for the free energy, we can directly proceed to evaluate

$$\inf_{\tilde{x}\in\Psi^+} \sum_{i=1}^{n}(1 - \beta a_i)x_i,$$

where

$$\Psi^+ = \left\{ \widetilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} |x_i| \leq \sum_{i=1}^{k} p_i \log 2,\ \ 1 \leq k \leq n \right\} \cap \{\tilde{x} : x_i \geq 0 \text{ for } 1 \leq i \leq n\}.$$

This is what we will do now. To calculate the infimum, let us set $r_0 = 0$ and for $k = 1, 2, \cdots$, let us define β_k, r_k as follows:

$$\beta_1 = \min\{\frac{1}{a_i} : 1 \leq i \leq n\}$$
$$r_1 = \max\{i : \frac{1}{a_i} = \beta_1\}$$

and in general, for $k > 1$,

$$\beta_k = \min\{\frac{1}{a_i} : r_{k-1} < i \leq n\}$$
$$r_k = \max\{i : r_{k-1} < i \leq n,\ \frac{1}{a_i} = \beta_k\}.$$

Obviously this process stops at a finite stage say at K, so that $\beta_K = \frac{1}{a_n}$ and $r_K = n$. We put $\beta_{K+1} = \infty$. For example, if $a_1 > a_2 > \cdots > a_n$ then $\beta_k = \frac{1}{a_k}$, $r_k = k$ for $k = 1, 2, \cdots, n$, and $K = n$. On the other hand if $a_1 < a_2 < \cdots < a_n$ then $\beta_1 = \frac{1}{a_n}$, $r_1 = n$ and $K = 1$.

Clearly, $0 = \beta_0 < \beta_1 < \beta_2 \cdots < \beta_K < \beta_{K+1} = \infty$.

Remark 2.6.3. The case $\gamma = 1$ can also be recovered as a limiting case from the previous section. We can proceed by defining β_k as done in the last subsection. But now we have to take limit $\lim_{\gamma \downarrow 1} \left(\frac{(p_j + \cdots + p_k)\gamma \log 2}{a_j^{\frac{\gamma}{\gamma-1}} + \cdots + a_k^{\frac{\gamma}{\gamma-1}}} \right)^{\frac{\gamma-1}{\gamma}}$ to define $B(j,k)$. A simple calculation shows that, $B(j,k) = \frac{1}{\max_{j \leq i \leq k} a_i}$. Hence the β_ks defined in the earlier subsection lead to the same formula as above when $\beta \downarrow 1$.

Now, fix $j \leq K$ and let $\beta \in [\beta_j, \beta_{j+1})$. Define $\widetilde{\overline{x}} \in \Psi^+$ as follows:

$$\begin{aligned} \overline{x}_i &= \sum_{j=r_{(l-1)}+1}^{r_l} p_j \log 2 \quad \text{if } i = r_l \text{ for some } l, 1 \leq l \leq j \\ &= 0 \qquad\qquad\qquad\qquad \text{otherwise.} \end{aligned} \tag{2.6.14}$$

Claim: $\inf_{\widetilde{x}\in\Psi^+} \sum_{i=1}^{n}(1-\beta a_i)x_i$ occurs at $\widetilde{\overline{x}}$.

In case $j = 0$ that is $\beta_j = \beta_0 = 0$, the claim is obvious. Indeed, for $\beta < \beta_1$, $(1-\beta a_i)$ is positive for all i, the infimum occurs at $\widetilde{\overline{x}}$ with $\overline{x}_i = 0$ for all i. So let us assume that $j \geq 1$. First note that, since $\beta \geq \beta_j$, we have $\beta \geq \beta_l$ for $1 \leq l \leq j$ and $(1-\frac{\beta}{\beta_l}) = (1-\beta a_{r_l}) \leq 0$. Moreover since β_l are strictly increasing with l, the numbers a_{r_l} are strictly decreasing, that is, $a_{r_1} > a_{r_2} > \cdots > a_{r_K}$. Now to prove the claim, fix any $\widetilde{x} \in \Psi^+$.

$$\begin{aligned}
&\sum_{i=1}^{n}(1-\beta a_i)x_i - \sum_{i=1}^{n}(1-\beta a_i)\overline{x}_i \\
\geq &\sum_{i=1}^{r_j}(1-\beta a_i)x_i - \sum_{i=1}^{r_j}(1-\beta a_i)\overline{x}_i
\end{aligned}$$

(Since $(1-\beta a_i) \geq 0$ and $\overline{x}_i = 0$ for $i > r_j$)

$$= \sum_{l=1}^{j}\left(\sum_{i=r_{(l-1)}+1}^{r_l}(1-\beta a_i)x_i - (1-\beta a_{r_l})\overline{x}_{r_l}\right)$$

(By definition of $\overline{x}_i$)

$$\geq \sum_{l=1}^{j}(1-\beta a_{r_l})\left(\sum_{i=r_{(l-1)}+1}^{r_l} x_i - \overline{x}_{r_l}\right)$$

(Since by definition $a_{r_l} \geq a_i$ for $r_{(l-1)}+1 \leq i \leq r_l$)

$$\begin{aligned}
=&(1-\beta a_{r_j})\left(\sum_{1}^{r_j} x_i - \sum_{1}^{j}\overline{x}_{r_l}\right) + \sum_{l=1}^{j-1}\beta(a_{r_{(l+1)}} - a_{r_l})\left(\sum_{1}^{r_l} x_i - \sum_{1}^{l}\overline{x}_{r_i}\right) \\
=&(1-\beta a_{r_j})\sum_{1}^{r_j}(x_i - p_i\log 2) + \sum_{l=1}^{j-1}\beta(a_{r_{(l+1)}} - a_{r_l})\sum_{1}^{r_l}(x_i - p_i\log 2) \\
\geq&0
\end{aligned}$$

The last inequality follows from the facts that (i) by definition $\sum_{1}^{r_l}(x_i - p_i\log 2) \leq 0$, (ii) $(a_{r_{(l+1)}} - a_{r_l}) < 0$ for $1 \leq l \leq j$ and (iii) $(1-\beta a_{r_j}) \leq 0$. Hence, the proof of the claim.

Here then is the formula for the free energy.

Theorem 2.6.3. *For two sided exponential GREM, almost surely,*

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \begin{cases} \log 2 & \text{if } \beta < \beta_1 \\ \log 2 + \sum_{l=1}^{j} (\beta a_{r_l} - 1) \sum_{r_{l-1}+1}^{r_l} p_i \log 2 & \text{if } \beta_j \leq \beta < \beta_{j+1}. \end{cases}$$

Remark 2.6.4. Once again, the free energy for the case $\gamma = 1$ can be recovered from that of $\gamma > 1$ as a limiting case. It is quite easy to check that with notation of β_l and r_l as in subsection 2.6.1, $\lim_{\gamma \downarrow 1} (\beta_l a_i)^{\frac{1}{\gamma-1}} = \frac{1}{k} \sum_{j=r_{(l-1)}+1}^{r_l} p_j \log 2$ where $k = \#\{i : a_i = \max_{r_{(l-1)}+1 \leq i \leq r_l} a_i \,\&\, r_{(l-1)} + 1 \leq i \leq r_l\}$. Moreover, for $\beta a_i < 1$, we have $\lim_{\gamma \downarrow 1} (\beta a_i)^{\frac{1}{\gamma-1}} = 0$. So $y_i = \lim_{\gamma \downarrow 1} \overline{x}_i(\gamma)$, where $\overline{x}_i$ as given by (2.6.6), may not give $\overline{x}_i$ as defined in (2.6.14). The only difference will be that $\overline{x}_i = 0$ for $r_{(l-1)} + 1 \leq i < r_l$ and $\overline{x}_{r_l} = \sum_{j=r_{(l-1)}+1}^{r_l} p_j \log 2$ whereas, with the same notation of k as above, $y_i = \frac{1}{k} \sum_{j=r_{(l-1)}+1}^{r_l} p_j \log 2$ for those i where $r_{(l-1)} + 1 \leq i \leq r_l$ and $a_i = \max_{r_{(l-1)}+1 \leq i \leq r_l} a_i$. But it is easy to see that if $\inf_{\widetilde{x} \in \Psi^+} \sum_{i=1}^{n} (1 - \beta a_i) x_i$ occurs at $\widetilde{\overline{x}}$, then it will occur also at $\widetilde{y} = (y_i)$. Thus the limiting free energy in the case of $\gamma = 1$ is nothing but the limiting $(\gamma \to 1)$ case of the limiting free energy of the GREM where $\gamma > 1$.

Thus Remarks 2.6.3 and 2.6.4 will lead to the following

Theorem 2.6.4. *If $\mathcal{E}_\gamma(\beta)$ and $\mathcal{E}(\beta)$ denote the limiting free energy as given by Theorems 2.6.1 and 2.6.3 respectively, then for all $\beta \geq 0$, almost surely,*

$$\lim_{\gamma \downarrow 1} \mathcal{E}_\gamma(\beta) = \mathcal{E}(\beta).$$

As in the case of $\gamma = 2$, for the case $\gamma = 1$ also two special situations are worth mentioning

Corollary 2.6.5. *i) Let $a_1 > a_2 > \cdots > a_n$. Then a.s.*

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \begin{cases} \log 2 & \text{if } \beta < \frac{1}{a_1} \\ \log 2 + \sum_{i=1}^{k} (\beta a_i - 1) p_i \log 2 & \text{if } \frac{1}{a_k} \le \beta < \frac{1}{a_{k+1}} \\ \beta \sum_{i=1}^{n} a_i p_i \log 2 & \text{if } \beta \ge \frac{1}{a_n}. \end{cases}$$

ii) Let $a_1 \le a_2 \le \cdots \le a_n$. Then a.s.

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \begin{cases} \log 2 & \text{if } \beta < \frac{1}{a_n} \\ \beta a_n \log 2 & \text{if } \beta \ge \frac{1}{a_n}. \end{cases}$$

Remark 2.6.5. Returning to Theorem 2.6.3, it is interesting to note that exponential GREM with parameters $(p_1, \cdots, p_n, a_1, \cdots, a_n)$ is equivalent to GREM with parameters $(p'_1, \cdots, p'_K, a'_1, \cdots, a'_K)$ where $p'_1 = \sum_{1}^{r_1} p_j,\ p'_2 = \sum_{r_1+1}^{r_2} p_j, \cdots, p'_K = \sum_{r_{(K-1)}+1}^{n} p_j$ and $a'_1 = a_{r_1},\ a'_2 = a_{r_2}, \cdots, a'_K = a_{r_K}$. This is evident from Theorem 2.6.3. Here 'equivalent' is used in the sense that for every β, both systems have the same free energy. Thus, in order that an n-level GREM does not collapse to a lower level GREM it is necessary and sufficient that the weights a_i be strictly decreasing. One should keep in mind that we are using the same distribution at all levels of the GREM.

The purpose of the following remark is to show that the energy function determines the parameters of the model. One could characterize functions that arise as energy functions for exponential GREM.

Remark 2.6.6. As observed in the previous Remark, an n level GREM may reduce to a K level GREM for some $K < n$. In the exponential GREM, some weights a_i do not appear in the formula for free energy. When such a thing happens it is clearly not possible to recover the weights from the formula for energy. It is interesting to note that when the GREM is in reduced form, we can recover the parameters from the energy function. Here is the precise statement.

Since an exponential GREM is in reduced form if and only if $a_1 > \cdots > a_n > 0$

and $p_i \neq 0$ for $1 \leq i \leq n$, let us assume this to be the case. Let $\mathcal{E}(\beta)$ be the energy function, that is $\mathcal{E}(\beta) = \lim_N \frac{1}{N} \log Z_N(\beta)$. From Theorem 2.6.3, it is easy to see that $\mathcal{E}(\beta)$ is a piecewise linear continuous function of β taking value $\log 2$ near zero. Further, its derivative $\mathcal{E}'(\beta) = \sum_{i=1}^{k} a_i p_i \log 2$ in $(\frac{1}{a_k}, \frac{1}{a_{k+1}})$. These properties are good enough to show the following: $\mathcal{E}(\beta)$ uniquely determines all the quantities p_i and a_i. In other words, the energy function identifies the parameters.

If $0 < x_1 < \cdots < x_n$ be the points where the left and right derivatives of $\mathcal{E}(\beta)$ are unequal, then $a_i = \frac{1}{x_i}$. Further, if $\mathcal{E}'(\beta) = c_i$ in (x_i, x_{i+1}) then $p_i = \frac{x_i(c_i - c_{i-1})}{\log 2}$ for $1 \leq i \leq n$. Here $x_0 = 0$ and $x_{n+1} = \infty$.

In fact the above considerations lead to a characterization of energy functions for exponential GREM. Suppose f is a continuous function on $[0, \infty)$ with $f(0) = \log 2$. Further suppose that there are finitely many points $0 < x_1 < \cdots < x_n$ where the left and right derivatives are unequal and f' is a constant, say, c_i in (x_i, x_{i+1}). Here $x_0 = 0$ and $x_{n+1} = \infty$. Then f is the energy function for some exponential GREM iff

$$0 = c_0 < c_1 < \cdots < c_n \quad \text{and} \quad \sum_{i=1}^{n} x_i(c_i - c_{i-1}) = \log 2. \tag{2.6.15}$$

2.6.3 $0 < \gamma < 1$

Now we come to the case $\gamma < 1$. Unlike in the above two subsections, here we have not been able to derive the closed form expression of the free energy for general n level trees. For $\gamma < 1$, the function $\sum_{i=1}^{n} \left(\frac{x_i^\gamma}{\gamma} - \beta a_i x_i \right)$ is not a convex function, rather a concave function. Moreover the domain $\Psi = \{\widetilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} |x_i|^\gamma \leq \sum_{i=1}^{k} \gamma p_i \log 2, \quad 1 \leq k \leq n\}$ is also a non-convex set. Hence in order to calculate

$$\inf_{\tilde{x} \in \Psi^+} \sum_{i=1}^{n} \left(\frac{x_i^\gamma}{\gamma} - \beta a_i x_i \right) \tag{2.6.16}$$

with

$$\Psi^+ = \{\widetilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} |x_i|^\gamma \leq \sum_{i=1}^{k} \gamma p_i \log 2, \,\&\, x_k \geq 0 \text{ for } 1 \leq k \leq n\}, \tag{2.6.17}$$

we can not use the convex analysis, as we did for the case $\gamma > 1$. However, by change of variables, the problem can be brought back to optimizing a convex function over a convex set. Now we specialize to the case $n = 2$. We shall calculate

$$\inf_{\tilde{x}\in\Psi^+} \left\{ (\frac{1}{\gamma}x_1^\gamma - \beta a_1 x_1) + (\frac{1}{\gamma}x_2^\gamma - \beta a_2 x_2) \right\} \tag{2.6.18}$$

with

$$\Psi^+ = \{(x_1, x_2) \geq 0 : \, x_1^\gamma \leq \gamma p_1 \log 2, x_1^\gamma + x_2^\gamma \leq \gamma(p_1 + p_2)\log 2\}. \tag{2.6.19}$$

To do so, we transform the problem by denoting $\frac{1}{\gamma}x_1^\gamma = x$, $\frac{1}{\gamma}x_2^\gamma = y$, $a_1\gamma^{\frac{1}{\gamma}} = a$, $a_2\gamma^{\frac{1}{\gamma}} = b$, $p_1 \log 2 = c$, $p_2 \log 2 = d$ and $\alpha = \frac{1}{\gamma}$ so that, $\alpha > 1$ and we need to calculate

$$-\sup_{\tilde{x}\in\Psi^+} \{(\beta a x^\alpha - x) + (\beta b y^\alpha - y)\} \tag{2.6.20}$$

with

$$\Psi^+ = \{(x, y) \geq 0 : \, x \leq c, x + y \leq c + d\}. \tag{2.6.21}$$

Let $f(x, y) = (\beta a x^\alpha - x) + (\beta b y^\alpha - y)$. Since $f(x, y)$ is a convex function and we are looking for supremum over a convex set, the supremum occurs at the boundary points. Note that Ψ^+ is a polygon. Where as for any $\epsilon \in \mathbb{R}$, the set $\{f(x, y) = \epsilon\}$ is either empty set or a smooth curve. Hence the above supremum occurs at one of the corner points, $A \equiv (0, 0)$, $B \equiv (c, 0)$, $C \equiv (c, d)$ and $D \equiv (0, c + d)$, of Ψ^+. Now

$$f(A) \gtreqless f(B) \quad \text{iff } \beta \lesseqgtr \frac{1}{ac^{\alpha-1}}, \tag{2.6.22}$$

$$f(A) \gtreqless f(C) \quad \text{iff } \beta \lesseqgtr \frac{c+d}{ac^\alpha + bd^\alpha}, \tag{2.6.23}$$

$$f(A) \gtreqless f(D) \quad \text{iff } \beta \lesseqgtr \frac{1}{b(c+d)^{\alpha-1}}, \tag{2.6.24}$$

$$f(B) \gtreqless f(C) \quad \text{iff } \beta \lesseqgtr \frac{1}{bd^{\alpha-1}}, \tag{2.6.25}$$

$$f(B) \gtreqless f(D) \quad \text{iff } b(c+d)^\alpha > ac^\alpha \text{ and } \beta \lesseqgtr \frac{d}{b(c+d)^\alpha - ac^\alpha}, \tag{2.6.26}$$

$$f(B) > f(D) \quad \text{if } b(c+d)^\alpha \leq ac^\alpha, \tag{2.6.27}$$

$$f(C) \gtreqless f(D) \quad \text{iff } ac^\alpha + bd^\alpha \gtreqless b(c+d)^\alpha. \tag{2.6.28}$$

Note that the last two relations do not depend on β. Now comparing all the possibilities, we obtain the following three scenarios:

$b(c+d)^\alpha \le ac^\alpha + bd^\alpha$

Let us assume $b(c+d)^\alpha \le ac^\alpha + bd^\alpha$. Then it is easy to see that

$$b(c+d)^\alpha \le ac^\alpha + bd^\alpha \Rightarrow \begin{cases} f(C) \ge f(D), \\ \frac{c+d}{ac^\alpha+bd^\alpha} \le \frac{1}{b(c+d)^{\alpha-1}} < \frac{1}{bd^{\alpha-1}}. \end{cases}$$

Now

$$\frac{c+d}{ac^\alpha + bd^\alpha} < \frac{1}{bd^{\alpha-1}} \Rightarrow \frac{1}{ac^{\alpha-1}} < \frac{1}{bd^{\alpha-1}},$$

and

$$\frac{1}{ac^{\alpha-1}} \lesseqqgtr \frac{c+d}{ac^\alpha + bd^\alpha} \Leftrightarrow \frac{1}{ac^{\alpha-1}} \lesseqqgtr \frac{1}{bd^{\alpha-1}}$$

implies

$$\frac{1}{ac^{\alpha-1}} < \frac{c+d}{ac^\alpha + bd^\alpha} \le \frac{1}{b(c+d)^{\alpha-1}} < \frac{1}{bd^{\alpha-1}}.$$

Hence we get

$$\sup_{\Psi^+} f(x,y) = \begin{cases} f(A) & \text{if } 0 \le \beta \le \frac{1}{ac^{\alpha-1}} \\ f(B) & \text{if } \frac{1}{ac^{\alpha-1}} \le \beta \le \frac{1}{bd^{\alpha-1}} \\ f(C) & \text{if } \beta \ge \frac{1}{bd^{\alpha-1}}. \end{cases} \tag{2.6.29}$$

$b(c+d)^\alpha > ac^\alpha + bd^\alpha$ & $b(c+d)^{\alpha-1} \le ac^{\alpha-1}$

Let us assume $b(c+d)^\alpha > ac^\alpha + bd^\alpha$ & $b(c+d)^{\alpha-1} \le ac^{\alpha-1}$. Then it is easy to see that

$$b(c+d)^\alpha > ac^\alpha + bd^\alpha \Rightarrow \begin{cases} f(D) > f(C), \\ \frac{1}{b(c+d)^{\alpha-1}} < \frac{c+d}{ac^\alpha+bd^\alpha}, \\ \frac{d}{b(c+d)^\alpha - ac^\alpha} < \frac{1}{bd^{\alpha-1}}. \end{cases}$$

Moreover,

$$b(c+d)^{\alpha-1} \le ac^{\alpha-1} \Rightarrow \frac{1}{ac^{\alpha-1}} \le \frac{1}{b(c+d)^{\alpha-1}} \le \frac{d}{b(c+d)^\alpha - ac^\alpha}.$$

Thus

$$\frac{1}{ac^{\alpha-1}} \le \frac{1}{b(c+d)^{\alpha-1}} \le \frac{d}{b(c+d)^\alpha - ac^\alpha} < \frac{1}{bd^{\alpha-1}}.$$

Hence we get

$$\sup_{\Psi^+} f(x,y) = \begin{cases} f(A) & \text{if } 0 \le \beta \le \frac{1}{ac^{\alpha-1}} \\ f(B) & \text{if } \frac{1}{ac^{\alpha-1}} \le \beta \le \frac{d}{b(c+d)^\alpha - ac^\alpha} \\ f(D) & \text{if } \beta \ge \frac{d}{b(c+d)^\alpha - ac^\alpha}. \end{cases} \tag{2.6.30}$$

$b(c+d)^\alpha > ac^\alpha + bd^\alpha$ & $b(c+d)^{\alpha-1} > ac^{\alpha-1}$

Let us assume $b(c+d)^\alpha > ac^\alpha + bd^\alpha$ & $b(c+d)^{\alpha-1} > ac^{\alpha-1}$. Then it is easy to see that

$$b(c+d)^\alpha > ac^\alpha + bd^\alpha \Rightarrow \begin{cases} f(D) > f(C), \\ \frac{1}{b(c+d)^{\alpha-1}} < \frac{c+d}{ac^\alpha + bd^\alpha}. \end{cases}$$

Moreover,

$$b(c+d)^{\alpha-1} > ac^{\alpha-1} \Rightarrow \frac{1}{ac^{\alpha-1}} > \frac{1}{b(c+d)^{\alpha-1}} > \frac{d}{b(c+d)^\alpha - ac^\alpha}.$$

Thus

$$\frac{d}{b(c+d)^\alpha - ac^\alpha} < \frac{1}{b(c+d)^{\alpha-1}} < \frac{c+d}{ac^\alpha + bd^\alpha}.$$

Hence we get

$$\sup_{\Psi^+} f(x,y) = \begin{cases} f(A) & \text{if } 0 \le \beta \le \frac{1}{b(c+d)^{\alpha-1}} \\ f(D) & \text{if } \beta \ge \frac{1}{b(c+d)^{\alpha-1}}. \end{cases} \tag{2.6.31}$$

We can conclude the above three cases in the following:

Theorem 2.6.6. *For two level GREM with driving distribution having density $\phi_{N,\gamma}$ as defined in (2.6.1) with $0 < \gamma < 1$, we have, almost surely,*

1. *if $a_2 \le a_1(p_1 \log 2)^{\frac{1}{\gamma}} + a_2(p_2 \log 2)^{\frac{1}{\gamma}}$, then the limiting free energy is*

$$\begin{cases} \log 2 & \text{for } 0 \le \beta \le \frac{1}{a_1 \gamma^{\frac{1}{\gamma}} (p_1 \log 2)^{\frac{1-\gamma}{\gamma}}}, \\ p_2 \log 2 + \beta a_1 (\gamma p_1 \log 2)^{\frac{1}{\gamma}} & \text{for } \frac{1}{a_1 \gamma^{\frac{1}{\gamma}} (p_1 \log 2)^{\frac{1-\gamma}{\gamma}}} \le \beta \le \frac{1}{a_2 \gamma^{\frac{1}{\gamma}} (p_2 \log 2)^{\frac{1-\gamma}{\gamma}}}, \\ \beta(a_1(\gamma p_1 \log 2)^{\frac{1}{\gamma}} + a_2(\gamma p_2 \log 2)^{\frac{1}{\gamma}}) & \text{for } \beta \ge \frac{1}{a_2 \gamma^{\frac{1}{\gamma}} (p_2 \log 2)^{\frac{1-\gamma}{\gamma}}}. \end{cases}$$

2. *if $a_2 > a_1(p_1 \log 2)^{\frac{1}{\gamma}} + a_2(p_2 \log 2)^{\frac{1}{\gamma}}$ and $a_2 \le a_1(p_1 \log 2)^{\frac{1-\gamma}{\gamma}}$, then the limiting free energy is*

$$\begin{cases} \log 2 & \textit{for } 0 \le \beta \le \frac{1}{a_1\gamma^{\frac{1}{\gamma}}(p_1\log 2)^{\frac{1-\gamma}{\gamma}}}, \\ p_2\log 2 + \beta a_1(\gamma p_1 \log 2)^{\frac{1}{\gamma}} & \textit{for } \frac{1}{a_1\gamma^{\frac{1}{\gamma}}(p_1\log 2)^{\frac{1-\gamma}{\gamma}}} \le \beta \le \frac{p_2\log 2}{a_2(\gamma\log 2)^{\frac{1}{\gamma}} - a_1(\gamma p_1\log 2)^{\frac{1}{\gamma}}}, \\ \beta a_2(\gamma\log 2)^{\frac{1}{\gamma}}) & \textit{for } \beta \ge \frac{p_2\log 2}{a_2(\gamma\log 2)^{\frac{1}{\gamma}} - a_1(\gamma p_1\log 2)^{\frac{1}{\gamma}}}. \end{cases}$$

3. *if* $a_2 > a_1(p_1\log 2)^{\frac{1}{\gamma}} + a_2(p_2\log 2)^{\frac{1}{\gamma}}$ *and* $a_2 > a_1(p_1\log 2)^{\frac{1-\gamma}{\gamma}}$*, then the limiting free energy is*

$$\begin{cases} \log 2 & \textit{for } 0 \le \beta \le \frac{1}{a_2\gamma^{\frac{1}{\gamma}}(\log 2)^{\frac{1-\gamma}{\gamma}}}, \\ \beta a_2(\gamma\log 2)^{\frac{1}{\gamma}}) & \textit{for } \beta \ge \frac{1}{a_2\gamma^{\frac{1}{\gamma}}(\log 2)^{\frac{1-\gamma}{\gamma}}}. \end{cases}$$

Remark 2.6.7. Note that in a 2 level double exponential GREM (in the earlier subsection) with weights a_1 and a_2, we had at most two cases, namely, $\frac{1}{a_1} \le \frac{1}{a_2}$ and $\frac{1}{a_1} > \frac{1}{a_2}$. Where as for $\gamma < 1$, we have three cases.

Though there are three cases, we can think of them as two cases like the double exponential GREM, namely, $\frac{1}{a_1(p_1\log 2)^{\frac{1-\gamma}{\gamma}}} \le \frac{1}{a_2(\log 2)^{\frac{1-\gamma}{\gamma}}}$ and $\frac{1}{a_1(p_1\log 2)^{\frac{1-\gamma}{\gamma}}} > \frac{1}{a_2(\log 2)^{\frac{1-\gamma}{\gamma}}}$, where the first case has two more subcases, namely, $\frac{1}{a_2(p_2\log 2)^{\frac{1-\gamma}{\gamma}}} \le \frac{p_2\log 2}{a_2(\log 2)^{\frac{1}{\gamma}} - a_1(p_1\log 2)^{\frac{1}{\gamma}}}$ and $\frac{1}{a_2(p_2\log 2)^{\frac{1-\gamma}{\gamma}}} > \frac{p_2\log 2}{a_2(\log 2)^{\frac{1}{\gamma}} - a_1(p_1\log 2)^{\frac{1}{\gamma}}}$.

2.7 Level-dependant Distributions

We already mentioned that the LDP holds good even when the driving distributions at various levels are different. To be precise, fix numbers $\gamma_1, \cdots, \gamma_n$; each greater than 0. Consider an n level GREM with the driving distribution at the i-th level being ϕ_{N,γ_i} given by (2.6.1). That is, at the first level for each edge σ_1 the associated random variable $\xi(\sigma_1)$ has density ϕ_{N,γ_1}. In general, for any edge $\sigma_1\cdots\sigma_i$ at the i-th level, the associated random variables $\xi(\sigma_1\cdots\sigma_i)$ has density ϕ_{N,γ_i}. Then the map $\Sigma_N \to \mathbb{R}^n$ by

$$\sigma \mapsto \left(\frac{\xi(\sigma_1,\omega)}{N}, \frac{\xi(\sigma_1\sigma_2,\omega)}{N}, \cdots, \frac{\xi(\sigma_1\cdots\sigma_n,\omega)}{N}\right)$$

induce random probability $\mu_N(\omega)$ on $\mathbb{R}^n$ by transporting the uniform probability on Σ_N. Theorem 2.4.6 suggest that in this case the free energy of the system will be

$$\lim_N \frac{1}{N}\log Z_N(\beta) = \log 2 - \inf_{\widetilde{x}\in\Psi^+}\left\{\sum_{i=1}^n \left(\frac{x_i^{\gamma_i}}{\gamma_i} - \beta a_i x_i\right)\right\}, \tag{2.7.1}$$

where Ψ^+ is the intersection of

$$\Psi = \{\widetilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} \frac{|x_i|^{\gamma_i}}{\gamma_i} \leq \sum_{i=1}^{k} p_i \log 2, \;\; 1 \leq k \leq n\},$$

with the positive orthant of $\mathbb{R}^n$ and $p_i = \lim\limits_{N\to\infty} \frac{k(i,N)}{N}$.

In its generality, it is very difficult to have a closed form expression for the above infimum. May be there is no general closed form expression, for the infimum and hence for the free energy of the system. To make a beginning and to see what one can expect, we now specialize to the case $n = 2$. The limiting frequencies $\lim\limits_{N} \frac{k(i,N)}{N}$ are p_i for $i = 1, 2$. The weights for the two level are a_1 and a_2 respectively. We assume p_1, p_2, a_1, a_2 are strictly positive.

2.7.1 Exponential - Gaussian GREM

In this case we consider the distributions at the first level to be $\phi_{N,1}$ and at the second level to be $\phi_{N,2}$ — that is, exponential and Gaussian respectively. So from (2.7.1), the expression for the free energy for this case will read as follows:

$$\begin{aligned}\mathcal{E}(\beta) &= \lim_{N} \frac{1}{N} \log Z_N(\beta) \\ &= \log 2 - \inf\{f(x,y) : x, y \geq 0;\; x \leq p_1 \log 2;\; x + \frac{1}{2}y^2 \leq \log 2\} \qquad (2.7.2)\end{aligned}$$

where

$$f(x,y) = x(1 - \beta a_1) + \frac{1}{2}y^2 - \beta a_2 y. \qquad (2.7.3)$$

To calculate $\mathcal{E}(\beta)$ explicitly we proceed as follows. First we discuss the case $\beta \leq \frac{1}{a_1}$. Then we discuss $\beta > \frac{1}{a_1}$. This later case leads to three subcases. In each subcase combining the conclusion along with the case $\beta \leq \frac{1}{a_1}$, we give a full picture of $\mathcal{E}(\beta)$ for all values of β.

I. $\boldsymbol{\beta \leq \frac{1}{a_1}}$

On the interval $[0, \infty)$, the function $\frac{1}{2}y^2 - \beta a_2 y$ decreases up to βa_2 and then increases. So when $\beta \leq \frac{1}{a_1}$, that is when $1 - \beta a_1 \geq 0$, the above function attains its minimum at the point, $(0, \beta a_2 \wedge \sqrt{2 \log 2})$.

II. $\boldsymbol{\beta > \frac{1}{a_1}}$

If $\beta > \frac{1}{a_1}$, here is how to calculate the infimum. The function $g(y) = \inf_x f(x,y)$ is given by

$$g(y) = \begin{cases} p_1(1-\beta a_1)\log 2 + \frac{1}{2}y^2 - \beta a_2 y & \text{for } 0 \le y \le \sqrt{2p_2 \log 2} \\ (1-\beta a_1)\log 2 + \frac{1}{2}\beta a_1 y^2 - \beta a_2 y & \text{for } \sqrt{2p_2\log 2} \le y \le \sqrt{2\log 2}. \end{cases}$$

This is because, when $0 \le y \le \sqrt{2p_2\log 2}$, $\inf_x f(x,y)$ is attained at $x = p_1 \log 2$, whereas in the other case the infimum is attained at $x = \log 2 - \frac{1}{2}y^2$.

Since the required infimum of f is just the infimum of $g(y)$, one has to calculate $\inf_{0\le y\le\sqrt{2\log 2}} g(y)$ by analyzing g in the two intervals separately. This is what we do below. First note that the function g is continuous. Now we have the following three scenarios.

A1: $\boldsymbol{\frac{a_2}{a_1} < \sqrt{2p_2\log 2}}$

Let us assume $\frac{a_2}{a_1} < \sqrt{2p_2\log 2}$. First let us consider β such that $\frac{1}{a_1} < \beta \le \frac{\sqrt{2p_2\log 2}}{a_2}$. In particular, $\beta a_2 \le \sqrt{2p_2\log 2}$ where as $\frac{a_2}{a_1} < \sqrt{2p_2\log 2}$. So the function $\frac{1}{2}y^2 - \beta a_2 y$ is decreasing up to βa_2 in $[0, \sqrt{2p_2\log 2}]$ and then increasing. Thus in $[0,\sqrt{2p_2\log 2}]$, g attains its minimum at βa_2. On $[\sqrt{2p_2\log 2}, \sqrt{2\log 2}]$ the function $\frac{1}{2}\beta a_1 y^2 - \beta a_2 y = \beta a_1(\frac{1}{2}y^2 - \frac{a_2}{a_1}y)$ is increasing. Hence, g being continuous, for the values of β under consideration, the infimum will occur at $y = \beta a_2$.

Now β be such that $\beta > \frac{\sqrt{2p_2\log 2}}{a_2}$ so that $\beta a_2 > \sqrt{2p_2\log 2}$. Since $\frac{a_2}{a_1} < \sqrt{2p_2\log 2}$ the function $\frac{1}{2}\beta a_1 y^2 - \beta a_2 y$ is increasing on $[\sqrt{2p_2\log 2}, \sqrt{2\log 2}]$. The function $\frac{1}{2}y^2 - \beta a_2 y$ is decreasing on $[0,\sqrt{2p_2\log 2}]$ attaining infimum at $y = \sqrt{2p_2\log 2}$. As a consequence, for $\beta > \frac{\sqrt{2p_2\log 2}}{a_2}$, the infimum of $g(y)$ is occurs at $\sqrt{2p_2\log 2}$.

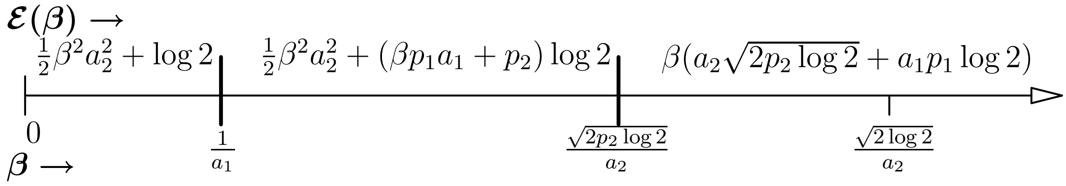
Thus combining **I.** and above para we conclude that if $\frac{a_2}{a_1} < \sqrt{2p_2\log 2}$ then phase transitions take place at $\beta = \frac{1}{a_1}$ and $\beta = \frac{\sqrt{2p_2\log 2}}{a_2}$. So, substituting this corresponding arguments where minimum is attained in (2.7.3), we have the following

Theorem 2.7.1. *In the Exponential-Gaussian GREM, if $\frac{a_2}{a_1} < \sqrt{2p_2\log 2}$ then almost surely,*

$$\lim_{N\to\infty}\frac{1}{N}\log Z_N(\beta) = \begin{cases} \log 2 + \frac{1}{2}\beta^2 a_2^2 & \text{if } \beta \le \frac{1}{a_1} \\ p_2\log 2 + \frac{1}{2}\beta^2 a_2^2 + \beta p_1 a_1 \log 2 & \text{if } \frac{1}{a_1} < \beta \le \frac{\sqrt{2p_2\log 2}}{a_2} \\ \beta(a_2\sqrt{2p_2\log 2} + a_1 p_1 \log 2) & \text{if } \beta > \frac{\sqrt{2p_2\log 2}}{a_2}. \end{cases}$$

We can picture the value of $\mathcal{E}(\beta)$ against β as given below. The values of β are given under the line and values of $\mathcal{E}(\beta)$ are given above the line. The phase transitions occur at the dark lines.

Subcase A1



This case seems rather peculiar. This is indeed a sum of two REMs, as follows. Imagine placing exponential random variables ξ_{σ_1} at the first level and one i.i.d bunch $\{\xi_{\sigma_1\sigma_2}\}$ is placed below each first level node. In other words, consider $\{\eta_{\sigma_2} : \sigma_2 \in 2^{k(2,N)}\}$ i.i.d $\mathcal{N}(0, N)$ and set $\xi_{\sigma_1\sigma_2} = \eta_{\sigma_2}$ for all σ_1, σ_2. Consider the corresponding Hamiltonian $H_N(\sigma) = a_1\xi_{\sigma_1} + a_2\xi_{\sigma_1\sigma_2} = a_1\xi_{\sigma_1} + a_2\eta_{\sigma_2}$. Let us set $Z_N^1 = \sum\limits_{\sigma_1} e^{\beta a_1 \xi_{\sigma_1}}$, the partition function for the $k(1, N)$-particles system consisting of exponential Hamiltonian with weight a_1. Let $Z_N^2 = \sum\limits_{\sigma_2} e^{a_2\eta_{\sigma_2}}$, the partition function for $k(2, N)$ particle system consisting of Gaussian, $\mathcal{N}(0, N)$ Hamiltonian with weight a_2. Clearly, $Z_N = Z_N^1 \cdot Z_N^2$. If, for $i = 1, 2$; $\mathcal{E}_i = \lim\limits_N \frac{1}{N}\log Z_N^i$ then the exponential REM formula [28, 29] yields, a.s.,

$$\mathcal{E}_1(\beta) = \begin{cases} p_1 \log 2 & \text{if } \beta \leq \frac{1}{a_1} \\ \beta p_1 a_1 \log 2 & \text{if } \beta > \frac{1}{a_1}. \end{cases} \tag{2.7.4}$$

The Gaussian REM formula (keeping in mind that for N fixed, the $k(2, N)$ particle system has $\mathcal{N}(0, N)$ Hamiltonians as opposed to $\mathcal{N}(0, k(2, N))$ yields, a.s,

$$\mathcal{E}_2(\beta) = \begin{cases} p_2 \log 2 + \frac{1}{2}a_2^2\beta^2 & \text{if } \beta \leq \frac{\sqrt{2p_2\log 2}}{a_2} \\ \beta a_2\sqrt{2p_2\log 2} & \text{if } \beta > \frac{\sqrt{2p_2\log 2}}{a_2}. \end{cases} \tag{2.7.5}$$

One can now verify that, a.s.

$$\mathcal{E}(\beta) = \mathcal{E}_1(\beta) + \mathcal{E}_2(\beta).$$

In other words the GREM behaves like sum of two independent REMs, one exponential and other Gaussian. The word independent is used here in the sense that there is no interaction between these two REMs – that is, there is no interaction between the $k(1, N)$ particles and the $k(2, N)$ particles, as if there is a barrier between these two sets of particles. Of course, this is so as long as $\frac{a_2}{a_1} < \sqrt{2p_2 \log 2}$.

A2: $\sqrt{2p_2 \log 2} \leq \frac{a_2}{a_1} < \sqrt{2 \log 2}$

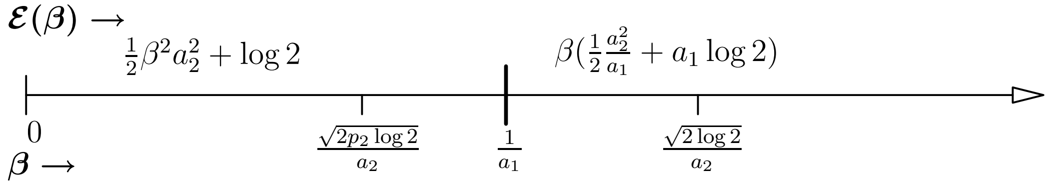
Let us assume $\sqrt{2p_2 \log 2} \leq \frac{a_2}{a_1} < \sqrt{2 \log 2}$. Then $\beta > \frac{1}{a_1}$ means $\beta a_2 > \frac{a_2}{a_1} \geq \sqrt{2p_2 \log 2}$ where as $\frac{a_2}{a_1} < \sqrt{2 \log 2}$. So the function $\frac{1}{2}y^2 - \beta a_2 y$ is decreasing on $[0, \sqrt{2p_2 \log 2}]$ and the other function $\frac{1}{2}\beta a_1 y^2 - \beta a_2 y = \beta a_1(\frac{1}{2}y^2 - \frac{a_2}{a_1}y)$ is decreasing up to $\frac{a_2}{a_1}$ in $[\sqrt{2p_2 \log 2}, \leq \sqrt{2 \log 2}]$ and then increasing. Hence, as g is continuous, the infimum will occur at $y = \frac{a_2}{a_1}$. Thus, the phase transition takes place at $\beta = \frac{1}{a_1}$. So we have the following

Theorem 2.7.2. *In the Exponential-Gaussian GREM, if $\sqrt{2p_2 \log 2} \leq \frac{a_2}{a_1} < \sqrt{2 \log 2}$ then almost surely,*

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \begin{cases} \log 2 + \frac{1}{2}\beta^2 a_2^2 & \text{if } \beta \leq \frac{1}{a_1} \\ \beta\left(\frac{1}{2}\frac{a_2^2}{a_1} + a_1 \log 2\right) & \text{if } \beta > \frac{1}{a_1} \end{cases}$$

As earlier, we can picture the value of $\mathcal{E}(\beta)$ against β as given below. The values of β are given under the line and values of $\mathcal{E}(\beta)$ are given above the line. The phase transitions occur at the dark lines.

Subcase A2



In this case, we observe that the free energy for inverse temperature up to $\frac{1}{a_1}$ is given by $\log 2 + \frac{1}{2}\beta^2 a_2^2$. This can be thought of as the Gaussian REM energy but not going all the way up to $\beta \leq \frac{\sqrt{2 \log 2}}{a_2}$ but cut short at $\frac{1}{a_1}$. This can also be thought of as the sum of the two energies $\mathcal{E}_1$ and $\mathcal{E}_2$ as in (2.7.4) and (2.7.5), but then the Gaussian effect is prolonged up to $\beta \leq \frac{1}{a_1}$ instead of stopping at $\frac{\sqrt{2p_2 \log 2}}{a_2}$. We do not know which is the correct interpretation. For $\beta > \frac{1}{a_1}$, the system exhibits a new phenomenon which we are unable to explain. The term $\beta a_1 \log 2$ is reminiscent of the exponential REM energy. The other term $\frac{1}{2}\beta\frac{a_2^2}{a_1}$ appears to be new.

A3: $\sqrt{2 \log 2} \leq \frac{a_2}{a_1}$

Let us assume $\sqrt{2 \log 2} \leq \frac{a_2}{a_1}$. Then $\beta > \frac{1}{a_1}$ means $\beta a_2 > \frac{a_2}{a_1} \leq \sqrt{2 \log 2}$. So both the functions $\frac{1}{2}y^2 - \beta a_2 y$ and $\frac{1}{2}\beta a_1 y^2 - \beta a_2 y = \beta a_1(\frac{1}{2}y^2 - \frac{a_2}{a_1}y)$ are decreasing on

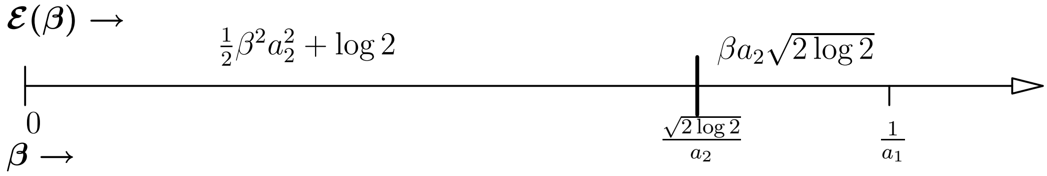
$[0, \sqrt{2p_2 \log 2}]$ and $[\sqrt{2p_2 \log 2}, \leq \sqrt{2 \log 2}]$ respectively. Hence the infimum will occur at $y = \sqrt{2 \log 2}$. Being $\beta a_2 > \sqrt{2 \log 2}$, the phase transition takes place at $\beta = \frac{\sqrt{2 \log 2}}{a_2}$. Hence we have the following

Theorem 2.7.3. *In the Exponential-Gaussian GREM, if* $\frac{a_2}{a_1} \geq \sqrt{2 \log 2}$ *then almost surely,*

$$\lim_{N \to \infty} \frac{1}{N} \log Z_N(\beta) = \begin{cases} \log 2 + \frac{1}{2}\beta^2 a_2^2 & \text{if } \beta \leq \frac{\sqrt{2 \log 2}}{a_2} \\ \beta a_2 \sqrt{2 \log 2} & \text{if } \beta > \frac{\sqrt{2 \log 2}}{a_2} \end{cases}$$

As earlier, we can picture the value of $\mathcal{E}(\beta)$ against β as given below. The values of β are given under the line and values of $\mathcal{E}(\beta)$ are given above the line. The phase transitions occur at the dark lines.

Subcase A3



Thus in subcase A3, the system behaves like a REM with Gaussian distributions [16] having weight a_2, that is, as if $H_N(\sigma)$ are i.i.d centered Gaussian with variance $a_2^2 N$. For example, when $a_1 = a_2$ then this is just the standard Gaussian REM. It does not depend on the quantities p_1 and p_2. Even when $p_2 = 0.0001$ (very small) the first level exponentials do not show up in the limit. Further the GREM reduces to a REM. Of course, this is so as long as $\sqrt{2 \log 2} < \frac{a_2}{a_1}$. This should be contrasted with subcase A1 where the entire system behaves like sum of two independent REM, one Gaussian and other exponential.

2.7.2 Gaussian - Exponential GREM

Let us consider the situation where the driving distributions at the first level are Gaussian, $\phi_{N,2}$ and at the second level they are exponential, $\phi_{N,1}$. Moreover, as earlier a_1 and a_2 are the weights at the first and second level respectively. We will use the same notation for $k(1, N), k(2, N)$ and for p_1, p_2. In this case the general formula of Theorem 2.4.6 reduces to the following:

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf\{\widetilde{f}(x, y) : x, y \geq 0;\ x \leq \sqrt{2p_1 \log 2};\ \frac{1}{2}x^2 + y \leq \log 2\}$$

almost surely, where

$$\widetilde{f}(x,y) = \frac{1}{2}x^2 - \beta a_1 x + y(1-\beta a_2).$$

In this case to calculate the infimum we proceed as follows. Put $g(x) = \inf_y \widetilde{f}(x,y)$. Since $(1\beta a_2) \geq 0$ for $\beta \leq \frac{1}{a_2}$, we have

$$\begin{aligned} g(x) &= \frac{1}{2}x^2 - \beta a_1 x && \text{if } \beta \leq \frac{1}{a_2} \\ &= \frac{1}{2}x^2 - \beta a_1 x + (1-\beta a_2)(\log 2 - \frac{1}{2}x^2) && \text{if } \beta > \frac{1}{a_2} \end{aligned}$$

that is

$$\begin{aligned} g(x) &= \frac{1}{2}x^2 - \beta a_1 x && \text{if } \beta \leq \frac{1}{a_2} \\ &= \beta a_2 \left(\frac{1}{2}x^2 - \frac{a_1}{a_2}x\right) + (1-\beta a_2)\log 2 && \text{if } \beta > \frac{1}{a_2}. \end{aligned}$$

Since infimum of $f(x,y)$ is same as that of infimum of g over x, one has to calculate $\inf_{0 \leq x \leq \sqrt{2p_1 \log 2}} g(x)$. Here, we will have the following two scenarios.

B1: $\frac{a_1}{a_2} \leq \sqrt{2p_1 \log 2}$

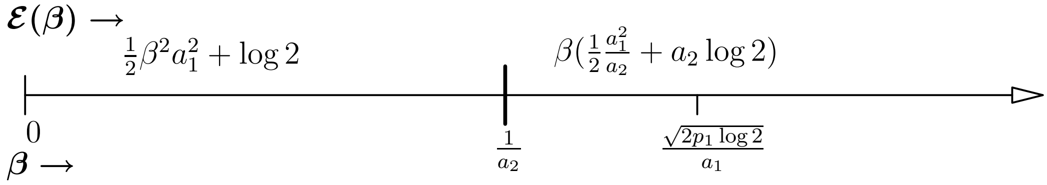
Let us assume $\frac{a_1}{a_2} \leq \sqrt{2p_1 \log 2}$. If $\beta \leq \frac{1}{a_2}$, then $\beta a_1 \leq \sqrt{2p_1 \log 2}$. The function $\frac{1}{2}x^2 - \beta a_1 x$ decreases up to βa_1 an then increases. Hence when $\beta \leq \frac{1}{a_2}$ the infimum occurs at $x = \beta a_1$. For $\beta > \frac{1}{a_2}$ as $\frac{a_1}{a_2} \leq \sqrt{2p_1 \log 2}$, the infimum will occur at $x = \frac{a_1}{a_2}$. So we have the following

Theorem 2.7.4. *In the Gaussian-Exponential GREM, if $\frac{a_1}{a_2} \leq \sqrt{2p_1 \log 2}$ then almost surely,*

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \begin{cases} \log 2 + \frac{1}{2}\beta^2 a_1^2 & \text{if } \beta \leq \frac{1}{a_2} \\ \beta\left(\frac{1}{2}\frac{a_1^2}{a_2} + a_2 \log 2\right) & \text{if } \beta > \frac{1}{a_2} \end{cases}$$

As earlier, we can picture the value of $\mathcal{E}(\beta)$ against β as given below. The values of β are given under the line and values of $\mathcal{E}(\beta)$ are given above the line. The phase transitions occur at the dark lines.

Subcase B1



B2: $\sqrt{2p_1 \log 2} < \frac{a_1}{a_2}$

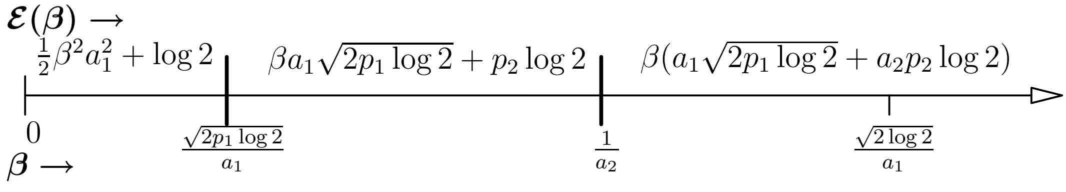
Let us assume $\frac{a_1}{a_2} > \sqrt{2p_1 \log 2}$. If $\beta \leq \frac{1}{a_2}$ we have $\beta a_1 \leq \frac{a_1}{a_2}$. So the quantity βa_1 will be in $[0, \sqrt{2p_1 \log 2}]$ as long as $\beta \leq \frac{\sqrt{2p_1 \log 2}}{a_1}$. As the function $\frac{1}{2}x^2 - \beta a_1 x$ decreases up to βa_1 an then increases, for $\beta \leq \frac{\sqrt{2p_1 \log 2}}{a_1}$ the infimum occurs at $x = \beta a_1$. But for and for $\frac{\sqrt{2p_1 \log 2}}{a_1} < \beta \leq \frac{1}{a_2}$, the infimum occurs at $x = \sqrt{2p_1 \log 2}$. For $\beta > \frac{1}{a_2}$ consider the function $\frac{1}{2}x^2 - \frac{a_1}{a_2}x$ which decreases up to $\frac{a_1}{a_2}$ and then increases. As $\frac{a_1}{a_2} > \sqrt{2p_1 \log 2}$, the infimum will occur at $x = \sqrt{2p_1 \log 2}$. Thus we have the following

Theorem 2.7.5. *In the Gaussian-Exponential GREM, if $\frac{a_1}{a_2} > \sqrt{2p_1 \log 2}$ then almost surely,*

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) = \begin{cases} \log 2 + \frac{1}{2}\beta^2 a_1^2 & \text{if } \beta \leq \frac{\sqrt{2p_1 \log 2}}{a_2} \\ p_2 \log 2 + \beta a_1 \sqrt{2p_1 \log 2} & \text{if } \frac{\sqrt{2p_1 \log 2}}{a_2} < \beta \leq \frac{1}{a_2} \\ \beta \left(a_1 \sqrt{2p_1 \log 2} + a_2 p_2 \log 2 \right) & \text{if } \beta > \frac{1}{a_2} \end{cases}$$

As earlier, we can picture the value of $\mathcal{E}(\beta)$ against β as given below. The values of β are given under the line and values of $\mathcal{E}(\beta)$ are given above the line. The phase transitions occur at the dark lines.

Subcase B2



Remarks similar to Exponential-Gaussian GREM apply here as well. Subcase B1 is similar to subcase A2. Here also the term $\frac{1}{2}\beta\frac{a_1^2}{a_2}$ is not reminiscent of anything we know.

Subcase B2 is similar to that of subcase A1. That is in subcase B2, the limiting free energy is sum of two REM free energies – one is of Gaussian REM and other is of exponential REM. To be precise, the Gaussian REM limiting free energy (keeping in mind that for N fixed, the $k(1, N)$ particle system has $\mathcal{N}(0, a^2 N)$ Hamiltonian as opposed to $\mathcal{N}(0, a^2 k(1,N))$ yields, a.s,

$$\widetilde{\mathcal{E}}_1(\beta) = \begin{cases} p_1 \log 2 + \frac{1}{2} a_1^2 \beta^2 & \text{if } \beta \le \frac{\sqrt{2p_1 \log 2}}{a_1} \\ \beta a_1 \sqrt{2p_1 \log 2} & \text{if } \beta > \frac{\sqrt{2p_1 \log 2}}{a_1}. \end{cases} \tag{2.7.6}$$

On the other hand, for fixed N, if we have configurations $2^{k(1,N)}$ then the exponential REM limiting free energy, with Hamiltonian as a_2 times double exponential random variable yields, a.s.,

$$\widetilde{\mathcal{E}}_2(\beta) = \begin{cases} p_2 \log 2 & \text{if } \beta \le \frac{1}{a_2} \\ \beta p_2 a_2 \log 2 & \text{if } \beta > \frac{1}{a_2}. \end{cases} \tag{2.7.7}$$

Now it is easy verify that, in subcase B2, a.s.

$$\mathcal{E}(\beta) = \widetilde{\mathcal{E}}_1(\beta) + \widetilde{\mathcal{E}}_2(\beta).$$

The reader should note that to compare subcase B2 with subcase A1, we interchange a_2 with a_1 and p_2 with p_1 (to maintain the same weights and proportions for the exponential and Gaussian levels).

The last interesting note is that in Gaussian-Exponential GREM, the system never reduces completely to a Gaussian REM as happened in subcase A3.

Thus the large deviation technique allows the use of different distributions at different levels leading to some interesting phenomenons. The conclusions of Exponential-Gaussian GREM differ from those of Exponential-Gaussian. The system may reduce to a Gaussian REM even with a very small weight is associated to that level. Even the system may appear as a system of two independent REMs separated by a big wall preventing them to interact between each other. Moreover, there are situations where we could not explain the terms present in the expression for energy.

Chapter 3

More Tree Structures including Randomness

In this chapter, we will consider several models similar to that of Generalized Random Energy Model. In the previous chapter, we formulated GREM in general tree set up with out giving any examples of general tree structures. The set up also allows us to randomize the tree structure. First we consider regular trees but the trees are random, driven by Poisson random variables. Then we consider non-regular random trees again driven by Poisson random variables. We prove that in both the cases the free energy exists for almost every tree sequences and they are same as that of usual deterministic tree GREMs for almost every sample point. Also we consider Multinomial trees. These will be explained later.

The usual GREM has hierarchical structure, and it is so in all the above mentioned models. In 2006, Bolthausen and Kistler [3] defined a model which is a generalization of the GREM where the model is no longer hierarchical. They called the model as non-hierarchical version of GREM and prove the existence of the free energy by using second moment method. Surprisingly, the energy expression is again the same as that of the usual GREM. So the non-hierarchy does not play a role in the limiting free energy. We produce an alternative proof of their result through large deviation techniques and show that the free energy of this model is minimum of certain hidden GREMs. Then we introduce another model, block tree GREM where the energy is maximum over certain GREM energies. We present further generalization in a model, in the next chapter, through which we can get all the models REM, GREM, Bolthausen-Kistler model and their versions with the external field.

3.1 Regular Poisson GREM

In generalized random energy model, we have randomness coming from the driving distributions. The reformulation of GREM in general tree structure allows us to introduce another randomness at the tree level which is independent of the randomness of the Hamiltonians. As usual for N particles system, let $\{k(i,N),\ 1 \le i \le n\}$ be a partition of N into n (the level of the tree) positive integers. Consider, for each N, independent random variables $L_{1N}, \cdots, L_{nN}$ where $L_{iN} \sim P(2^{k(i,N)})$, i.e. a Poisson random variable with parameter $2^{k(i,N)}$ for $1 \le i \le n$. Let us construct a random tree with $(1 + L_{iN})$ nodes at the i-th level below each node of the (i-1)-th level. That is, at the first level there will be $1 + L_{1N}$ many edges and at the second level there will be total $(1 + L_{1N})(1 + L_{2N})$ many edges. Here we are considering $1 + L_{iN}$ instead of L_{iN} itself, to take care of the situation $L_{iN} = 0$ so that each branch in the tree is of length n. Once again we denote the edges at the first level by σ_1 and the second level edges below σ_1 as $\sigma_1\sigma_2$ and so on. The weight of the i-th level is $a_i > 0$. Similarly we will associate independent random variable $\xi(\sigma_1 \cdots \sigma_i)$ with the edge $\sigma_1 \cdots \sigma_i$. In this case for N particle system, instead of 2^N configurations we will have $(1 + L_{1N})(1 + L_{2N}) \cdots (1 + L_{nN})$ many configurations. Of course this is also a regular tree, but random, and could be called regular Poisson tree. The corresponding GREM model, where the Hamiltonian for the configuration $\sigma = (\sigma_1, \cdots, \sigma_n)$ is defined as $\sum_{i=1}^{n} a_i \xi(\sigma_1 \cdots \sigma_i)$, can be called a *regular Poisson tree GREM* with parameter $\tilde{k} = (k(1,N), \cdots, k(n,N))$. The next result says that if the same conditions as in Corollary 2.3.2 hold then even with randomization of tree, the conclusion holds for almost every tree sequence.

Proposition 3.1.1. *Consider a regular Poisson tree GREM with parameter $\tilde{k}$. The following is true:*

a) If $\sum_{N \ge 1} 2^{k(1,N)+\cdots+k(i,N)} q_{1N} \cdots q_{iN} < \infty$, for some $i, 1 \le i \le n$ then for a.e. tree sequence, a.s. eventually, $\mu_N(\triangle) = 0$.

b) If $\sum_{N \ge 1} 2^{-(k(1,N)+\cdots+k(i,N))} q_{1N}^{-1} \cdots q_{iN}^{-1} < \infty$, for each $i = 1, \cdots, n$, then for a.e. tree sequence the following is true: for any $\epsilon > 0$, a.s. eventually,

$$(1 - \epsilon) q_{1N} \cdots q_{nN} \le \mu_N(\triangle) \le (1 + \epsilon) q_{1N} \cdots q_{nN}.$$

Proof. a) It suffices to verify the hypothesis of Theorem 2.3.1(a) holds for almost every tree sequence, that is, $\sum_{N \ge n} B_{iN} q_{1N} \cdots q_{iN} < \infty$ for some i. Recall that B_{iN} is the number of branches at the i-th level.

But we could prove a stronger statement, namely, if for some i with $1 \leq i \leq n$, $\sum_{N \geq n} 2^{k(1,N)+\cdots+k(i,N)} q_{1N} \cdots q_{iN} < \infty$, then $\mathbf{E}_T \sum_{N \geq n} B_{iN} q_{1N} \cdots q_{iN} < \infty$ for that i where $\mathbf{E}_T$ is the tree expectation. Since the tree randomness is independent of the Hamiltonian randomness, in view of the hypothesis, it suffices to show

$$\mathbf{E}_T B_{iN} \leq 2^{k(1,N)+\cdots+k(i,N)+i}. \tag{3.1.1}$$

Using independence of the random variables $(L_{jN},\ 1 \leq j \leq i)$, we get

$$\mathbf{E}_T B_{iN} = \mathbf{E} \prod_{j=1}^{i} (1 + L_{jN}) = \prod_{j=1}^{i} \mathbf{E}(1 + L_{jN}) = \prod_{j=1}^{i} (1 + 2^{k(j,N)}) \leq 2^i \prod_{j=1}^{i} 2^{k(j,N)}. \tag{3.1.2}$$

b) It is enough to show that for fixed $\epsilon > 0$, almost every tree sequence satisfies the stated conclusion. This is achieved by verifying that the hypothesis of Theorem 2.3.1(b) holds for almost every tree sequence, that is, $\sum_{N \geq n} \frac{s_{iN}^2}{B_N^2 q_{1N} \cdots q_{iN}} < \infty$. Recall that, $s_{iN}^2 = \sum_{\sigma_1, \cdots, \sigma_i} e^2(\sigma_1, \cdots, \sigma_i)$ where $e(\sigma_1 \sigma_2 \cdots \sigma_i)$ denotes the number of nodes at the n-th level below the node $\sigma_1 \sigma_2 \cdots \sigma_i$ and B_N is the number of leaves or the the total number of branches in the tree.

Here also we prove a stronger statement, namely, $\mathbf{E}_T \sum_{N \geq n} \frac{s_{iN}^2}{B_N^2 q_{1N} \cdots q_{iN}} < \infty$ for each i where $\mathbf{E}_T$ is the tree expectation. Again, since the tree randomness is independent of the Hamiltonian randomness, in view of the hypothesis, it suffices to show

$$\mathbf{E}_T \left(\frac{s_{iN}^2}{B_N^2} \right) \leq 2^{-(k(1,N)+\cdots+k(i,N))}. \tag{3.1.3}$$

But due to regularity of the tree $s_{iN}^2 = \prod_{j=1}^{i} (1 + L_{jN}) \prod_{j=i+1}^{n} (1 + L_{jN})^2$ and $B_N^2 = \prod_{j=1}^{n} (1 + L_{jN})^2$. Hence

$$\frac{s_{iN}^2}{B_N^2} = \prod_{j=1}^{i} \frac{1}{1 + L_{jN}}.$$

Thus using the independence of the random variables $(L_{jN},\ 1 \leq j \leq n)$, we get

$$\mathbf{E}_T\left(\frac{s_{iN}^2}{B_N^2}\right) = \prod_{j=1}^{i} \mathbf{E}\left(\frac{1}{1+L_{jN}}\right). \tag{3.1.4}$$

Since for a Poisson random variables X with parameter λ, $\mathbf{E}\frac{1}{1+X} = \frac{1}{\lambda}\left(1-e^{-\lambda}\right)$ and since $L_{jN} \sim P(2^{k(i,N)})$ we have,

$$\mathbf{E}\left(\frac{1}{1+L_{jN}}\right) = 2^{-k(j,N)}\left(1-e^{-2^{k(j,N)}}\right) \leq 2^{-k(j,N)}. \tag{3.1.5}$$

Substituting (3.1.5) in (3.1.4) we get (3.1.3). □

Now further if we assume that $\frac{k(i,N)}{N} \to p_i\,(>0)$ for $1 \leq i \leq n$ and the random variables $\xi(\sigma_1 \cdots \sigma_i)$ are distributed like ϕ_{N,γ_i} as defined in (2.4.1), the rest of the proof for existence of the free energy is the same as that of Theorem 2.4.6. Thus if the sequence $\left\{\frac{\xi(\sigma_1\cdots\sigma_i)}{N}\right\}$ satisfies LDP with good rate function $\mathcal{I}_i$ for each i with scale parameter N, then we have the following.

Theorem 3.1.2. *Assume the setup as in the above paragraph. For regular Poisson tree GREM, for almost every tree sequences, almost surely,*

$$\lim_N \frac{1}{N}\log Z_N(\beta) = \log 2 - \inf_{\widetilde{x}\in\Psi}\left\{\sum_{i=1}^{n}\left(\mathcal{I}_i(x_i) - \beta a_i x_i\right)\right\},$$

where

$$\Psi = \left\{\widetilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k}\mathcal{I}_i(x_i) \leq \sum_{i=1}^{k} p_i \log 2,\ \ 1 \leq k \leq n\right\}.$$

Thus, though we have another randomness in the setup of the model, the limiting free energy remains the same. That is why, in the original setup of GREM by Derrida, though (α_i^N) may not be an integer, with out any loss of generality one can consider the number of branches at the i-th level to be $[\alpha_i^N]$. We make this more precise in Remark 3.3.1.

We also note that, in this model the number of configurations in the configuration space may not be of the form α^N, where α ia a natural number. Instead, it is of the form $l_1 l_2 \cdots l_n$.

In the above model, there are n random variables controlling the number of nodes at the n level of the GREM. Since n is fixed and $N \to \infty$, one may get the impression that this extra randomness is not showing up in the final conclusion, namely, in the expression for the free energy. The next model shows that such an impression is not correct.

3.2 Poisson GREM

In the above model we randomized the tree sequences so that the underlying trees once again remain regular. But the general formulation allow us to consider a non-regular tree sequence. In the previous model, for each N, we randomized the tree using n many Poisson random variables corresponding to the levels of the tree. But it is conceivable to use independent Poisson variables at each of the nodes to construct the tree as well as the configuration space. This is what we do now. As in the previous model, let $\{k(1,N),\cdots,k(n,N)\}$ be a partition of N. Unlike in that model, now let us consider an n-level tree with $P(2^{k(i,N)})+1$, many nodes below each of the nodes at the $(i-1)$-th level for $1 \le i \le n$. Here $P(2^{k(i,N)})$ denotes a Poisson random variable with parameter $2^{k(i,N)}$. In other words, instead of fixing one random variable and taking so many nodes below each of the $(i-1)$-th level nodes, we now fix one random variable for each node of the $(i-1)$-th level and take so many nodes below that. Let us assume all these Poisson random variables are independent. As in the previous model, we denote a typical edge at the i-th level by $\sigma_1\cdots\sigma_i$ below the edge $\sigma_1\cdots\sigma_{i-1}$ and we associate independent random variables $\xi(\sigma_1\cdots\sigma_i)$ to it. We assume that this family $\{\xi(\sigma_1\cdots\sigma_i)\}$ is independent of the above Poisson family. For $1 \le i \le n$, we have a positive number a_i denoting the weights for the i-th level of the tree. Now we define the Hamiltonian for the configuration $\sigma = (\sigma_1,\cdots,\sigma_n)$ as

$$H_N(\sigma) = \sum_{i=1}^{n} a_i \xi(\sigma_1\cdots\sigma_i).$$

This model can be called a true *Poisson tree GREM* with parameter $\tilde{k} = (k(1,N),\cdots,k(n,N))$. Here we randomize, rather Poissonize, the tree in its full form. Even in this case also the model behaves the same way and we get the same conclusions as that of the above model, that is, (a) and (b) of Proposition 3.1.1 remain true. This is the content of the next proposition.

Proposition 3.2.1. *Consider a Poisson tree GREM with parameter $\tilde{k}$. The following is true:*

a) If $\sum_{N\ge1} 2^{k(1,N)+\cdots+k(i,N)} q_{1N}\cdots q_{iN} < \infty$, for some $i, 1 \le i \le n$ then for a.e. tree sequence, a.s. eventually, $\mu_N(\triangle) = 0$.

b) If $\sum_{N\ge1} 2^{-(k(1,N)+\cdots+k(i,N))} q_{1N}^{-1}\cdots q_{iN}^{-1} < \infty$, for each $i = 1,\cdots,n$, then for a.e.

tree sequence the following is true: for any $\epsilon > 0$, a.s. eventually,

$$(1-\epsilon)q_{1N}\cdots q_{nN} \le \mu_N(\triangle) \le (1+\epsilon)q_{1N}\cdots q_{nN}.$$

We need the following two inequalities to prove the proposition.

Let $a \ge 1$, $b \ge 1$ and $\lambda > 0$. Suppose that $X \sim P(a\lambda)$ and $Y \sim P(b\lambda)$ are independent random variables. Then

$$\mathbf{E}\left(\frac{X+a}{X+Y+a+b}\right)^2 \le 2\left(\frac{a}{a+b}\right)^2, \tag{3.2.1}$$

and

$$\mathbf{E}\frac{X+a}{(X+Y+a+b)^2} \le \frac{a}{(a+b)^2}\frac{1}{\lambda}. \tag{3.2.2}$$

Both these rely on conditioning. Since X and Y are independent Poisson random variables, given $X+Y=l$, the conditional distribution of X is binomial $\left(l, \frac{a}{a+b}\right)$. So

$$\begin{aligned}
&\mathbf{E}\left(\frac{X+a}{X+Y+a+b}\right)^2 \\
=&\mathbf{E}\left\{\frac{1}{(X+Y+a+b)^2}\mathbf{E}\left[(X+a)^2 \mid X+Y\right]\right\} \\
=&\mathbf{E}\left\{\frac{1}{(X+Y+a+b)^2}\frac{a^2(X+Y+a+b)^2+(X+Y)ab}{(a+b)^2}\right\} \\
\le&\frac{a^2}{(a+b)^2} \qquad \text{since } a \ge 1,
\end{aligned}$$

and

$$\begin{aligned}
&\mathbf{E}\left(\frac{X+a}{(X+Y+a+b)^2}\right) \\
=&\mathbf{E}\left\{\frac{1}{(X+Y+a+b)^2}\mathbf{E}[(X+a) \mid X+Y]\right\} \\
=&\mathbf{E}\left\{\frac{1}{(X+Y+a+b)^2}\frac{a(X+Y+a+b)}{(a+b)}\right\} \\
=&\frac{a}{(a+b)}\mathbf{E}\frac{1}{(X+Y+a+b)}
\end{aligned}$$

$$
\begin{aligned}
&< \frac{a}{(a+b)} \mathbf{E} \frac{1}{(X+Y+1)} \qquad \text{since } a+b>1 \\
&< \frac{a}{\lambda(a+b)^2}.
\end{aligned}
$$

Proof of Proposition 3.2.1. The proof is routine but involves rather cumbersome notation. To describe the random tree for the N-particle system, let $L_0 \sim P(2^{k(1,N)})$, the number of edges at the first level. For $1 \leq \sigma_1 \leq L_0 + 1$, let $L_{\sigma_1} \sim P(2^{k(2,N)})$, the number of edges at the second level below the first level edge σ_1. In general, for $\sigma_1\sigma_2\cdots\sigma_i$, with $1 \leq \sigma_1 \leq L_0+1$, $1 \leq \sigma_2 \leq L_{\sigma_1}+1, \cdots, 1 \leq \sigma_i \leq L_{\sigma_1\cdots\sigma_{i-1}}+1$, let $L_{\sigma_1\cdots\sigma_i} \sim P(2^{k(i+1,N)})$, the number of edges at the $(i+1)$-th level below the edge $\sigma_1\cdots\sigma_{i-1}$ at the i-th level.

To prove part (a), it suffices to show, as in Proposition 3.1.1, that

$$
\mathbf{E}_T B_{iN} \leq 2^{k(1,N)+\cdots+k(i,N)+i}.
$$

Since, in this model

$$
B_{iN} = \sum_{\sigma_1} \cdots \sum_{\sigma_{i-1}} (L_{\sigma_1\cdots\sigma_{i-1}}+1),
$$

the proof is immediate.

To prove (b), as in Proposition 3.1.1, it suffices to show that for each i,

$$
\mathbf{E}_T \left(\frac{s_{iN}^2}{B_N^2} \right) \leq 2^n 2^{-(k(1,N)+\cdots+k(i,N))}.
$$

But in this model,

$$
\frac{s_{iN}^2}{B_N^2} = \sum_{\sigma_1} \cdots \sum_{\sigma_i} \left(\frac{\sum_{\sigma_{i+1}} \cdots \sum_{\sigma_{n-1}} (L_{\sigma_1\cdots\sigma_{n-1}}+1)}{\sum_{\sigma_1} \cdots \sum_{\sigma_{n-1}} (L_{\sigma_1\cdots\sigma_{n-1}}+1)} \right)^2.
$$

To calculate the expectation we proceed as follows. Let $\mathcal{F}_o$ be the σ-field generated by L_0, $\mathcal{F}_1$ be the σ-field generated by $\{L_0, L_{\sigma_1} : 1 \leq \sigma_1 \leq L_0+1\}$ and in general

$\mathcal{F}_i$ be the σ-field generated by $\{L_0, L_{\sigma_1}, \cdots, L_{\sigma_1\cdots\sigma_i} : 1 \le \sigma_1 \le L_0+1, 1 \le \sigma_2 \le L_{\sigma_1}+1, \cdots, 1 \le \sigma_i \le L_{\sigma_1,\cdots,\sigma_{i-1}}\}$ for $i = 0, \cdots, n-2$. Let $\mathbf{E}_i$ be the conditional expectation given $\mathcal{F}_i$. Then (3.2.1) suggests that,

$$\begin{aligned}\mathbf{E}_{n-2}\left(\frac{s_{iN}^2}{B_N^2}\right) &= \mathbf{E}_{n-2}\sum_{\sigma_1}\cdots\sum_{\sigma_i}\left(\frac{\sum_{\sigma_{i+1}}\cdots\sum_{\sigma_{n-1}} L_{\sigma_1\cdots\sigma_{n-1}} + \sum_{\sigma_{i+1}}\cdots\sum_{\sigma_{n-1}} 1}{\sum_{\sigma_1}\cdots\sum_{\sigma_{n-1}} L_{\sigma_1\cdots\sigma_{n-1}} + \sum_{\sigma_1}\cdots\sum_{\sigma_{n-1}} 1}\right)^2 \\ &\le 2\sum_{\sigma_1}\cdots\sum_{\sigma_i}\left(\frac{\sum_{\sigma_{i+1}}\cdots\sum_{\sigma_{n-2}}(L_{\sigma_1\cdots\sigma_{n-2}}+1)}{\sum_{\sigma_1}\cdots\sum_{\sigma_{n-2}}(L_{\sigma_1\cdots\sigma_{n-2}}+1)}\right)^2.\end{aligned}$$

Similarly,

$$\mathbf{E}_{n-3}\mathbf{E}_{n-2}\left(\frac{s_{iN}^2}{B_N^2}\right) \le 2^2\sum_{\sigma_1}\cdots\sum_{\sigma_i}\left(\frac{\sum_{\sigma_{i+1}}\cdots\sum_{\sigma_{n-3}}(L_{\sigma_1\cdots\sigma_{n-3}}+1)}{\sum_{\sigma_1}\cdots\sum_{\sigma_{n-3}}(L_{\sigma_1\cdots\sigma_{n-3}}+1)}\right)^2,$$

and thus

$$\mathbf{E}_i\cdots\mathbf{E}_{n-2}\left(\frac{s_{iN}^2}{B_N^2}\right) \le 2^{n-i-1}\sum_{\sigma_1}\cdots\sum_{\sigma_i}\left(\frac{1}{\sum_{\sigma_1}\cdots\sum_{\sigma_i}(L_{\sigma_1\cdots\sigma_i}+1)}\right)^2.$$

Now we can use (3.2.2) to calculate further conditional expectations so that

$$\mathbf{E}_{i-1}\mathbf{E}_i\cdots\mathbf{E}_{n-2}\left(\frac{s_{iN}^2}{B_N^2}\right) \le 2^{n-i-1}\frac{1}{2^{k(i,N)}}\sum_{\sigma_1}\cdots\sum_{\sigma_{i-1}}\left(\frac{1}{\sum_{\sigma_1}\cdots\sum_{\sigma_i}(L_{\sigma_1\cdots\sigma_{i-1}}+1)}\right)^2,$$

and so on to get

$$\mathbf{E}_0\mathbf{E}_1\cdots\mathbf{E}_{n-2}\left(\frac{s_{iN}^2}{B_N^2}\right) \le 2^{n-i-1}\frac{1}{2^{k(i,N)+\cdots+k(1,N)}}.$$

Since $\mathbf{E}_T\left(\frac{s_{iN}^2}{B_N^2}\right) = \mathbf{E}_0\mathbf{E}_1\cdots\mathbf{E}_{n-2}\left(\frac{s_{iN}^2}{B_N^2}\right)$, the proof is complete. □

Once again to verify the existence of limiting free energy, one has to verify all the steps involved in section 2.4. To be precise, if we assume that $\frac{k(i,N)}{N} \to p_i\,(>0)$ for $1 \le i \le n$, the sequence $\left\{\frac{\xi(\sigma_1\cdots\sigma_i)}{N}\right\}$ satisfies LDP with good rate function $\mathcal{I}_i$ for each i, then we get

Theorem 3.2.2. *For almost every tree sequences, almost surely,*

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{\widetilde{x} \in \Psi} \left\{ \sum_{i=1}^{n} (\mathcal{I}_i(x_i) - \beta a_i x_i) \right\},$$

where

$$\Psi = \left\{ \widetilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} \mathcal{I}_i(x_i) \leq \sum_{i=1}^{k} p_i \log 2, \;\; 1 \leq k \leq n \right\}.$$

Remark 3.2.1. Recall that a tree is regular if for any i, the number of nodes below a $(i-1)$-th level node depends only on i and not on the specific nodes. We say that a tree sequence is regular if after some stage each tree in the sequence is regular. Under suitable conditions $--$ for instance, when $\sum_N e^{-k(i,N)} < \infty$ for some i $--$ it is possible to show that almost every tree sequence ceases to be regular. Though in this case the probability that the tree sequence will consist of regular trees is very small, we did not get any further new result except that, now the limiting free energy is constant for almost every tree sequences as well as almost every sample points.

3.3 Multinomial tree GREM

In the above two models, we randomized the number of nodes at each level keeping the average fixed. It is also possible to randomize the vector $\tilde{k}$ suitably. To do that, we fix $p_i > 0$ for $1 \leq i \leq n$ with $\sum_1^n p_i = 1$. Now consider an n-faced die with p_i being the chance of face i appearing in a throw. Now we can consider two experiments with this die. *Firstly*, we can consider an indefinite throws of the die and for N particle system let $K(i, N)$ be the number of times face i appears in first N throws. In the *second*, experiment for N particle system we will throw the die independently N times and observe the outcomes. With the same notation, let $K(i, N)$ be the number of times face i appears. Clearly, in both the cases $K(i, N) \geq 0$ and $\sum_{i=1}^n K(i, N) = N$. We can consider GREM with parameter $\tilde{K}$, that is where the under lying tree has $2^{K(i,N)}$ many edges below each of the $(i-1)$ level node. These can be called *multinomial tree GREM* of first kind and *multinomial tree GREM* of second kind with parameter $\tilde{p} = (p_1, \cdots, p_n)$ respectively. With the same notation as in section 2.3, in this case we have

$$\mathbf{E}_T B_{iN} = \mathbf{E}\prod_{j=1}^{i} 2^{K(j,N)} = \sum_{k=0}^{N} 2^k \binom{N}{k} \left(\sum_1^i p_j\right)^k \left(1-\sum_1^i p_j\right)^{N-k} = \left(1+\sum_1^i p_j\right)^N.$$

But for any x, as $(1+x)^N \leq e^{Nx}$, we have

$$\mathbf{E}_T B_{iN} \leq e^{N\sum_1^i p_j} = 2^{\frac{N}{\log 2}(p_1+\cdots+p_i)}. \tag{3.3.1}$$

On the other hand, in this case, $s_{iN}^2 = 2^{K(1,N)+\cdots+K(i,N)}2^{2(K((i+1),N)+\cdots+K(n,N))}$ and $B_N^2 = 2^{2\sum_{j=1}^{n} K(j,N)}$ so that $\left(\frac{s_{iN}^2}{B_N^2}\right) = 2^{-\sum_1^i K(j,N)}$. Once again using the fact that $\sum_1^i K(j,N)$ is binomial with parameters N and $\sum_1^i p_j$, we see that

$$\mathbf{E}2^{-\sum_1^i K(j,N)} = \left(1-\frac{1}{2}\sum_1^i p_j\right)^N.$$

Hence by same inequality as earlier, we have

$$\mathbf{E}2^{-\sum_1^i K(j,N)} \leq e^{-\frac{N}{2}\sum_1^i p_j} = 2^{-\frac{N}{2\log 2}\sum_1^i p_j}. \tag{3.3.2}$$

Combining the above observations (3.3.1) and (3.3.2), we have the following

Corollary 3.3.1. *Consider a multinomial tree GREM either of first kind or of second kind with parameter $\tilde{p}$. Let $\triangle = \triangle_1 \times \cdots \times \triangle_n$ be a box in $\mathbb{R}^n$ and $q_{iN} = P(\frac{\xi(\sigma_1\cdots\sigma_i)}{N}) \in \triangle_i$.*

a) If $\sum_{N\geq 1} 2^{\frac{N}{\log 2}(p_1+\cdots+p_i)} q_{1N}\cdots q_{iN} < \infty$, for some $i, 1 \leq i \leq n$ then for a.e. tree sequence, a.s. eventually, $\mu_N(\triangle) = 0$.

b) If $\sum_{N\geq 1} 2^{-\frac{N}{2\log 2}(p_1+\cdots+p_i)} q_{1N}^{-1}\cdots q_{iN}^{-1} < \infty$, for each $i = 1, \cdots, n$, then for a.e. tree sequence the following is true: for any $\epsilon > 0$, a.s. eventually,

$$(1-\epsilon)q_{1N}\cdots q_{nN} \leq \mu_N(\triangle) \leq (1+\epsilon)q_{1N}\cdots q_{nN}.$$

Notice the difference in the hypothesis of (a) and (b) in the above corollary. More specifically, there is a factor $\frac{1}{2}$ extra in the exponent of 2 in part (b). This difference will not help us to obtain the exact support of the empirical measure μ_N. For the variational problem this support is very essential. But, of course, we have strong law of large number in our hand. By SLLN, almost surely, $\frac{1}{N}(K(1,N),\cdots,K(n,N)) \to (p_1,\cdots,p_n)$ in both the models, first or second kind. In the first case, we use SLLN for a sequence of i. i. d. random variables and in the second case we use SLLN for array of rowwise independent random variables [27]. According to Hu et al in [27]: *If* $\{X_{nk}\}$ *be an array of rowwise independent random variables such that* $EX_{nk} = 0$ *and there exists a random variable* X *with* $EX^2 < \infty$ *so that for all* n *and* k *and for all* $t > 0$, $P(|X_{nk}| > t) \leq P(|X| > t)$, *then* $\frac{1}{n}\sum_{k=1}^{n} X_{nk} \to 0$ *almost surely.* Since we can write $K(i,N)$ as $\sum_{k=1}^{N} X_{Nk}^{(i)}$ where $X_{Nk}^{(i)}$ are Bernoulli with success probability p_i, the above result is applicable for the second kind model. Thus, in either of the cases, for every $\epsilon > 0$ almost every tree sequences after some stage

$$N(p_1 + \cdots + p_i - \epsilon) < K(1,N) + \cdots + K(i,N) < N(p_1 + \cdots + p_i + \epsilon)$$

for $i = 1,\cdots,n$. That is for almost every tree sequences and for any arbitrary $\epsilon > 0$, $B_{iN} \leq 2^{N(p_1+\cdots+p_i+\epsilon)}$ and $\frac{s_{iN}^2}{B_N^2} < 2^{-N(p_1+\cdots+p_i-\epsilon)}$ for $i = 1,\cdots,n$. As a consequence, we can restate Theorem 2.3.1 as follows

Corollary 3.3.2. *Consider a multinomial tree GREM with parameter* $\tilde{p}$. *Let* $\triangle = \triangle_1 \times \cdots \times \triangle_n$ *be a box in* $\mathbb{R}^n$ *and* $q_{iN} = P(\frac{\xi(\sigma_1\cdots\sigma_i}{N}) \in \triangle_i$. *Let* $\epsilon > 0$.

a) If $\sum_{N\geq 1} 2^{N\epsilon}2^{N(p_1+\cdots+p_i)}q_{1N}\cdots q_{iN} < \infty$, *for some* $i, 1 \leq i \leq n$ *then for a.e. tree sequence, a.s. eventually,* $\mu_N(\triangle) = 0$.

b) If $\sum_{N\geq 1} 2^{N\epsilon}2^{-N(p_1+\cdots+p_i)}q_{1N}^{-1}\cdots q_{iN}^{-1} < \infty$, *for each* $i = 1,\cdots,n$, *then for a.e. tree sequence the following is true: a.s. eventually,*

$$(1-\epsilon)q_{1N}\cdots q_{nN} \leq \mu_N(\triangle) \leq (1+\epsilon)q_{1N}\cdots q_{nN}.$$

Now the proof of existence of the asymptotic free energy for this model is routine and for almost every tree sequence the expression for free energy will be same as that of the deterministic tree model where $\frac{k(i,N)}{N} \to p_i$ for $1 \leq i \leq n$. To state precisely, let us we assume that the sequence $\left\{\frac{\xi(\sigma_1\cdots\sigma_i)}{N}\right\}$ satisfies LDP with good rate function $\mathcal{I}_i$ for each i, then we have the following.

Theorem 3.3.3. *With the setup as in the above paragraph, for almost every tree* *sequence, almost surely,*

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{\widetilde{x}\in\Psi} \left\{ \sum_{i=1}^{n} (\mathcal{I}_i(x_i) - \beta a_i x_i) \right\},$$

where

$$\Psi = \left\{ \widetilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} \mathcal{I}_i(x_i) \leq \sum_{i=1}^{k} p_i \log 2, \ \ 1 \leq k \leq n \right\}.$$

Remark 3.3.1. Going back to Theorem 2.3.1, let (T_N) and $(\tilde{T}_N)$ be two sequences of trees. Suppose there are numbers $C > c > 0$ such that for each i, $c \leq \frac{\tilde{s}_{iN}}{s_{iN}} \leq C$ and $c \leq \frac{\tilde{B}_{iN}}{B_{iN}} \leq C$. Then it is easy to see that, hypothesis of Theorem 2.3.1(b) holds for (T_N) iff it holds for $(\tilde{T}_N)$. Accordingly, the conclusion of Theorem 2.3.1(b) holds for (T_N) iff it holds for $(\tilde{T}_N)$. Same remark applies for Theorem 2.3.1(a).

3.4 Bolthausen - Kistler GREM

In 2006, Bolthausen and Kistler proposed a model where they tried to go beyond the natural *ultrametricity* of the GREM model. To recall, a metric d is *ultrametric* if in the metric property one replaces the triangle inequality by $d(x, z) \leq \max(d(x, y), d(y, z))$. In all of the above GREM model one can define a metric on the configuration space Σ_N of the N particle system through the covariance structure of the Hamiltonian. To be precise for two configurations σ and τ in Σ_N,

$$d(\sigma, \tau) = \sqrt{E(H_N(\sigma) - H_N(\tau))^2}. \tag{3.4.1}$$

Here as usual, $H_N(\sigma) = \sum_{i=1}^{n} a_i \xi(\sigma_1 \cdots \sigma_i)$ with the usual GREM notation. Let $\sigma = (\sigma_1, \cdots, \sigma_n)$ and $\tau = (\tau_1, \cdots, \tau_n)$. If $\sigma = \tau$ then $d(\sigma, \tau) = 0$. If $\sigma_i = \tau_i$ for $1 \leq i \leq k < n$ but $\sigma_{(k+1)} \neq \tau_{(k+1)}$ then $d(\sigma, \tau) = \sqrt{2 \sum_{i=k+1}^{n} a_i^2 E\xi(\sigma_1 \cdots \sigma_i)^2}$, assuming that the ξ's are symmetric with finite variance. The distance between any two configuration will be maximum, when they differ at the first level of the tree. The longer the initial segment of σ and τ coincide, the closer they are. Also it is quite easy to verify that this metric indeed is an ultrametric. To see this, it is enough to see the level of difference among the configurations. Suppose, we have any three configurations σ, τ and η in Σ_N. Let k_i, k_2 and k_3 be the maximum non-negative integers so that $\sigma_i = \tau_i$ for $1 \leq i \leq k_1$; $\sigma_i = \eta_i$ for $1 \leq i \leq k_2$ and $\tau_i = \eta_i$ for $1 \leq i \leq k_3$ respectively. To show $d(\sigma, \tau) \leq \max(d(\sigma, \eta), d(\eta, \tau))$, we only need to show that $k_1 \geq \min(k_2, k_3)$.

Without loss of generality if we assume $k_2 \leq k_3$ then $\sigma = \eta_1 \cdots \eta_{k_2}\sigma_{k_2+1} \cdots \sigma_n$ and $\tau = \eta_1 \cdots \eta_{k_3}\tau_{k_3+1} \cdots \tau_n$. If $k_1 < k_2$, then $\sigma_{k_1+1} = \eta_{k_1+1} = \tau_{k_1+1}$ will contradict the maximality of k_1.

We will denote the model of Bolthausen and Kistler as BK-GREM. The set up in the BK-GREM is the following: For a fixed number $n \in \mathbb{N}$, they consider the set $I = \{1, 2, \cdots, n\}$ and a collection of non-negative real numbers $\{a_J\}_{J \subset I}$ such that $\sum\limits_{J \subset I} a_J = 1$ with $a_\varnothing = 0$. There may be subsets J of I for which $a_J = 0$, so they consider $\mathcal{P}_I$ as that collection of subsets J of I for which $a_J > 0$ that is, $\mathcal{P}_I = \{J : a_J > 0\}$. Like in the usual GREM, they fix n positive real numbers γ_i for $i \in I$ so that $\sum\limits_{i=1}^{n} \gamma_i = 1$ and split the configurations space $\Sigma_N = \{-1, 1\}^N$ in to products $\Sigma_{\gamma_1 N} \times \Sigma_{\gamma_2 N} \times \cdots \times \Sigma_{\gamma_n N}$ with $\Sigma_{\gamma_i N} = \{-1, 1\}^{\gamma_i N}$ for each i. Since $\gamma_i N$ may not be an integer one needs to use $[\gamma_i N]$ instead of $\gamma_i N$, and $\prod_{i=1}^{n} \Sigma_{[\gamma_i N]}$ as the configuration space etc. Since we shall soon reformulate this model we do not elaborate on these points. So a configuration σ can be written as $(\sigma_1, \cdots, \sigma_n)$. For $J = \{j_1, \cdots, j_k\} \subset I$, denote $\Sigma_{J,N}$ for $\prod\limits_{l=1}^{k} \Sigma_{\gamma_{j_l} N}$ and σ_J for the projected configuration $(\sigma_j)_{j \in J} \in \Sigma_{J,N}$. In this setup, the random Hamiltonian is defined as

$$H_N(\sigma) = \sum_{J \in \mathcal{P}_I} \xi_J(\sigma_J),$$

where for $J \in \mathcal{P}_I$ and $\sigma_J \in \Sigma_{J,N}$ the random variables $\xi_J(\sigma_J)$ are independent centered Gaussian random variables with variance $a_J N$. It is quite easy to verify that, in this model if we define the metric on the configurations space by the same formula as in (3.4.1) then the metric will not be always an ultrametric. But yet they have shown that the limiting free energy is again a GREM free energy. To be precise, define a *chain* $(A_0, A_1, \cdots, A_k)$ to be an increasing sequence of subsets of I with $\varnothing = A_0 \subset A_1 \subset \cdots \subset A_k = I$. What they have shown is that for any BK-GREM there exists a chain $(A_0, A_1, \cdots, A_k)$ and positive constants $\widetilde{a}_i$ for $1 \leq i \leq k$ with $\sum\limits_{i=1}^{k} \widetilde{a}_i = 1$ such that the following holds: the free energy of this BK-GREM is same as that of a Gaussian k level GREM energy where random variables at the i-th level have variance $\widetilde{a}_i N$. This means, once again the limiting free energy does not go beyond the GREM one. We will present here an alternative and elegant proof of this thanks to large deviation results. To do that we will reformulate the model in the next section.

3.4.1 Reformulation

We formulate BK-GREM as follows. Fix a set $I = \{1, 2, \cdots, n\}$ with $n \geq 1$. Let $N \geq n$ be the number of particles, each of which can have two states/spins $+1, -1$; so that the configuration space is $\Sigma_N = 2^N$. Consider a partition of N into integers $k(i, N), 1 \leq i \leq n$ with each $k(i, N) \geq 1$ and $\sum\limits_{i} k(i, N) = N$. We will as usual think of 2^N as $\prod_{i\in I} 2^{k(i,N)}$ and $\sigma \in 2^N$ is $\sigma_1 \cdots \sigma_n$ where $\sigma_i \in 2^{k(i,N)}$. Let S be the collection of non-empty subset of I. For each element s in S we denote $2^{K_{sN}} = \prod_{i\in s} 2^{k(i,N)}$. With this notation $2^N = 2^{K_{IN}}$. The map $\sigma \in 2^{K_{IN}} \to \sigma(s) \in 2^{K_{sN}}$ is the projection map via s. For $s = \{i_1, i_2, \cdots, i_k\} \in S$ where $i_1 < i_2 < \cdots < i_k$ and $\sigma = \sigma_1 \cdots \sigma_n \in 2^{K_{IN}}$, we denote $\sigma(s) = \sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_k} \in 2^{K_{sN}}$, the projection of σ via s. Now for fixed N, we have a bunch of independent random variables $\xi(s, \sigma(s))$ as s varies over S and $\sigma(s)$ varies over $2^{K_{sN}}$.

For each $\sigma \in 2^N$ one can think of a lattice isomorphic to the lattice of power set of I where $\sigma(s)$ corresponds to the edge of the lattice joining the nodes $s = \{i_1, i_2, \cdots, i_k\}$ and $\{i_1, i_2, \cdots, i_{k-1}\}$. Now for each σ associate random variables $\xi(s, \sigma(s))$ to each of the lattice edge $\sigma(s)$. We associate weights $a_s \geq 0$ to each edges $\sigma(s)$. These are not random. In a configuration $\sigma = \sigma_1 \cdots \sigma_n$ the Hamiltonian is defined as

$$H_N(\sigma) = N \sum_{s\in S} a_s \xi(s, \sigma(s)).$$

For $\beta > 0$ the partition function is

$$Z_N(\beta) = 2^N \mathbf{E}_\sigma e^{-\beta H_N(\sigma)}.$$

Here $\mathbf{E}_\sigma$ stands for expectation with respect to σ when 2^N has uniform distribution. In other words, $\mathbf{E}_\sigma$ is simply the usual average over σ.

Since ξ's are random variables both H_N and Z_N are random variables. We suppress the parameter ω that comes with the random variables ξ. As usual $\frac{1}{N}\log Z_N(\beta)$ is the free energy of the N-particle system. As N changes, the distribution of the ξ's would in general depends on N. So strictly speaking we should be using superscript N for the random variables. But for ease in reading we suppress the superscript. This should be borne in mind. We assume that all our random variables are defined on one probability space.

3.4.2 LDP Approach

In this subsection, we outline how large deviation principle can be used. Since we will prove a more general result in the next chapter (see section 4.3), we refrain from giving complete details. Let us consider the map $\Sigma_N \to \mathbb{R}^S$ (recall S is the collection of non-empty subsets of I) defined by

$$\sigma \mapsto \xi_\sigma = (\xi(s, \sigma(s)))_{s\in S}.$$

Let μ_N be the induced probability on $\mathbb{R}^S$ when Σ_N has uniform distribution, that is, each $\sigma \in \Sigma_N$ has probability $\frac{1}{2^N}$. In other words, for any Borel set $A \subset \mathbb{R}^S$,

$$\mu_N(A) = \frac{1}{2^N}\#\{\sigma : \xi_\sigma \in A\}.$$

In particular, if A is a box, say $\triangle = \prod_{s\in S} \triangle_s$, with each $\triangle_s \subseteq \mathbb{R}$, then

$$\mu_N(\triangle) = \frac{1}{2^N}\sum_{\sigma}\prod_{s\in S}\mathbf{1}_{\triangle_s}(\xi(s,\sigma(s))).$$

Here now is the basic observation similar to that of Theorem 2.3.1.

Theorem 3.4.1. *Let $\triangle = \prod_{s\in S}\triangle_s \subset \mathbb{R}^S$. Denote $q_{sN} = P(\xi(s,\sigma(s)) \in \triangle_s)$ for $s \in S$. For $t \in S$ we denote $\prod_{s\subseteq t} q_{sN}$ by Q_{tN} and $\prod_{i\in t} k(i,N)$ by K_{tN}.*

a) If $\sum_{N\geq 1} 2^{K_{tN}} Q_{tN} < \infty$, for some $t \in S$ then a.s. eventually, $\mu_N(\triangle) = 0$.

b) If for all $t \in S$, $\sum_{N\geq 1} 2^{-K_{tN}} Q_{tN}^{-1} < \infty$, then for any $\epsilon > 0$ a.s. eventually,

$$(1-\epsilon)\mathbf{E}\mu_N(\triangle) \leq \mu_N(\triangle) \leq (1+\epsilon)\mathbf{E}\mu_N(\triangle).$$

There is no new idea, we need to verify that arguments of the previous chapter go through.

Proof. a) Let t be such that $\sum_{N\geq 1} 2^{K_{tN}} Q_{tN} < \infty$. Then

$$\begin{aligned}\mu_N(\triangle) &= \frac{1}{2^N}\sum_{\sigma}\prod_{s\in S}\mathbf{1}_{\triangle_s}(\xi(s,\sigma(s))) \\ &\leq \frac{2^{N-K_{tN}}}{2^N}\sum_{\sigma(t)}\prod_{s\subseteq t}\mathbf{1}_{\triangle_s}(\xi(s,\sigma(s))) \\ &= \frac{1}{2^{K_{tN}}}\sum_{\sigma(t)}\prod_{s\subseteq t}\mathbf{1}_{\triangle_s}(\xi(s,\sigma(s))) = G_N, \text{(say)}.\end{aligned}$$

Let $A_N = \{G_N = 0\}$. Observe that

$$A_N^c = \left\{\sum_{\sigma(t)}\prod_{s\subseteq t}\mathbf{1}_{\triangle_s}(\xi(s,\sigma(s))) \geq 1\right\}.$$

Now by Chebyshev's inequality,

$$\mathbf{P}(A_N^c) < \mathbf{E}\sum_{\sigma(t)}\prod_{s\subseteq t}\mathbf{1}_{\triangle_s}(\xi(s,\sigma(s))) = 2^{K_tN}Q_{tN}.$$

Thus by assumption and Borel-Cantelli, A_N will occur a.s. eventually. i.e. $G_N = 0$ and hence $\mu_N(\triangle) = 0$.

b)

$$\begin{aligned}
&\mathrm{Var}(\mu_N(\triangle))\\
=&\mathbf{E}(\mu_N(\triangle))^2 - (\mathbf{E}\mu_N(\triangle))^2\\
=&\frac{1}{2^{2N}}\sum_{\sigma,\tau}\left[\mathbf{E}\prod_{s\in S}\mathbf{1}_{\triangle_s}(\xi(s,\sigma(s)))\mathbf{1}_{\triangle_s}(\xi(s,\tau(s))) - Q_{IN}^2\right]\\
=&\frac{1}{2^{2N}}\sum_{t\in S}\sum_{\substack{\sigma(t)=\tau(t)\\ \sigma_i\neq\tau_i,\forall i\in t^c}}\left[\mathbf{E}\prod_{s\subseteq t}\mathbf{1}_{\triangle_s}(\xi(s,\sigma(s)))\prod_{s\nsubseteq t}\mathbf{1}_{\triangle_s}(\xi(s,\sigma(s)))\mathbf{1}_{\triangle_s}(\xi(s,\tau(s))) - Q_{IN}^2\right]\\
\leq&\frac{1}{2^{2N}}\sum_{t\in S}\prod_{s\subseteq t}q_{sN}\prod_{s\nsubseteq t}q_{sN}^2\sum_{\substack{\sigma(t)=\tau(t)\\ \sigma_i\neq\tau_i,\forall i\in t^c}}1\\
\leq&\frac{1}{2^{2N}}\sum_{t\in S}\frac{Q_{IN}^2}{Q_{tN}}2^{K_{tN}}2^{2(N-K_{tN})}\\
=&\sum_{t\in S}\frac{Q_{IN}^2}{2^{K_{tN}}Q_{tN}}.
\end{aligned}$$

Hence for any $\epsilon > 0$, by Chebyshev's inequality

$$\mathbf{P}(|\mu_N(\triangle) - \mathbf{E}\mu_N(\triangle)| > \epsilon\mathbf{E}\mu_N(\triangle)) < \frac{1}{\epsilon^2}\sum_{t\in S}\frac{1}{2^{K_{tN}}Q_{tN}}.$$

But, in view of the assumption, the sum over N of the right side is finite. So by Borel-Cantelli lemma, a.s. eventually,

$$(1-\epsilon)\mathbf{E}\mu_N(\triangle) \leq \mu_N(\triangle) \leq (1+\epsilon)\mathbf{E}\mu_N(\triangle).$$

□

Suppose that for each element $s \in S$, we have a sequence of probabilities $\{\lambda_N^s : N \geq n\}$ obeying LDP with a good convex rate function $\mathcal{I}_s(x)$. We now consider reformulated BK-model where each $\xi(s, \sigma(s))$ has distribution λ_N^s. Thus for fixed N, we have a bunch of independent random variables $\xi(s, \sigma(s))$ as s and $\sigma(s)$ vary. For example, for the N particle system, one can consider for each $s \in S$; $\xi(s, \sigma(s))$ to be i.i.d. having density

$$\phi(x) = \frac{1}{2\Gamma(\frac{1}{\gamma_s})} \left(\frac{\gamma_s}{N}\right)^{\frac{\gamma_s - 1}{\gamma_s}} e^{-N\frac{|x|^{\gamma_s}}{\gamma_s}} \qquad -\infty < x < \infty. \tag{3.4.2}$$

Let us denote

$$\Psi = \{\widetilde{x} \in \mathbb{R}^S : \sum_{s \subseteq t} \mathcal{I}_s(x_s) \leq \sum_{i \in t} p_i \log 2, \ \ \forall t \in S\} \tag{3.4.3}$$

and the map $\mathcal{J} : \mathbb{R}^S \to \mathbb{R}$, defined by,

$$\begin{aligned} \mathcal{J}(\tilde{x}) &= \sum_{s \in S} \mathcal{I}_s(x_s) \quad \text{if } \tilde{x} \in \Psi \\ &= \infty \qquad\qquad\quad \text{otherwise.} \end{aligned}$$

Then with the help of Theorem 3.4.1, one can mimic the steps in Theorem 2.5.1 to get

Theorem 3.4.2. *In the reformulated BK-GREM, let* $\frac{k(i,N)}{N} \to p_i > 0$ *as* $N \to \infty$ *for* $1 \leq i \leq n$. *Then almost surely, the sequence* $\{\mu_N, N \geq 1\}$ *satisfies LDP with rate function* $\mathcal{J}$ *defined above.*

In this way, once again Varadhan's lemma will ensure the existence of the limiting free energy in this case also. Thus though we don't have ultrametricity on the configuration space, the simple LDP technique works.

Theorem 3.4.3. *In reformulated BK-GREM, almost surely*

$$\lim_{N \to \infty} \frac{1}{N} \log Z_N(\beta) = \log 2 - \inf_{\widetilde{x} \in \Psi} \sum_{s \in S} \left(\beta a_s x_s + \mathcal{I}_s(x_s)\right).$$

Remark 3.4.1. One may feel that the reformulation of BK-GREM is not exactly similar to the original version of Bolthausen and Kistler. They consider only those subsets s

of I for which $a_s \neq 0$ whereas in the above reformulation all the non-empty subsets are considered. But from the above theorem it is easy to check that in calculating infimum, $\mathcal{I}_s$ being non-negative and $\mathcal{I}_s(0) = 0$, the terms corresponding to those $s \in S$ for which $a_s = 0$ will not contribute.

In [3], Bolthausen and Kistler identified the free energy of this model as minimum of several GREMs associated with, what they call, chains. We shall show that there are $n!$ many n level usual GREMs hidden in the above model. The method used above also identifies the free energy of the BK-GREM as the minimum of the free energies of these $n!$ GREMs. This is what we do in the next section.

3.5 Hidden Tree GREMs

In this section, we consider the BK-GREM, that is, we take S to be the set of all increasing sequences of elements of I with the Gaussian driving distributions. As mentioned earlier, this is nothing but the Bolthausen-Kistler's model since such sequences correspond to non-empty subsets of I. Suppose now for $s \in S$ the associated weight is a_s. Here we evaluate the explicit expression for the limiting free energy of BK-GREM. Though it is possible to consider different driving distributions for each $s \in S$, a general closed form expression appears to be difficult. Of course, we could also start with some more general driving distributions than Gaussian, like distribution having density ϕ as in (3.4.2) with $\gamma > 1$ at all the levels. Since in that case, there is no new idea needed, we restrict ourselves to Gaussian case for notational simplicity. It is worth mentioning here that, in [3], Bolthausen and Kistler evaluate the expression of the limiting free energy in two steps. In the first step, they define a *chain* as a sequence of strictly increasing sequences of subsets $(A_0, A_1, \cdots, A_K)$ of I so that $\varnothing = A_0 \subset A_1 \subset \cdots \subset A_K = I$. For such a chain they associated a K level GREM with appropriate weights calculated from the weights of the original model. Then by second moment estimates, they have shown that the limiting free energy of each such GREM associated to a chain is an almost sure upper bound for the limiting free energy of their model. In the second step, they constructed a chain in which the free energy of the BK-GREM is attained.

Here we get the expression for the limiting free energy by calculating

$$\inf_{\widetilde{x} \in \Psi} \sum_{s \in S} \left(\beta a_s x_s + \frac{1}{2} x_s^2 \right), \tag{3.5.1}$$

where

$$\Psi = \{\widetilde{x} \in \mathbb{R}^S : \sum_{t \subseteq s} x_t^2 \leq \sum_{i \in s} 2p_i \log 2, \ \forall s \in S\}.$$

Note that Ψ is same as that of (3.4.3) with $\mathcal{I}_s(x_s) = \frac{1}{2}x_s^2$. As earlier, the above infimum is same as

$$\inf_{\widetilde{x}\in\Psi^+} \sum_{s\in S} \left(\frac{1}{2}x_s^2 - \beta a_s x_s\right), \tag{3.5.2}$$

where

$$\Psi^+ = \{\widetilde{x} \in \mathbb{R}^S : \sum_{t\subseteq s} x_t^2 \leq \sum_{i\in s} 2p_i \log 2 \ \&\ x_s \geq 0, \ \forall s \in S\}. \tag{3.5.3}$$

To evaluate (3.5.2), we start with some notations. Here the ideas are very much similar to that of Bolthausen and Kistler. A new idea is the introduction of permutations of $\{1, 2, \cdots, n\}$ justifying the title of this section. For $A \subseteq I$, let us define

$$p_A = \sum_{i\in A} p_i$$

and

$$w_A^2 = \sum_{\substack{s\subseteq A\\ s\in S}} a_s^2,$$

with $w_\varnothing^2 = 0$. Let $\mathcal{P}_0 = \mathcal{P}_I$ denote the set of permutations of I. With this notation, for $\pi \in \mathcal{P}_I$ and $0 \leq i < j \leq n$, denote

$$B_{ij}^\pi = \sqrt{\frac{2(p_{\pi(i+1)} + \cdots + p_{\pi(j)}) \log 2}{w^2_{\{\pi(1),\cdots,\pi(j)\}} - w^2_{\{\pi(1),\cdots,\pi(i)\}}}}, \tag{3.5.4}$$

where, for $i = 0$ the set $\{\pi(1), \cdots, \pi(i)\}$ that appears in the denominator is treated as the empty set.

Note that earlier to evaluate the explicit energy expression for γ-GREM with $\gamma > 1$, in subsection 2.6.1, we consider only one triangular array of numbers defined as $B(j,k)$ in (2.6.4) for $1 \leq j \leq k \leq n$. Now here we are considering $n!$ many triangular arrays corresponding to each permutation π.

Now let,

$$\beta_1 = \min_{\pi\in\mathcal{P}_I} \min_{0<j\leq n} B_{0j}^\pi = \min_{(\pi,j)} B_{0j}^\pi. \tag{3.5.5}$$

Also note that, in subsection 2.6.1, we define β_1 in (2.6.5) as the minimum over all the entries in the first row of the triangular array $B(j,k)$ in (2.6.4). Here we are defining β_1 as the minimum over all the entries in the first rows of all the $n!$ triangular arrays.

It may be further noted that in subsection 2.6.1, to define β_1 if the minimum occurred at two places, we had taken the maximum index (see (2.6.4) for the definition of r_l). We now implement the same plan in the present setting also. Since the minimum may be attained in the first lines of two different triangular arrays, one needs to know what is meant by maximum index. This will be done now.

Suppose the minimum occurs at two places, say at (θ, k) and (ϱ, l), that is, $\beta_1 =$

$B^{\theta}_{0k} = B^{\varrho}_{0l}$. Let $G = \{\theta(1), \cdots, \theta(k)\}$; $H = \{\varrho(1), \cdots, \varrho(l)\}$. Let $|G \cup H| = m$ and let $\mathcal{P}_1$ denote the class of all permutations of I for which $\{\pi(1), \cdots, \pi(m)\} = G \cup H$. So $\mathcal{P}_1 \subset \mathcal{P}_I$.

Claim: $B^{\pi}_{0m} = \sqrt{\frac{2p_{G\cup H}\log 2}{w^2_{G\cup H}}} = \beta_1$, *for every* $\pi \in \mathcal{P}_1$.

To justify the claim, first of all note that

$$w^2_{G\cup H} \geq w^2_G + w^2_H - w^2_{G\cap H},$$

whereas,

$$p_{G\cup H} = p_G + p_H - p_{G\cap H}.$$

Then for $\pi \in \mathcal{P}_1$,

$$\begin{aligned}
&2p_{G\cup H}\log 2 - \beta_1^2 w^2_{G\cup H} \\
&\leq 2(p_G + p_H - p_{G\cap H})\log 2 - \beta_1^2\left(w^2_G + w^2_H - w^2_{G\cap H}\right) \\
&= \left(2p_G \log 2 - \beta_1^2 w^2_G\right) + \left(2p_H \log 2 - \beta_1^2 w^2_H\right) + \left(\beta_1^2 w^2_{G\cap H} - p_{G\cap H}\right)
\end{aligned}$$

as $\beta_1 = B^{\theta}_{0k} = B^{\varrho}_{0l}$, first two terms are zero,

$$\begin{aligned}
&= \beta_1^2 w^2_{G\cap H} - p_{G\cap H} \\
&\leq 0.
\end{aligned}$$

The last inequality follows from the fact that β_1 is obtained by taking the minimum over all possible choice of (π, j). This shows $\beta_1^2 \geq \frac{2p_{G\cup H}\log 2}{w^2_{G\cup H}}$.

Once again β_1 being the minimum over all possible choice of (π, j), we conclude that β_1^2 actually equals $\frac{2p_{G\cup H}\log 2}{w^2_{G\cup H}}$ proving the claim.

If the minimum in (3.5.5) occurs at more than two places, we can use induction to conclude that there exists a unique maximal set, say, $G_1 \subseteq I$ such that the following holds. Let $|G_1| = l_1$ and $\mathcal{P}_1 =$ all permutations that map $\{1, 2, \cdots l_1\}$ on to G_1. Then for any $\pi \in \mathcal{P}_1$, $B^{\pi}_{0l_1} = \beta_1$.

It may happen that $G_1 = I$, then we will stop. Otherwise, let us define

$$\beta_2 = \min_{\pi\in\mathcal{P}_1} \min_{l_1<j\leq n} B^{\pi}_{l_1 j} = \min_{(\pi,j)} B^{\pi}_{l_1 j}.$$

Of course, the last minimum is only over $\pi \in \mathcal{P}_1$.

Once again going back to subsection 2.6.1, to define β_2 in (2.6.5), we look only the entries from the $r_1 + 1$-th row of the triangular array $B(j, k)$ in (2.6.4). Here, we are looking the entries of $l_1 + 1$-th rows (as 0 corresponds the first row) of the triangular arrays corresponding to each $\pi \in \mathcal{P}_1$.

If possible, suppose the minimum occurs at two places, say, at (θ, k) and at (ϱ, l). So $\theta, \varrho \in \mathcal{P}_1$; $l_1 < k, l \leq n$ and $\beta_2 = B^{\theta}_{l_1 k} = B^{\varrho}_{l_1 l}$. Let $G = \{\theta(1), \cdots, \theta(k)\}$; $H = \{\varrho(1), \cdots, \varrho(l)\}$ and $G_2 = G \cup H$. Let $|G_2| = l_2$ and denote by $\mathcal{P}_2$, the class of all permutations of $\mathcal{P}_1$ for which $\{\pi(1), \cdots, \pi(l_2)\} = G_2$. Since π in $\mathcal{P}_1$ already maps $\{1, 2, \cdots, l_1\}$ onto G_1, this extra condition only means that π moreover maps $\{l_1 + 1, \cdots, l_2\}$ onto $G_2 - G_1$. Clearly, $\mathcal{P}_2 \subset \mathcal{P}_1$.

Claim: $B^{\pi}_{l_1 l_2} = \sqrt{\dfrac{2p_{G_2 - G_1} \log 2}{w^2_{G_2} - w^2_{G_1}}} = \beta_2$, *for every* $\pi \in \mathcal{P}_2$.

The justification of this claim is similar to that of the earlier one. Once again note that,

$$w^2_{G_2} \geq w^2_G + w^2_H - w^2_{G \cap H},$$

but

$$p_{G_2 - G_1} = p_{G - G_1} + p_{H - G_1} - p_{G \cap H - G_1}.$$

So for $\pi \in \mathcal{P}_2$, we have

$$\begin{aligned}
&2p_{G_2 - G_1} \log 2 - \beta_2^2 \left(w^2_{G_2} - w^2_{G_1}\right) \\
&\leq 2 \left(p_{G - G_1} + p_{H - G_1} - p_{G \cap H - G_1}\right) \log 2 - \beta_2^2 \left(w^2_G + w^2_H - w^2_{G \cap H} - w^2_{G_1}\right) \\
&= 2 \left(p_{G - G_1} + p_{H - G_1} - p_{G \cap H - G_1}\right) \log 2 - \\
&\qquad\qquad \beta_2^2 \left(w^2_G - w^2_{G_1} + w^2_H - w^2_{G_1} - w^2_{G \cap H} + w^2_{G_1}\right) \\
&= \left(2p_{G - G_1} \log 2 - \beta_2^2 \left(w^2_G - w^2_{G_1}\right)\right) + \left(2p_{H - G_1} \log 2 - \beta_2^2 \left(w^2_H - w^2_{G_1}\right)\right) + \\
&\qquad\qquad \left(\beta_2^2 \left(w^2_{G \cap H} - w^2_{G_1}\right) - 2p_{G \cap H - G_1} \log 2\right) \\
&= \beta_2^2 \left(w^2_{G \cap H} - w^2_{G_1}\right) - 2p_{G \cap H - G_1} \log 2 \\
&\leq 0.
\end{aligned}$$

Hence $\beta_2^2 \geq \dfrac{2p_{G_2 - G_1} \log 2}{w^2_{G_2} - w^2_{G_1}}$ and once again β_2 being the minimum over all possible choice of $\pi \in \mathcal{P}_1$ and $l_1 < j \leq n$, the only possibility left is the equality. That is $\beta_2^2 = \dfrac{2p_{G_2 - G_1} \log 2}{w^2_{G_2} - w^2_{G_1}}$ and hence the claim is proved.

If the minimum occurs at more than two places, we can use induction to conclude that there exists a unique maximal set, say, G_2, such that $G_1 \subset G_2 \subseteq I$ and the following holds. Let $|G_2| = l_2$ and $\mathcal{P}_2$ be all permutations of $\mathcal{P}_I$ that map $\{l_1 + 1, \cdots, l_2\}$ onto $G_2 - G_1$. Then $B^{\pi}_{l_1 l_2}(G_1) = \beta_2$ for all $\pi \in \mathcal{P}_2$. Of course, all the quantity l_2, β_2 depend on G_1.

Proceeding by induction can summarize:

There is a (unique) integer K with $1 \leq K \leq n$ and for every i with $1 \leq i \leq K$ there are β_i, l_i, G_i and $\mathcal{P}_i$ satisfying the following:

1. $\varnothing = G_0 \subset G_1 \subset \cdots \subset G_K = I$ with $|G_i| = l_i$ so that $1 \leq l_1 < l_2 < \cdots < l_K = n$.

2. $\mathcal{P}_i$ is the set of permutations π of I that maps $\{1, 2, \cdots, l_j\}$ onto G_j for each $j \leq i$ so that $\mathcal{P}_1 \supset \mathcal{P}_2 \supset \cdots \supset \mathcal{P}_K$.

3. $\beta_i = B^{\pi}_{l_1 l_2}$ for every $\pi \in \mathcal{P}_i$ and this common value is also same as $\min\limits_{\pi \in \mathcal{P}_{i-1}} \min\limits_{l_{i-1} < j \leq n} B^{\pi}_{l_{i-1} j}$.

So note that for any $\pi \in \mathcal{P}_K$, we can trace out the β_i for $1 \leq i \leq K$, as

$$\beta_i = \sqrt{\frac{2(p_{\pi(l_{i-1}+1)} + \cdots + p_{\pi(l_i)}) \log 2}{w^2_{\{\pi(1), \cdots, \pi(l_i)\}} - w^2_{\{\pi(1), \cdots, \pi(l_{i-1})\}}}}. \tag{3.5.6}$$

Moreover, the infimum in (3.5.2) reduces to the following:

$$\inf_{\widetilde{x} \in \Psi} \sum_{s \in S} \left(\frac{1}{2} x_s^2 - \beta a_s x_s \right), \tag{3.5.7}$$

and

$$\Psi = \left\{ \widetilde{x} : \; X_C^2 \leq 2 p_C \log 2, \; \forall (\varnothing \neq) C \subseteq I \right\} \tag{3.5.8}$$

where we used the notation

$$X_C^2 = \sum_{s \subseteq C, s \in S} x_s^2 \quad \text{and} \quad p_C = \sum_{i \in C} p_i.$$

Now we prove that, if $\beta_j \leq \beta < \beta_{j+1}$, the above infimum is attained at $\widetilde{x}^* = (x_s^*; s \in S) \in \mathbb{R}^S$ given by

$$x_s^* = \begin{cases} \beta_1 a_s & \text{if } s \subseteq G_1 \\ \beta_2 a_s & \text{if } s \subseteq G_2, s \nsubseteq G_1 \\ \vdots & \\ \beta_j a_s & \text{if } s \subseteq G_j, s \nsubseteq G_{j-1} \\ \beta a_s & \text{if } s \nsubseteq G_j. \end{cases}$$

First of all note that $\widetilde{x}^* \in \Psi$. For, $C \subseteq I$ implies

$$X_C^{*\,2}$$

$$= \sum_{s \subseteq C, s \in S} x_s^{*\,2}$$

$$
\begin{aligned}
&= \sum_{i=1}^{(j+1)\wedge K} \sum_{\substack{s\subseteq G_i, s\nsubseteq G_{i-1}\\ s\subseteq C, s\in S}} x_s^{*2} \\
&= \sum_{i=1}^{j} \sum_{\substack{s\subseteq G_i, s\nsubseteq G_{i-1}\\ s\subseteq C, s\in S}} \beta_i^2 a_s^2 \;+\; 1_{\{j+1\le K\}} \sum_{\substack{s\nsubseteq G_j\\ s\subseteq C, s\in S}} \beta^2 a_s^2 \\
&\le \sum_{i=1}^{j} \sum_{\substack{s\subseteq G_i, s\nsubseteq G_{i-1}\\ s\subseteq C, s\in S}} \beta_i^2 a_s^2 \;+\; 1_{\{j+1\le K\}} \sum_{\substack{s\nsubseteq G_j\\ s\subseteq C, s\in S}} \beta_{j+1}^2 a_s^2 \\
&= \sum_{i=1}^{j} \sum_{\substack{s\subseteq G_i, s\nsubseteq G_{i-1}\\ s\subseteq C, s\in S}} \frac{2p_{G_i-G_{i-1}}\log 2}{w_{G_i}^2 - w_{G_{i-1}}^2} a_s^2 \;+\; 1_{\{j+1\le K\}} \sum_{\substack{s\nsubseteq G_j\\ s\subseteq C, s\in S}} \frac{2p_{G_{j+1}-G_j}\log 2}{w_{G_{j+1}}^2 - w_{G_j}^2} a_s^2 \\
&\le \sum_{i=1}^{j} \sum_{\substack{s\subseteq G_i, s\nsubseteq G_{i-1}\\ s\subseteq C, s\in S}} \frac{2p_{(C\cup G_{i-1})-G_{i-1}}\log 2}{w_{C\cup G_{i-1}}^2 - w_{G_{i-1}}^2} a_s^2 \;+\; 1_{\{j+1\le K\}} \sum_{\substack{s\nsubseteq G_j\\ s\subseteq C, s\in S}} \frac{2p_{(C\cup G_j)-G_j}\log 2}{w_{C\cup G_j}^2 - w_{G_j}^2} a_s^2 \\
&\le \sum_{i=1}^{j} 2p_{(C\cup G_{i-1})-G_{i-1}}\log 2 \;+\; 1_{\{j+1\le K\}} 2p_{(C\cup G_j)-G_j}\log 2 \\
&= 2p_C \log 2.
\end{aligned}
$$

Secondly, note that for any $\widetilde{x}\in\Psi$, we have

$$\sum_{i=1}^{j} \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} \beta a_s (x_s^* - x_s) \ge \sum_{i=1}^{j} \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} x_s^*(x_s^* - x_s).$$

For, by Holder's inequality we have

$$\sum_{i=1}^{j} \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} x_s^* x_s \le \sqrt{\sum_{i=1}^{j} \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} x_s^{*2}} \sqrt{\sum_{i=1}^{j} \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} x_s^{2}} \le \sum_{i=1}^{j} \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} x_s^{*2},$$

where the last inequality follows from the fact that $\widetilde{x}\in\Psi$. Hence

$$\sum_{i=1}^{j} \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} x_s^*(x_s^* - x_s) \ge 0.$$

Since $\beta > \beta_j$, we have $\left(\frac{\beta}{\beta_i} - 1\right) > 0$ for $1 \le i \le j$. Moreover β_i being increasing in i, these numbers $\left(\frac{\beta}{\beta_i} - 1\right)$ are decreasing and hence we get,

$$\sum_{i=1}^{j} \left(\frac{\beta}{\beta_i} - 1\right) \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} x_s^*(x_s^* - x_s) \ge 0.$$

In other words using the definition of x_s^* we get the observation

$$\sum_{i=1}^{j} \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} \beta a_s(x_s^* - x_s) \ge \sum_{i=1}^{j} \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} x_s^*(x_s^* - x_s).$$

Now by using the above inequality, we have

$$\begin{aligned}
&\sum_{s\subseteq G_j} \left(\tfrac{1}{2}x_s^2 - \beta a_s x_s\right) - \sum_{s\subseteq G_j} \left(\tfrac{1}{2}x_s^{*2} - \beta a_s x_s^*\right)\\
&= \sum_{s\subseteq G_j} \left(\tfrac{1}{2}x_s^2 - \beta a_s(x_s - x_s^*) - \tfrac{1}{2}x_s^2\right)\\
&\ge \sum_{s\subseteq G_j} \left(\tfrac{1}{2}x_s^2 - x_s^*(x_s - x_s^*) - \tfrac{1}{2}x_s^2\right)\\
&= \tfrac{1}{2} \sum_{s\subseteq G_j} (x_s - x_s^*)^2 \ge 0.
\end{aligned}$$

Moreover, using the definition of x_s^* for $s \nsubseteq G_j$, we have

$$\begin{aligned}
&\sum_{s\nsubseteq G_j} \left(\tfrac{1}{2}x_s^2 - \beta a_s x_s\right) - \sum_{s\nsubseteq G_j} \left(\tfrac{1}{2}x_s^{*2} - \beta a_s x_s^*\right)\\
&= \sum_{s\nsubseteq G_j} \left(\tfrac{1}{2}x_s^2 - \beta a_s x_s + \tfrac{1}{2}\beta^2 a_s^2\right)\\
&= \tfrac{1}{2} \sum_{s\nsubseteq G_j} (x_s - \beta x_s)^2 \ge 0.
\end{aligned}$$

Thus combining the above two inequality,

$$\sum_{s\in S} \left(\frac{1}{2}x_s^2 - \beta a_s x_s\right) - \sum_{s\in S} \left(\frac{1}{2}x_s^{*2} - \beta a_s x_s^*\right) \ge 0$$

and hence the infimum occurs at $\widetilde{x}^*$.

Denote $\beta_0 = 0$ and $\beta_{K+1} = \infty$. Suppose $1 \le j \le K$ and $\beta \in [\beta_j, \beta_{j+1})$ then the infimum in (3.5.7) becomes

$$\begin{aligned}
&\sum_{i=1}^{j} \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} \left(\tfrac{1}{2}\beta_i^2 a_s^2 - \beta\beta_i a_s^2\right) + \sum_{s\nsubseteq G_j} \left(\tfrac{1}{2}\beta^2 a_s^2 - \beta^2 a_s^2\right) \\
&= \sum_{i=1}^{j} \tfrac{1}{2}\beta_i^2 \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} a_s^2 - \beta \sum_{i=1}^{j} \beta_i \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} a_s^2 - \tfrac{1}{2}\beta^2 \sum_{s\nsubseteq G_j} a_s^2 \\
&= p_{G_j} \log 2 - \beta \sum_{i=1}^{j} \beta_i \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} a_s^2 - \tfrac{1}{2}\beta^2 \sum_{s\nsubseteq G_j} a_s^2
\end{aligned}$$

We can summarize the above discussion in the following

Theorem 3.5.1. *In the Gaussian BK-GREM, almost surely,*

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \sum_{i\notin G_j} p_i \log 2 + \beta \sum_{i=1}^{j} \beta_i \sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} a_s^2 + \frac{1}{2}\beta^2 \sum_{s\nsubseteq G_j} a_s^2,$$

if $\beta \in [\beta_j, \beta_{j+1})$ *for* $0 \le j \le K$.

We shall now describe for each $\pi \in \mathcal{P}_I$ an n level GREM. In what follows $\pi \in \mathcal{P}_I$ is fixed. For the N particle system there are $2^{k(\pi(i),N)}$ furcations at the ith level, below each node of the $(i-1)$th level. The weights at the i-th level in this GREM are $w(\pi, i)$ which are defined by $w(\pi, 1) = a_{\pi(1)}$, and in general, for $1 \le i \le n$

$$w^2(\pi, i) = \sum_{\substack{s\subseteq\{\pi(1),\cdots,\pi(i)\} \\ s\nsubseteq\{\pi(1),\cdots,\pi(i-1)\}}} a_s^2. \tag{3.5.9}$$

Let $\mathcal{E}(\pi, \beta)$ be the almost sure limiting free energy of this GREM. This exists by Theorem 2.4.6. As done in subsection 2.6.1, we set $r_0^\pi = 0$ and let

$$\beta_i^\pi = \min_{k > r_{i-1}} B^\pi_{r^\pi_{i-1}, k}$$

with $r_i^\pi = \max\{l > r_{i-l} : B(r_{i-1}^\pi, l) = \beta_i\}$ for $1 \le i \le K^\pi$ with $r_{K^\pi} = n$. Also denote $\beta_0^\pi = 0$ and $\beta^\pi_{K^\pi+1} = \infty$. Then by Theorem 2.6.1, we have for $\beta \in \left[\beta_j^\pi, \beta_{j+1}^\pi\right)$ with $0 \le j \le K^\pi$,

$$\mathcal{E}(\pi, \beta) = \sum_{i=r_j^\pi+1}^{n} p_{\pi(i)} \log 2 + \frac{1}{2}\beta^2 \sum_{i=r_j^\pi+1}^{n} w^2(\pi, i) + \beta \sum_{i=1}^{r_j^\pi} \beta_i^\pi w^2(\pi, i). \tag{3.5.10}$$

Now let us consider $\pi \in \mathcal{P}_K$. Then note that $l_i = r_i^\pi$ for all $1 \le i \le K$ and by definition β_i^π is same as β_i . Hence $\sum_{i=r_j^\pi+1}^{n} p_{\pi(i)} = \sum_{i\notin G_j} p_i$; $\sum_{s\subseteq G_i, s\nsubseteq G_{i-1}} a_s^2 = w^2(\pi, i)$ and $\sum_{s\nsubseteq G_j} a_s^2 = \sum_{i=r_j^\pi+1}^{n} w^2(\pi, i)$, so that $\mathcal{E}(\beta) = \mathcal{E}(\pi, \beta)$.

Thus for every $\pi \in \mathcal{P}_K$, the GREM associated in the above paragraph has the same energy, namely, $\mathcal{E}(\beta)$, the energy of the BK-GREM.

We now go on to show that if π is any permutation then the energy of the GREM associated with π, namely $\mathcal{E}(\pi, \beta)$, is larger than $\mathcal{E}(\beta)$. So fix a permutation π.

Denote $H_i^\pi = \{s : \ s \subseteq \{\pi(1), \cdots, \pi(i)\} \& s \nsubseteq \{\pi(1), \cdots, \pi(i-1)\}\}$, that is, H_i^π consists of all subsets of $\{\pi(1), \cdots, \pi(i)\}$ that include $\pi(i)$. Then

$$\begin{aligned}
&\sum_{s\in S} \left(\frac{1}{2}x_s^2 - \beta a_s x_s\right) \\
=&\sum_{i=1}^{n} \sum_{\substack{s\in S, s\subseteq\{\pi(1),\cdots,\pi(i)\} \\ s\nsubseteq\{\pi(1),\cdots,\pi(i-1)\}}} \left(\frac{1}{2}x_s^2 - \beta a_s x_s\right) \\
\geq&\sum_{i=1}^{n} \left(\sum_{\substack{s\subseteq\{\pi(1),\cdots,\pi(i)\} \\ s\nsubseteq\{\pi(1),\cdots,\pi(i-1)\}}} \frac{1}{2}x_s^2 - \beta \left(\sum_{\substack{s\subseteq\{\pi(1),\cdots,\pi(i)\} \\ s\nsubseteq\{\pi(1),\cdots,\pi(i-1)\}}} a_s^2 \right)^{\frac{1}{2}} \left(\sum_{\substack{s\subseteq\{\pi(1),\cdots,\pi(i)\} \\ s\nsubseteq\{\pi(1),\cdots,\pi(i-1)\}}} x_s^2 \right)^{\frac{1}{2}} \right)
\end{aligned}$$

since for $C \subseteq I$, $\sum_{s\in S, s\subseteq C} a_s x_s \leq \left(\sum_{s\in S, s\subseteq C} a_s^2\right)^{\frac{1}{2}} \left(\sum_{s\in S, s\subseteq C} x_s^2\right)^{\frac{1}{2}} = w_C X_C$,

$$=\sum_{i=1}^{n} \left(\frac{1}{2}X_{H_i^\pi}^2 - \beta w(\pi, i) X_{H_i^\pi}\right).$$

Moreover, for $\pi \in \mathcal{P}_I$, let us denote

$$\begin{aligned}
\Psi_\pi &= \left\{X_{\{\pi(1),\cdots,\pi(i)\}}^2 \leq 2p_{\{\pi(1),\cdots,\pi(i)\}} \log 2, \quad 1 \leq i \leq n\right\} \\
&= \left\{\sum_{i=1}^{k} X_{H_i^\pi}^2 \leq \sum_{i=1}^{k} 2p_{\pi(i)} \log 2, \quad \forall 1 \leq k \leq n\right\} \subset \mathbb{R}^S.
\end{aligned}$$

Then $\Psi \subseteq \Psi_\pi$ for every $\pi \in \mathcal{P}_I$. Hence for every $\pi \in \mathcal{P}_I$, we have

$$\inf_{\Psi} \sum_{s\in S} \left(\frac{1}{2}x_s^2 - \beta a_s x_s\right) \geq \inf_{\Psi_\pi} \sum_{i=1}^{n} \left(\frac{1}{2}X_{H_i^\pi}^2 - \beta w(\pi, i) X_{H_i^\pi}\right)$$

and hence

$$\begin{aligned}\mathcal{E}(\beta) &= \log 2 - \inf_{\Psi} \sum_{s\in S} \left(\tfrac{1}{2}x_s^2 - \beta a_s x_s\right) \\ &\leq \log 2 - \inf_{\Psi_\pi} \sum_{i=1}^{n} \tfrac{1}{2}X_{H_i^\pi}^2 - \beta\beta w(\pi,i) X_{H_i^\pi} \quad = \mathcal{E}(\pi,\beta).\end{aligned}$$

Thus we have proved the following.

Theorem 3.5.2. *Almost surely,*

$$\mathcal{E}(\beta) = \inf_{\pi} \mathcal{E}(\beta,\pi).$$

That is, the free energy of the Gaussian BK-GREM represents the free energy of an n level tree GREM. In fact, it represents the minimum out of all possible $n!$ many n-level Gaussian tree GREM energies with appropriately defined weights.

Remark 3.5.1. A closer look of the definition of B_{ij}^{π} reveals that if $a_s = 0$ for some $s \in S$ then such an s plays no role in the definition of β_is. Moreover, since $\sum_{s\in S}\left(\frac{1}{2}x_s^2 - \beta a_s x_s\right) = \sum_{s\in S, a_s\neq 0}\left(\frac{1}{2}x_s^2 - \beta a_s x_s\right) + \sum_{s\in S, a_s=0}\frac{1}{2}x_s^2$, in calculating infimum of $\sum_{s\in S}\left(\frac{1}{2}x_s^2 - \beta a_s x_s\right)$ on ψ we will be quite justified to put $x_s = 0$ for all those $s \in S$ for which $a_s = 0$. This will lead to the calculation of infimum of $\sum_{s\in S, a_s\neq 0}\left(\frac{1}{2}x_s^2 - \beta a_s x_s\right)$ on Ψ. In other words, we could consider S to be the collection of all those increasing sequences s for which $a_s \neq 0$, instead of all sequences. This is the setup of the original Bolthausen-Kistler model.

Bolthausen and Kistler have shown that the free energy is the minimum among the free energies of the tree GREMs associated with the all possible increasing chains of subsets of I. What the above argument shows is that one need not consider all chains. It is enough to consider $n!$ many n level GREMs. Can we reduce $n!$? Perhaps not in general. Incidentally, the argument also identifies all these n level GREMs which attains the minimum. In fact, the number of such n level GREM is precisely $|\mathcal{P}_K|$, cardinality of $\mathcal{P}_K$.

Remark 3.5.2. Though the BK-model is not a hierarchial model, yet when the driving distribution is Gaussian we are not able to get out of the tree GREM. That is, the tree GREM is in some way hidden in this model.

Now one can raise the question whether going out of Gaussian driving distributions leads to a BK-GREM that is not usual tree GREM (in the sense of energy). In this regard, it is worth mentioning, that if we consider that the driving distributions $\{\mu_N^s\}_N$ satisfying LDP with rate function $\mathcal{I}_s(x_s) = \frac{1}{\gamma}x_s^\gamma$ for some $\gamma \geq 1$ and every $s \in S$ then Theorem 3.5.2 remains true. To see this we follow the same line of proof as above with the appropriate changes as done in section 2.6.

Remark 3.5.3. Large deviation approach allows us to consider different driving distributions for different $s \in S$. This can be done with BK-GREM also and one can prove the existence of free energy. But it is not easy to obtain explicit formula.

3.6 Block Tree GREM

In the previous section we have shown that in the Gaussian BK-Model the limiting free energy is the minimum over all possible $n!$ many n-level tree GREMs with appropriate weights. Now we will conclude this chapter by exhibiting one model which includes again $n!$ many n-level GREMs and where the free energy is maximum over all those GREMs. To define the model we will use the notation $n, N, \sigma = \sigma_1 \cdots \sigma_n$ with the same interpretation as that of the earlier section. Let $a_1, \cdots, a_n$ be given non-negative weights. For any sequence $s = \langle j_1, \cdots, j_i \rangle$ of distinct elements from $I = \{1, 2, \cdots, n\}$ and for any $\sigma(s) = \langle \sigma_{j_1}, \cdots, \sigma_{j_i} \rangle \in 2^{k(j_1,N)} \times \cdots \times 2^{k(j_i,N)}$. We have random variables $\xi^s_{\sigma(s)}$ and these are independent $\mathcal{N}(0, N)$. Now depending on π, a permutation of I and $\sigma \in 2^N$, we define the Hamiltonian as

$$H_N(\sigma, \pi) = \sum_{i=1}^{n} a_{\pi(i)} \xi^{\pi(1)\pi(2)\cdots\pi(i)}_{\sigma_{\pi(1)}\sigma_{\pi(2)}\cdots\sigma_{\pi(i)}}. \tag{3.6.1}$$

Note that here the configuration space has $n! \times 2^N$ many points instead of usual 2^N many. We call this model as Block tree GREM. We define the partition function corresponding to inverse temperature $\beta > 0$ as

$$Z_N(\beta) = \sum_{\pi \in \mathcal{P}_I} \sum_{\sigma \in \Sigma_N} e^{-\beta H_N(\sigma,\pi)},$$

and the definition of free energy is $\frac{1}{N} \log Z_N(\beta)$.

So for $n = 3$ the model will look like as in Figure 3.1.

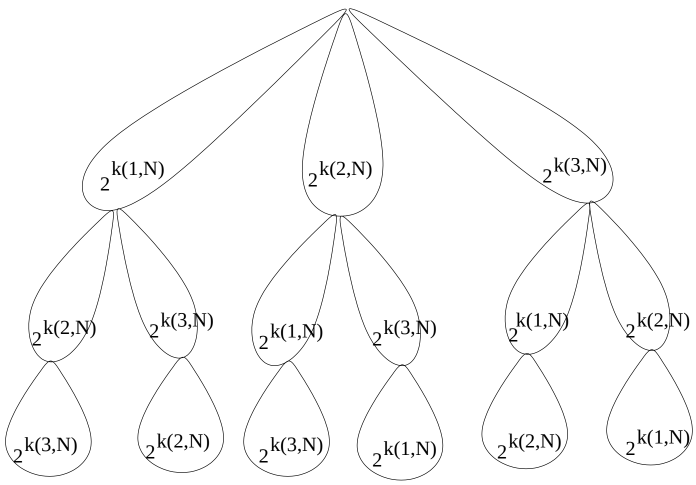


Figure 3.1: Block Tree GREM

Now for each $\pi \in \mathcal{P}_I$, let us denote

$$Z_N^\pi(\beta) = \sum_{\sigma \in \Sigma_N} e^{-\beta H_N(\sigma,\pi)}.$$

Note that for each π, Z_N^π denote the partition function for the n-level tree GREM with $2^{k(\pi(i),N)}$ furcations below each of node at the $(i-1)$-th level of the tree and with the associated weight in the i-th level being $a_{\pi(i)}$. So we can write

$$Z_N(\beta) = \sum_{\pi \in \mathcal{P}_I} Z_N^\pi(\beta).$$

Hence $\frac{1}{N}\log Z_N(\beta) = \frac{1}{N}\log \sum_{\pi\in\mathcal{P}_I} Z_N^\pi(\beta) = \frac{1}{N}\log \max_\pi Z_N^\pi(\beta) + \frac{1}{N}\log \sum_{\pi\in\mathcal{P}_I} \frac{Z_N^\pi(\beta)}{\max_\pi Z_N^\pi(\beta)}$.

Since limiting free energy exists almost surely corresponding to every π, let us denote $\mathcal{E}^\pi(\beta) = \lim_N \frac{1}{N}\log Z_N^\pi(\beta)$. Hence

$$\lim_N \frac{1}{N}\log Z_N(\beta) = \lim_N \frac{1}{N}\log \max_\pi Z_N^\pi(\beta).$$

Now log being increasing function we can bring the max out side log and the range of π being finite we can push the limit after max so that

$$\lim_N \frac{1}{N}\log Z_N(\beta) = \max_\pi \lim_N \frac{1}{N}\log Z_N^\pi(\beta) = \max_\pi \mathcal{E}^\pi(\beta).$$

Thus we have the following

Theorem 3.6.1. *In the block tree GREM, the limiting free energy $\mathcal{E}(\beta)$ exists almost surely and*

$$\mathcal{E}(\beta) = \max_{\pi} \mathcal{E}^{\pi}(\beta).$$

Remark 3.6.1. In the definition of weights, we fixed numbers $a_1, \cdots, a_n$ and weighted $\xi^{\pi(1)\pi(2)\cdots\pi(i)}_{\sigma_{\pi(1)}\sigma_{\pi(2)}\cdots\sigma_{\pi(i)}}$ with $a_{\pi(i)}$. Instead one could fix for each s, a sequence of distinct elements of I, a number a_s and then $\xi^{\pi(1)\pi(2)\cdots\pi(i)}_{\sigma_{\pi(1)}\sigma_{\pi(2)}\cdots\sigma_{\pi(i)}}$ could be weighted with $a_{\{\pi(1),\cdots,\pi(i)\}}$. Different driving distributions for different s can also be considered. Then also the above theorem remains true.

Remark 3.6.2. We now consider the weights $(a_s,\ s \subseteq I)$ as mentioned in the above remark. Using the notation of previous section, consider BK-GREM with these weights. Consider the GREM associated with $\pi \in \mathcal{P}_I$ in the BK-GREM and denotes its energy by $\mathcal{E}(\pi, \beta)$.

On the other hand, consider block tree GREM as mentioned in the above theorem with weights $\widetilde{a}_{\{\pi(1),\cdots,\pi(i)\}}$, where

$$\widetilde{a}^2_{\{\pi(1),\cdots,\pi(i)\}} = \sum_{\substack{s\subseteq\{\pi(1),\cdots,\pi(i)\} \\ s\not\subseteq\{\pi(1),\cdots,\pi(i-1)\}}} a_s^2.$$

By (3.5.9), we observe that for each $\pi \in \mathcal{P}_I$, $\widetilde{a}_{\{\pi(1),\cdots,\pi(i)\}} = w(\pi, i)$ for $1 \le i \le n$. Moreover, for a fixed $\pi \in \mathcal{P}_I$, in both the associated GREM model have $2^{k(\pi(i),N)}$ many edges at the i-th level below each node of the $(i-1)$-th level. Hence $\mathcal{E}^{\pi}(\beta) = \mathcal{E}(\pi, \beta)$ for each $\pi \in \mathcal{P}_I$. Thus the limiting free energy of this block tree GREM is larger than that of the BK-GREM.

Chapter 4

Word GREM with External Field

In this concluding chapter we discuss a more general version of random energy models, called Word GREM. This model includes Derrida's REM and GREM, also the model of Bolthausen and Kistler. Moreover the model is considered with external field. We apply this analysis to analyze the free energy of REM with external field.

4.1 Word GREM

In the previous chapters we have shown that the almost sure existence of the limiting free energy is assured through the simple LDP of certain empirical measures. This techniques is quite simple and neat. In this section, we present a general setup which includes all the models mentioned above. However, it is not just the generalization that should be noted. More importantly, we use the same large deviation technique which allows us to introduce external field in the model. To our knowledge these models are so far not discussed with external field except the REM by Derrida in [16]. Not only that, as already mentioned in the previous chapter this method allows consideration of different driving distributions. This in turn leads to diverse covariance structures for the Hamiltonian.

4.2 The Model

Let $I = \{\varsigma_1, \varsigma_2, \cdots, \varsigma_n\}$ be a set of n symbols where $n \geq 1$ is a positive integer. Let $S(I)$ be the set of all words formed by these n symbols. Let S be a finite subset of $S(I)$. So a typical word $s \in S$ of length l will look like $s = \varsigma_{i_1}\varsigma_{i_2}\cdots\varsigma_{i_l}$ where each $\varsigma_{i_j} \in I$. Occasionally we will use the symbol $s \in S$ as a word as well as a subset of I consisting of all the symbols in s. Since symbols may be repeated in a word, it is possible that two different words may correspond to the same subset of I. Moreover, without loss of generality we assume that each symbol appears in at least one word of S, that is $\bigcup_{s\in S} s = I$.

For $N \geq n$, the N particle system has configuration space, as usual, $\Sigma_N = \{+1,-1\}^N$ consisting of sequence of length N with entries $+1$ and -1. For $1 \leq i \leq n$, let $k(i,N) \geq 1$ be integers with $\sum_{i=1}^{n} k(i,N) = N$ and $\frac{k(i,N)}{N} \to p_i > 0$ as $N \to \infty$. Clearly, $\sum_1^n p_i = 1$.

For $\sigma = \langle \sigma_1, \cdots, \sigma_N \rangle \in \Sigma_N$, we denote $\sigma^1 = \langle \sigma_i : i \leq k(1,N) \rangle$, $\sigma^2 = \langle \sigma_i : k(1,N)+1 \leq i \leq k(1,N)+k(2,N) \rangle$, etc. Thus σ can also be written as $\sigma = \langle \sigma^1, \cdots, \sigma^n \rangle$. For each $s = \varsigma_{i_1}\varsigma_{i_2}\cdots\varsigma_{i_l} \in S$ and $\sigma = \langle \sigma^1, \cdots, \sigma^n \rangle$, we put, $\sigma(s) = \langle \sigma^{i_1}, \sigma^{i_2}, \cdots, \sigma^{i_l} \rangle$, $k(s,N) = \sum_{i=1}^{n} k(i,N) 1_{\{\varsigma_i \in s\}}$.

For each $s \in S$ and $\sigma \in \Sigma_N$ we have a random variables $\xi(s, \sigma(s))$. These are assumed to be independent random variables(distributions in general depend on N.) To make it more precise, denote $\Sigma_{iN} = \{+1,-1\}^{k(i,N)}$, for $1 \leq i \leq n$. For each $s = \varsigma_{i_1}\varsigma_{i_2}\cdots\varsigma_{i_l} \in S$ and $\sigma(s) = \langle \sigma^{i_1}, \cdots, \sigma^{i_l} \rangle \in \Sigma_{i_1 N} \times \cdots \times \Sigma_{i_l N}$, we have one random variable $\xi(s, \sigma(s))$. All these $\sum_{s\in S} 2^{k(s,N)}$ random variables are independent. Let us assume, for $s \in S$, all the $\xi(s,\sigma(s))$ have distribution λ_N^s on $\mathbb{R}$, that is, the distribution of $\xi(s,\sigma(s))$ depends on s but not on $\sigma(s)$. Let $f : \mathbb{R}^S \to \mathbb{R}$ be a continuous function. For the configuration $\sigma = \langle \sigma_1, \sigma_2, \cdots, \sigma_N \rangle$, the Hamiltonian of the system is defined as

$$H_N(\sigma, h) = N f(\xi(\sigma)) + h \sum_{i=1}^{N} \sigma_i, \tag{4.2.1}$$

where $\xi(\sigma) = (\xi(s,\sigma(s)))_{s\in S}$ and $h \geq 0$ is a number representing the intensity of the external field. The partition function of the system is

$$Z_N = \sum_{\sigma} e^{-\beta H_N(\sigma,h)},$$

with $\beta > 0$ being the inverse temperature. Once again the limiting free energy is $\lim_N \frac{1}{N} \log Z_N(\beta)$.

Remark 4.2.1. Observe that if $S = S_1$ consists of only one word $\varsigma_1\varsigma_2\cdots\varsigma_n$, and if $f(x) = x$ then this is just the REM. If $S = S_n$ consists of the n words $\{\varsigma_1, \varsigma_1\varsigma_2, \cdots, \varsigma_1\varsigma_2\cdots\varsigma_n\}$, and if $f((x_s)_s) = \sum_{s\in S} a_s x_s$ then this is just GREM. On the other hand, if $S = S_{2^n}$ consists of all the words $\varsigma_{i_1}\varsigma_{i_2}\cdots\varsigma_{i_l}$ with out repetition of symbols then S can clearly be identified as the collection of non-empty subsets of S. If moreover, $f((x_s)_s) = \sum_{s\in S} a_s x_s$ then this will lead to the BK-GREM. Of course, one could also take $S = S_{n!}$ consisting of all the $n!.n$ many words

$\{\varsigma_{\pi(1)}\varsigma_{\pi(2)}\cdots\varsigma_{\pi(l)} : 1 \le l \le n \,\&\, \pi$ a permutation of $\{1, 2, \cdots, n\}\}$.

Let $\overline{\sigma}^i$ denote the sum of the $k(i,N)$ many $+1$ and -1 appearing in σ^i. In other words, if $\sigma = \langle \sigma_1, \sigma_2, \cdots, \sigma_N \rangle$ then $\overline{\sigma}^i$ is the sum of all σ_j where j satisfies $k(1,N) + \cdots + k(i-1,N) + 1 \le j \le k(1,N) + \cdots k(i,N)$. Then, note from (4.2.1) that

$$\frac{H_N(\sigma,h)}{N} = f(\xi(\sigma)) + \frac{h}{N}\sum_{i=1}^{n} \overline{\sigma}^i, \tag{4.2.2}$$

and

$$Z_N(\beta,h) = 2^N E_\sigma e^{-N\beta \frac{H_N(\sigma,h)}{N}}, \tag{4.2.3}$$

where E_σ is the expectation with respect to the uniform probability on the configuration space.

Under certain assumptions we shall show that the limit $\lim\limits_N \frac{1}{N} \log Z_N(\beta)$ exists almost surely and is a non-random quantity. The essential assumptions are the following: Firstly the distributions of ξ should have exponential decay and secondly $\frac{k(i,N)}{N}$ converges.

Notations

We start with some notations which we will use in the rest of the chapter. A typical points in $\mathbb{R}^S \times \mathbb{R}^n$ will be denoted by $((x_s, s \in S), (y_i, i \le n))$ or simply as $(x_{_S}, y_{_I})$. In what follows, $\Box = \prod_{s\in S} \triangle_s \times \prod_{i=1}^n \nabla_i$ is a box in $\mathbb{R}^S \times \mathbb{R}^n$ where $\triangle_s$ for each $s \in S$ and ∇_i for $i \le n$ are open subintervals of $\mathbb{R}$.

For $A \subseteq I$, let $\mathcal{S}_A = \{s \in S : s \subseteq A\}$. So note that $\mathcal{S}_I = S$. We will denote $Q_{AN} = \prod_{s\in\mathcal{S}_A} q_{sN}$ where $q_{sN} = \lambda_N^s(\triangle_s)$. If $\mathcal{S}_A$ is empty for some A, we put $Q_{AN} = 1$. Also we will denote Q_{sN} with the same understanding as above considering s as a subset of I. Strictly speaking we should denote Q_{AN} and q_{sN} as $Q_{AN}(\Box)$ and $q_{sN}(\triangle_s)$ respectively, but for ease of writing we are not doing so. If $A = \{\varsigma_{i_1}, \varsigma_{i_2}, \cdots, \varsigma_{i_m}\}$, then sometimes we need only the indices $\{i_1, i_2, \cdots, i_m\}$ and we will denote them by $[A]$. For $A \subseteq I$, we denote $k(A,N) = \sum\limits_{i=1}^{n} k(i,N) 1_{\{\varsigma_i \in A\}} = \sum\limits_{i\in[A]} k(i,N)$ and $\alpha_{AN} = \frac{1}{2^{k(A,N)}} \sum\limits_{\langle\sigma^i :\, i\in[A]\rangle} \prod\limits_{i\in[A]} 1_{\nabla_i}(\frac{\overline{\sigma}^i}{N})$. We want to point out once again that time to time we will consider $s \in S$ as a subset of I. For example, if $s = \varsigma_{i_1}\varsigma_{i_2}\cdots\varsigma_{i_l} \in S$, we will use the notation $\alpha_{sN} = \frac{1}{2^{k(s,N)}} \sum_{\sigma(s)} \prod_{j=1}^{l} 1_{\nabla_{i_j}}\left(\frac{\overline{\sigma}^{i_j}}{N}\right)$.

4.3 A large deviation principle

For each $s \in S$, let us consider a probability λ^s on $\mathbb{R}$. If X is distributed like λ^s, let us denote $\Lambda_s(\rho) = \log E e^{\rho X}$ and $\mathcal{D}_{\Lambda_s} = \{\rho : \Lambda_s(\rho) < \infty\}$. Note that $0 \in \mathcal{D}_{\Lambda_s}$, but we want that $0 \in \mathcal{D}^0_{\Lambda_s}$. So from now on we will focus our attention on those λ^s for which 0 is an interior point in $\mathcal{D}_{\Lambda_s}$. As $0 \in \mathcal{D}^0_{\lambda^s}$, the mean $\overline{x}_s = \int x d\lambda^s(dx)$ exists and is finite quantity for each $s \in S$. Now if $X_1^s, X_2^s, \cdots$ are i.i.d. random variables having distribution λ^s, we will consider λ_N^s to be the law of $\frac{1}{N}(X_1^s + X_2^s + \cdots + X_N^s)$. Now by Cramer's theorem (Theorem 0.3.5) the sequence $\{\lambda_N^s\}$ satisfies large deviation principle with a good, convex rate function $\mathcal{I}_s$ given by $\mathcal{I}_s(x) = \sup_{\rho \in \mathbb{R}}\{\rho x - \Lambda_s(\rho)\}$. Note that this is a convex, good, non-negative lower semicontinuous function. Moreover, by property of good rate function, $\mathcal{I}_s(\overline{x}_s) = 0$ for every $s \in S$ so that the set $\mathcal{I}_s(x) < \alpha$ is non-empty for every $\alpha > 0$. We also want to point out that the functions $\mathcal{I}_s$ are increasing on $[\overline{x}_s, \infty)$ and decreasing on $(-\infty, \overline{x}_s]$.

Once again by Cramer's theorem, the arithmetic averages of i.i.d. mean zero, ± 1 valued random variables satisfy LDP with rate function $\mathcal{I}_0$ where $\mathcal{I}_0(y) = \infty$ for $|y| > 1$; $\mathcal{I}_0(\pm 1) = \log 2$ and for $-1 < y < 1$,

$$\begin{aligned} \mathcal{I}_0(y) &= y \tanh^{-1} y - \log\cosh(\tanh^{-1} y) \\ &= \tfrac{1+y}{2}\log(1+y) + \tfrac{1-y}{2}\log(1-y). \end{aligned} \tag{4.3.1}$$

Let us define the map from $\Sigma_N \to \mathbb{R}^S \times \mathbb{R}^n$ as follows:

$$\sigma = \left(\sigma^1, \sigma^2, \cdots, \sigma^n\right) \mapsto \left(\left(\xi(s, \sigma(s)),\, s \in S\right), \left(\frac{\overline{\sigma}^i}{N},\, 1 \le i \le n\right)\right),$$

where $\overline{\sigma}^i$ is the sum of the entries of σ^i.

Thus for each ω (sample point of the random variables ξ, which is suppressed so far), this map transports the uniform probability on Σ_N to $\mathbb{R}^S \times \mathbb{R}^n$. Denote this induced random probability by μ_N. Hence from (4.2.3), we have,

$$\frac{1}{N}\log Z_N(\beta, h) = \log 2 - \frac{1}{N}\log \int_{\mathbb{R}^S \times \mathbb{R}^n} e^{-N\beta\left(f(x_S) + h\sum_{i=1}^n y_i\right)} d\mu_N(x_S, y_I). \tag{4.3.2}$$

Proposition 4.3.1. *If for some $A \subseteq I$, $\sum_{N \ge n} 2^{k(A,N)} Q_{AN}\alpha_{AN} < \infty$ then almost surely eventually $\mu_N(\square) = 0$.*

Proof. Let A be such that $\sum_{N\geq n} 2^{k(A,N)} Q_{AN}\alpha_{AN} < \infty$. Then

$$\begin{aligned}
\mu_N(\Box) &= \frac{1}{2^N}\sum_{\sigma}\prod_{s\in S} 1_{\triangle_s}\left(\xi(s,\sigma(s))\right)\prod_{i\leq n} 1_{\triangledown_i}\left(\frac{\overline{\sigma}^i}{N}\right) \\
&\leq \frac{1}{2^N}\sum_{\sigma}\prod_{s\in\mathcal{S}_A} 1_{\triangle_s}\left(\xi(s,\sigma(s))\right)\prod_{i\in[A]} 1_{\triangledown_i}\left(\frac{\overline{\sigma}^i}{N}\right) \\
&= \frac{1}{2^{k(A,N)}}\sum_{\sigma^i:i\in[A]}\prod_{s\in\mathcal{S}_A} 1_{\triangle_s}\left(\xi(s,\sigma(s))\right)\prod_{i\in[A]} 1_{\triangledown_i}\left(\frac{\overline{\sigma}^i}{N}\right)
\end{aligned}$$

As a consequence,

$$\begin{aligned}
P(\mu_N(\Box) > 0) &= P\left(\sum_{\sigma_i:i\in[A]}\prod_{s\in\mathcal{S}_A} 1_{\triangle_s}\left(\xi(s,\sigma(s))\right)\prod_{i\in[A]} 1_{\triangledown_i}\left(\frac{\overline{\sigma}^i}{N}\right)\geq 1\right) \\
&\leq Q_{AN}\sum_{\sigma^i:i\in[A]}\prod_{i\in[A]} 1_{\triangledown_i}\left(\frac{\overline{\sigma}^i}{N}\right) \\
&= 2^{k(A,N)} Q_{AN}\alpha_{AN}.
\end{aligned}$$

The hypothesis and Borel-Cantelli lemma completes the proof. ☐

Let us note that, $E\mu_N(\Box) = Q_{IN}\alpha_{IN}$.

Proposition 4.3.2. *If for all non-empty $A\subseteq I$, $\sum_{N\geq n} 2^{-k(A,N)} Q_{AN}^{-1}\alpha_{AN}^{-1} < \infty$ then for all $\epsilon > o$, almost surely eventually,*

$$(1-\epsilon)Q_{IN}\alpha_{IN} \leq \mu_N(\Box) \leq (1+\epsilon)Q_{IN}\alpha_{IN}.$$

That is

$$(1-\epsilon)E\mu_N(\Box) \leq \mu_N(\Box) \leq (1+\epsilon)E\mu_N(\Box).$$

Proof. Note that

$$
\begin{aligned}
&\mathrm{Var}(\mu_N(\Box)) \\
=&\frac{1}{2^{2N}}\sum_{\sigma}\sum_{\tau} E\left(\prod_{s\in S} 1_{\triangle_s}\left(\xi(s,\sigma(s))\right) 1_{\triangle_s}\left(\xi(s,\tau(s))\right)\right)\prod_{i\le n} 1_{\triangledown_i}\left(\frac{\overline{\sigma}^i}{N}\right) 1_{\triangledown_i}\left(\frac{\overline{\tau}^i}{N}\right) \\
&\qquad - Q_{IN}^2\alpha_{IN}^2 \\
\le&\frac{1}{2^{2N}}\sum_{\substack{A\subseteq I\\ A\neq\phi}}\sum_{\sigma}\sum_{\substack{\tau_i=\sigma_i,\forall i\in[A]\\ \tau_i\neq\sigma_i,\forall i\in[A^c]}} E\left(\prod_{s\in S} 1_{\triangle_s}\left(\xi(s,\sigma(s))\right) 1_{\triangle_s}\left(\xi(s,\tau(s))\right)\right)\times \\
&\qquad \prod_{i\le n} 1_{\triangledown_i}\left(\frac{\overline{\sigma}^i}{N}\right) 1_{\triangledown_i}\left(\frac{\overline{\tau}^i}{N}\right)
\end{aligned}
$$

(since $Q_{IN}^2\alpha_{IN}^2$ cancels the terms corresponding to $\sigma_i\neq\tau_i$, $\forall i\in[I]$)

$$
=\frac{1}{2^{2N}}\sum_{\substack{A\subseteq I\\ A\neq\phi}}\frac{Q_{IN}^2}{Q_{AN}}\sum_{\sigma}\sum_{\substack{\tau_i=\sigma_i,\forall i\in[A]\\ \tau_i\neq\sigma_i,\forall i\in[A^c]}}\prod_{i\in[A]} 1_{\triangledown_i}\left(\frac{\overline{\sigma}^i}{N}\right)\prod_{i\in[A^c]} 1_{\triangledown_i}\left(\frac{\overline{\sigma}^i}{N}\right) 1_{\triangledown_i}\left(\frac{\overline{\tau}^i}{N}\right)
$$

(by definition of Q_{AN}, $A\subseteq I$)

$$
=\sum_{\substack{A\subseteq I\\ A\neq\phi}}\frac{Q_{IN}^2}{Q_{AN}}\frac{1}{2^{k(A,N)}}\frac{\alpha_{IN}^2}{\alpha_{AN}}.
$$

Now by Chebyshev's inequality for any $\epsilon>0$

$$
P(|\mu_N(\Box)-E\mu_N(\Box)|>\epsilon E\mu_N(\Box))<\frac{1}{\epsilon^2}\sum_{A\subseteq I}\frac{1}{2^{k(A,N)}Q_{AN}\alpha_{AN}}.
$$

Once again Borel-Cantelli lemma and the hypothesis yield that a.s. eventually,

$$
(1-\epsilon)E\mu_N(\Box)\le\mu_N(\Box)\le(1+\epsilon)E\mu_N(\Box).
$$

Hence the proof. ∎

Theorem 4.3.3. *For a.e. ω, the sequence $\{\mu_N(\omega), N \geq n\}$ satisfies LDP with rate function $\mathcal{J}$ given as follows:*

$$\mathcal{D}_{\mathcal{J}} = \left\{ (x_{_S}, y_{_I}) : \; \forall A \subseteq I, \sum_{t \in S_A} \mathcal{I}_t(x_t) + \sum_{i \in [A]} p_i \mathcal{I}_0 \left(\frac{y_i}{p_i}\right) \leq \sum_{k \in [A]} p_k \log 2 \right\}$$

and

$$\begin{aligned} \mathcal{J}(x_{_S}, y_{_I}) &= \sum_{s \in S} \mathcal{I}_s(x_s) + \sum_{i \in [I]} p_i \mathcal{I}_0 \left(\frac{y_i}{p_i}\right) \quad \textit{if } (x_{_S}, y_{_I}) \in \mathcal{D}_{\mathcal{J}} \\ &= \infty \qquad\qquad\qquad\qquad\qquad\qquad \textit{otherwise} \end{aligned}$$

Proof. In what follows, A denotes a non empty subset of I.

First of all note that, as $\mathcal{I}_0$ and $\mathcal{I}_s$ for $s \in S$ are convex, good rate functions, $\mathcal{D}_{\mathcal{J}}$ is a convex compact set.

Now let $\Box = \prod_{s \in S} \triangle_s \times \prod_{i=1}^n \nabla_i$ be an open box in $\mathbb{R}^S \times \mathbb{R}^n$ where $\triangle_s$ for each $s \in S$ and ∇_i for $i \leq n$ are subintervals of $\mathbb{R}$ with rational end points.

<u>**Step 1**</u> Suppose that closure of $\Box$ is disjoint with $\mathcal{D}_{\mathcal{J}}$, that is $\mathcal{D}_{\mathcal{J}} \cap \overline{\Box} = \phi$. In other words, for every $(x_S, y_I) \in \overline{\Box}$, there exists an $A \subseteq I$(depending on (x_S, y_I)) so that $\sum_{t \in S_A} \mathcal{I}_t(x_t) + \sum_{i \in [A]} p_i \mathcal{I}_0 \left(\frac{y_i}{p_i}\right) > \sum_{k \in [A]} p_k \log 2$. We shall show, almost surely eventually $\mu_N(\Box) = 0$.

Note that as $\mathcal{I}_0$ and $\mathcal{I}_s$ are lower semicontinuous functions for every $s \in S$ and $\overline{\nabla}_i$ and $\overline{\triangle}_s$ are compact sets, we can get $(x_S^0, y_I^0) \in \overline{\Box}$ so that $\mathcal{I}(y_i^0) = \mathcal{I}_0(\nabla_i)$ for $1 \leq i \leq n$ and $\mathcal{I}(x_s^0) = \mathcal{I}_s(\triangle_s)$ for every $s \in S$.

For this point $(x_S^0, y_I^0) \in \overline{\Box}$ there exists an $A \subseteq I$ so that $\sum_{t \in S_A} \mathcal{I}_t(x_t^0) + \sum_{i \in [A]} p_i \mathcal{I}_0 \left(\frac{y_i^0}{p_i}\right) > \sum_{k \in [A]} p_k \log 2$. We will prove that for this A the hypothesis of Proposition 4.3.1 is satisfied and hence for this $\Box$ almost surely eventually $\mu_N(\Box) = 0$ leading to $\lim_{N \to \infty} \frac{1}{N} \log \mu_N(\Box) = -\infty$.

Since $\{\lambda_N^s\}_N$ satisfies LDP with rate function $\mathcal{I}_s$, we have

$$\limsup \frac{1}{N} \log \lambda_N^s(\triangle_s) \leq -\mathcal{I}_s(\overline{\triangle}_s).$$

Let $\epsilon > 0$, to be chosen later. For all large N,

$$\frac{1}{N}\log \lambda_N^s(\triangle_s) < -\mathcal{I}_s(\overline{\triangle}_s) + \epsilon = -\mathcal{I}_s(x_s^0) + \epsilon,$$

that is $q_{sN} = \lambda_N^s(\triangle_s) < e^{-N(\mathcal{I}_s(x_s^0)-\epsilon)}$ eventually. And this is true for every $s \in S_A$. So eventually

$$Q_{AN} < e^{-N\sum\limits_{s\in S_A}(\mathcal{I}_s(x_s^0)-\epsilon)}.$$

Similarly, the law of $\frac{\overline{\sigma}^i}{N}$ satisfies LDP with rate $p_i\mathcal{I}_0(\frac{y_i}{p_i})$ and hence we will have eventually

$$\alpha_{AN} < e^{-N\sum\limits_{i\in[A]}(p_i\mathcal{I}_0(\frac{y_i^0}{p_i})-\epsilon)}.$$

Thus

$$2^{k(A,N)}Q_{AN}\alpha_{AN} < e^{-N\left[\sum\limits_{s\in S_A}(\mathcal{I}_s(x_s^0)-\epsilon)+\sum\limits_{i\in[A]}\left(p_i\mathcal{I}_0(\frac{y_i^0}{p_i})-\epsilon-\frac{k(i,N)}{N}\log 2\right)\right]}.$$

Now as $\frac{k(i,N)}{N} \to p_i$ and we have strict inequality in $\sum\limits_{t\in S_A}\mathcal{I}_t(x_t^0) + \sum\limits_{i\in[A]} p_i\mathcal{I}_0\left(\frac{y_i^0}{p_i}\right) > \sum\limits_{i\in[A]} p_i \log 2$, we can choose an ϵ so that

$$\sum_{N\geq n} 2^{k(A,N)}Q_{AN}\alpha_{AN} < \infty.$$

Hence by Proposition 4.3.1 we have, almost surely $\mu_N(\Box) = 0$. It is not difficult to see now that almost surely μ_N is eventually supported on a compact set.

Step 2 Let us now consider a $\Box$ which has non-empty intersection with $\mathcal{D}_{\mathcal{J}}$. We show that for this $\Box$, almost surely,

$$-\left[\sum_{s\in S}\mathcal{I}_s(\triangle_s)+\sum_{1\le i\le n}p_i\mathcal{I}_0(\frac{1}{p_i}\triangledown_i)\right]\le\liminf_{N\to\infty}\frac{1}{N}\log\mu_N(\square)$$
$$\le\limsup_{N\to\infty}\frac{1}{N}\log\mu_N(\square) \tag{4.3.3}$$
$$\le-\left[\sum_{s\in S}\mathcal{I}_s(\overline{\triangle}_s)+\sum_{1\le i\le n}p_i\mathcal{I}_0(\frac{1}{p_i}\overline{\triangledown}_i)\right].$$

Using LDP, we have

$$\liminf_{N}\frac{1}{N}\log\lambda^s_N(\triangle_s)\ge-\mathcal{I}_s(\triangle_s).$$

Hence for $\epsilon>0$ eventually,

$$\frac{1}{N}\log\lambda^s_N(\triangle_s)>-\mathcal{I}_s(\triangle_s)-\epsilon \tag{4.3.4}$$

for every $s\in S$. Moreover, eventually,

$$\frac{1}{N}\log P\left(\frac{\overline{\sigma}^i}{N}\in\triangledown_i\right)>-p_i\mathcal{I}_0\left(\frac{\triangledown_i}{p_i}\right)-\epsilon \tag{4.3.5}$$

for every $i\le n$. Hence for every $A\subseteq I$,

$$2^{-K(A,N)}Q^{-1}_{AN}\alpha^{-1}_{AN}=e^{-N\left(\sum\limits_{i\in[A]}\frac{k(i,N)}{N}\log 2+\sum\limits_{s\in S_A}\frac{1}{N}\log\lambda^s_N(\triangle_s)+\sum\limits_{i\in[A]}\frac{1}{N}\log P(\frac{\overline{\sigma}^i}{N}\in\triangledown_i)\right)}$$
$$<e^{-N\left[\sum\limits_{i\in[A]}\frac{k(i,N)}{N}\log 2-\sum\limits_{s\in S_A}(\mathcal{I}_s(\triangle_s)+\epsilon)-\sum\limits_{i\in[A]}\left(p_i\mathcal{I}_0\left(\frac{\triangledown_i}{p_i}\right)+\epsilon\right)\right]},$$

by (4.3.4) and (4.3.5).

As $\mathcal{D}_{\mathcal{J}}$ is a convex set and $\square$ is an non-empty open set, there exists at least one

(x_S^0, y_I^0) in $\mathcal{D}_{\mathcal{J}}^- \cap \Box$, where

$$\mathcal{D}_{\mathcal{J}}^- = \left\{ (x_{_S}, y_{_I}) : \ \forall A \subseteq I, \sum_{t \in S_A} \mathcal{I}_t(x_t) + \sum_{i \in [A]} p_i \mathcal{I}_0 \left(\frac{y_i}{p_i} \right) < \sum_{k \in [A]} p_k \log 2 \right\}.$$

Being a point in $\mathcal{D}_{\mathcal{J}}^-$, for every $A \subseteq I$, we have

$$\sum_{s \in S_A} \mathcal{I}_s(x_s^0) + \sum_{i \leq n} p_i \mathcal{I}_0 \left(\frac{y_i^0}{p_i} \right) < \sum_{i \in A} p_i \log 2.$$

That is

$$\sum_{s \in S_A} \mathcal{I}_s(\triangle_s) + \sum_{i \leq n} p_i \mathcal{I}_0 \left(\frac{\nabla_i}{p_i} \right) < \sum_{i \in A} p_i \log 2.$$

The above being a strict inequality, we can choose ϵ depending on $\Box$ so that for every $A \subseteq I$ the quantity

$$2^{-K(A,N)} Q_{AN}^{-1} \alpha_{AN}^{-1}$$

is summable over N. Now Proposition 4.3.2 yields (4.3.3). This completes step 2.

Towards step 3, let $\mathcal{A}$ be the collection of all open boxes $\Box$ with rational corner points satisfying either $\overline{\Box} \cap \mathcal{D}_{\mathcal{J}} = \emptyset$ or $\Box \cap \mathcal{D}_{\mathcal{J}}^- \neq \emptyset$. This collection is so rich that they form a base for the topology of $\mathbb{R}^S \times \mathbb{R}^I$. Note that $\mathcal{A}$ being a countable family, out side a grand null set, for every $\Box$ in $\mathcal{A}$ conclusions of Step 1 and Step 2 hold. In the next two steps, we show

$$\mathcal{J}(x_S, y_I) = \sup_{\Box : (x_S, y_I) \in \Box} \{ - \liminf_N \frac{1}{N} \log \mu_N(\Box) \} \tag{4.3.6}$$

$$= \sup_{\Box : (x_S, y_I) \in \Box} \{ - \limsup_N \frac{1}{N} \log \mu_N(\Box) \}. \tag{4.3.7}$$

<u>**Step 3**</u> Let $(x_S, y_I) \notin \mathcal{D}_{\mathcal{J}}$. Then $\mathcal{D}_{\mathcal{J}}$ being a closed set we can find a $\Box \in \mathcal{A}$ containing (x_S, y_I) so that $\overline{\Box}$ does not intersect with $\mathcal{D}_{\mathcal{J}}$. By Step 1, $\mu_N(\Box) = 0$ eventually so that $\lim_N \frac{1}{N} \log \mu_N(\Box) = \infty$. Also by definition of $\mathcal{J}$, we have $\mathcal{J}(x_S, y_I) = \infty$. Hence the above equalities hold when $(x_S, y_I) \notin \mathcal{D}_{\mathcal{J}}$.

Step 4 Now let $(x_S^0, y_I^0) \in \mathcal{D}_{\mathcal{J}}$ and $\mathcal{A}_{(x_S^0,y_I^0)} = \{\Box \in \mathcal{A} : (x_S^0, y_I^0) \in \Box\}$. Then, as observed in Step 2, for every $\Box \in \mathcal{A}_{(x_S^0,y_I^0)}$ we have eventually

$$\begin{aligned} -\left[\sum_{s\in S}\mathcal{I}_s(\triangle_s) + \sum_{1\le i\le n} p_i\mathcal{I}_0(\frac{1}{p_i}\nabla_i)\right] &\le \liminf_{N\to\infty}\frac{1}{N}\log\mu_N(\Box) \\ &\le \limsup_{N\to\infty}\frac{1}{N}\log\mu_N(\Box) \qquad (4.3.8) \\ &\le -\left[\sum_{s\in S}\mathcal{I}_s(\overline{\triangle}_s) + \sum_{1\le i\le n} p_i\mathcal{I}_0(\frac{1}{p_i}\overline{\nabla}_i)\right]. \end{aligned}$$

From the first part of the above inequality we have,

$$\liminf_N \frac{1}{N}\log\mu_N(\Box) \ge -\mathcal{J}(\Box).$$

And hence

$$\sup_{\Box\in\mathcal{A}_{(x_S^0,y_I^0)}}\{-\liminf_N\frac{1}{N}\log\mu_N(\Box)\} \le \sup_{\Box\in\mathcal{A}_{(x_S^0,y_I^0)}}\mathcal{J}(\Box) \;\le\; \mathcal{J}(x_S^0, y_I^0). \qquad (4.3.9)$$

On the other hand, for every $\Box \in \mathcal{A}_{(x_S^0,y_I^0)}$ using the right side inequality of (4.3.8), we have

$$\limsup_N \frac{1}{N}\log\mu_N(\Box) \le -\mathcal{J}(\overline{\Box}).$$

Let $\mathcal{A}'_{(x_S^0,y_I^0)} = \{\Box_k \in \mathcal{A} : k \ge 1\}$ be a subclass of $\mathcal{A}$ so that $\overline{\Box}_{k+1} \subset \Box_k$ for every k and $\cap_k\Box_k = \{(x_S^0, y_I^0)\}$. Then

$$\begin{aligned} \sup_{\Box\in\mathcal{A}_{(x_S^0,y_I^0)}}\{-\limsup_N\frac{1}{N}\log\mu_N(\Box)\} &\ge \sup_{\Box\in\mathcal{A}_{(x_S^0,y_I^0)}}\mathcal{J}(\overline{\Box}) \\ &\ge \sup_{\Box\in\mathcal{A}'_{(x_S^0,y_I^0)}}\mathcal{J}(\overline{\Box}) \\ &= \lim_k \mathcal{J}(\overline{\Box}_k) \\ &= \mathcal{J}(x_S^0, y_I^0). \qquad (4.3.10) \end{aligned}$$

The last equality follows as $\mathcal{J}$ is a good lower semicontinuous function (see Proposition 0.3.2).

Thus Step 3 and Step 4 complete the proof of (4.3.6) and (4.3.7).

Now proof of Theorem 4.3.3 is completed by appealing to Proposition 0.3.3 and observing that $\{\mu_N\}$ is eventually supported on a compact set. □

Remark 4.3.1. A closer look at the above proof shows that the convexity of the rate functions I_s is not an essential condition. Good rate functions on real line whose graph look like 'U' will suffice. That is we could take those non-negative functions $\mathcal{I}$ for which we will get at most two points $\underline{x} \leq \overline{x}$ so that $\mathcal{I}$ is zero on $[\underline{x}, \overline{x}]$; strictly decreasing on $(-\infty, \underline{x}] \cap \{\mathcal{I} \in (0,\infty)\}$ and strictly increasing on $[\overline{x}, \infty) \cap \{\mathcal{I} \in (0,\infty)\}$. With these conditions we can find (x_S^0, y_I^0) as stated in Step 1 and Step 2 for the proof to go through keeping all other Steps as it is.

For each s, we started with a distribution λ^s and appealed to Cramer's theorem (see the first paragraph of this section). Instead, we could start with $(\lambda_N^s,\ N \geq n)$ and assume that $\{\lambda_N^s\}$ satisfies LDP with rate function $\mathcal{I}_s$ having the properties mentioned in the above paragraph.

Since almost every sequence of probabilities μ_N is eventually supported on a compact set, we could use Varadhan's lemma with any continuous function. This will lead to the following (see equation (4.3.2)):

Theorem 4.3.4. *In the word GREM, almost surely*

$$\lim_N \frac{1}{N} \log Z_N(\beta, h) = \log 2 - \inf_{\mathcal{D}_\mathcal{J}} \left\{ \beta f(x_S) + \beta h \sum_{i=1}^{n} y_i + \mathcal{J}(x_S, y_I) \right\}.$$

Though we have taken any real valued continuous function f on $\mathbb{R}^S$, it is customary to consider $f(x_S) = \sum_{s\in S} a_s x_S$ so that the Hamiltonian becomes $H_N(\sigma) = N \sum_{s\in S} a_s \xi(s, \sigma(s)) + h\overline{\sigma}$ where $a_s,\ s \in S$ are non-negative weights and $h > 0$ is the strength of the external field. It is also customary to consider Gaussian driving distribution. In this setup if λ^s is standard normal, the above theorem will be applicable and will reduce to the following

Corollary 4.3.5. *In the Gaussian word GREM with external field, almost surely, the limiting free energy is*

$$\log 2 - \inf_{\mathcal{D}_{\mathcal{J}}} \left\{ \sum_{s\in S} \left(\tfrac{1}{2}x_s^2 + \beta a_s x_s\right) + \sum_{i=1}^{n} \left(\tfrac{p_i+y_i}{2} \log \tfrac{p_i+y_i}{p_i} + \tfrac{p_i-y_i}{2} \log \tfrac{p_i-y_i}{p_i} + \beta h y_i \right) \right\},$$

where $\mathcal{D}_{\mathcal{J}}$ is the set consisting of $(x_S, y_I) \in \mathbb{R}^S \times \mathbb{R}^n$ such that $\forall A \subseteq I$,

$$\sum_{s\in S_A} \tfrac{1}{2}x_s^2 + \sum_{i\in[A]} \left(\tfrac{p_i+y_i}{2} \log \tfrac{p_i+y_i}{p_i} + \tfrac{p_i-y_i}{2} \log \tfrac{p_i-y_i}{p_i} \right) \leq \sum_{k\in[A]} p_k \log 2.$$

Hence the use of large deviation techniques not only ensures the almost sure existence of the limiting free energy, the calculation of free energy of the system is then reduced to that of an optimization problem. Of course, it is not always possible to solve this optimization problem to arrive at a closed form expression when the external field is present or different driving distributions are considered for different $s \in S$. Even for $n = 2$ with Gaussian driving distribution it is difficult to obtain a closed form expression. The only case where we will get some 'closed' form expression is the case for $n = 1$ - that is the case of REM with external fields. This we will consider in the next section. But with no external field the situation is not that worse. In some of the cases, the method of calculation of the infimum will just reduce to what we did in section 3.5. In that case, though the model, to start with, was not an n level tree GREM, it reduced (as far as the free energy is concerned) to an n level tree GREM with appropriate weights [see §3.5]. It is quite conceivable that the present complicated model may always be equivalent to a tree GREM. We do not think so.

4.4 REM with external field

As mentioned in the last section, there is no general technique of obtaining a formula for the word GREM as well as tree GREM energy with external field. We will discuss here the simple REM with external field. Let us consider the word GREM where S consists of only one word, that is, S consists of the word $\varsigma_1\varsigma_2\cdots\varsigma_n$ where $I = \{\varsigma_1, \varsigma_2, \cdots, \varsigma_n\}$ (see beginning of the previous section). In such a case the word GREM reduces exactly to the usual REM with external field. Thus the Hamiltonian is

$$H_N(\sigma) = aN\xi_\sigma + h\sum_{i=1}^{N} \sigma_i$$

where ξ_σ are i.i.d. random variables (for each fixed N) and h, a are positive constants. Moreover for the Gaussian REM, ξ_σ are $\mathcal{N}(0, \frac{1}{N})$. Then by Corollary 4.3.5, the limiting free energy for the REM with external field exists almost surely and is given by

$$\begin{aligned}\mathcal{E}(\beta,h) &= \log 2 - \inf_{\mathcal{D}_{\mathcal{J}}}\left\{\tfrac{x^2}{2} + \tfrac{1+y}{2}\log(1+y) + \tfrac{1-y}{2}\log(1-y) + \beta(ax+hy)\right\} \\ &= \log 2 - \inf_{\mathcal{D}_{\mathcal{J}}}\left\{\tfrac{x^2}{2} + \mathcal{I}_0(y) + \beta(ax+hy)\right\},\end{aligned}$$

where $\mathcal{I}_0$ is given by (4.3.1) and

$$\begin{aligned}\mathcal{D}_{\mathcal{J}} &= \{(x,y) : \tfrac{x^2}{2} + \tfrac{1+y}{2}\log(1+y) + \tfrac{1-y}{2}\log(1-y) \leq \log 2\} \\ &= \{(x,y) : \tfrac{x^2}{2} + \mathcal{I}_0(y) \leq \log 2\}.\end{aligned}$$

In other words,

$$\mathcal{E}(\beta,h) = \log 2 - \inf_{\mathcal{D}_{\mathcal{J}}^+} f(x,y),$$

where $f(x,y) = \left\{\tfrac{x^2}{2} + \mathcal{I}_0(y) - \beta(ax+hy)\right\}$ and $\mathcal{D}_{\mathcal{J}}^+$ equals all points of $\mathcal{D}_{\mathcal{J}}$ with both coordinates non-negative.

To calculate the above infimum, first fix β, h and y with $0 \leq y \leq 1$. Then the range of x is $0 \leq x \leq \sqrt{2[\log 2 - \mathcal{I}_0(y)]}$. It is easy to see that if $\mathcal{I}_0(y) \leq \log 2 - \frac{1}{2}\beta^2 a^2$ then the $\inf_x f(x,y)$ is attained for $x = \beta a$ and if $\mathcal{I}_0(y) > \log 2 - \frac{1}{2}\beta^2 a^2$ then the infimum is attained for $x = \sqrt{2[\log 2 - \mathcal{I}_0(y)]}$. Since $\mathcal{I}_0$ is a non-negative function, the set $\{\mathcal{I}_0(y) \leq \log 2 - \frac{1}{2}\beta^2 a^2\}$ will be non-empty only when $\beta \leq \frac{1}{a}\sqrt{2\log 2}$. For $\beta > \frac{1}{a}\sqrt{2\log 2}$, we always have $\mathcal{I}_0(y) > \log 2 - \frac{1}{2}\beta^2 a^2$ so that the infimum is attained for $x = \sqrt{2[\log 2 - \mathcal{I}_0(y)]}$. Substituting these values of x in $f(x,y)$ we obtain the following expression for the infimum of $f(x,y)$ over x. First we need a notation. For $\beta \leq \frac{1}{a}\sqrt{2\log 2}$, let c_β be the solution of

$$\mathcal{I}_0(c_\beta) = \log 2 - \frac{1}{2}\beta^2 a^2. \tag{4.4.1}$$

Then

$$\varphi(y) = \inf_{0\leq x\leq\sqrt{2[\log 2 - \mathcal{I}_0(y)]}} f(x,y) = \begin{cases} g_1(y) & \text{if } \beta \leq \frac{1}{a}\sqrt{2\log 2} \text{ and } y \leq c_\beta, \\ g_2(y) & \text{if } \beta \leq \frac{1}{a}\sqrt{2\log 2} \text{ and } y > c_\beta, \\ g_2(y) & \text{if } \beta > \frac{1}{a}\sqrt{2\log 2}, \end{cases} \tag{4.4.2}$$

where

$$g_1(y) = -\frac{1}{2}\beta^2 a^2 + \mathcal{I}_0(y) - \beta h y$$

and

$$g_2(y) = \log 2 - \beta a\sqrt{2[\log 2 - \mathcal{I}_0(y)]} - \beta h y.$$

Since

$$g_1'(y) = \tanh^{-1}(y) - \beta h,$$

we have $g_1'(0) = -\beta h$ and

$$g_1'(y) \lesseqqgtr 0 \Leftrightarrow y \lesseqqgtr \tanh(\beta h).$$

On the other hand, as

$$g_2'(y) = \frac{\beta a \tanh^{-1}(y)}{\sqrt{2[\log 2 - \mathcal{I}_0(y)]}} - \beta h,$$

we have $g_2'(0) = -\beta h$ and by (4.4.1), $g_2'(c_\beta) = -\beta h + \tanh^{-1}(c_\beta)$. Thus $g_2'(c_\beta) \leq 0$ iff $c_\beta \leq \tanh(\beta h)$. Moreover,

$$g_2'(y) \lesseqqgtr 0 \Leftrightarrow \frac{a \tanh^{-1}(y)}{\sqrt{2[\log 2 - \mathcal{I}_0(y)]}} \lesseqqgtr h.$$

Let y_0 be the non-negative solution of

$$\frac{a \tanh^{-1}(y)}{\sqrt{2[\log 2 - \mathcal{I}_0(y)]}} = h. \tag{4.4.3}$$

Such a solution always exists since $\log 2 - \mathcal{I}_0(y) \to 0$ as $y \to 1$.

Since $\mathcal{I}_0$ is a strictly increasing function of $[0,1]$, from equations (4.4.1) and (4.4.3), we note that

$$\tanh(\beta h) \lesseqqgtr y_0 \Leftrightarrow y_0 \lesseqqgtr c_\beta.$$

Now if $\beta \leq \frac{1}{a}\sqrt{2\log 2}$ and $y_0 \leq c_\beta$ then $\tanh(\beta h) \leq y_0 \leq c_\beta$ and the function φ in (4.4.2) is decreasing up to $y = \tanh(\beta h)$ and then increasing. In such case, the $\inf_{0\leq y\leq 1} \varphi(y)$ will occur at $y = \tanh(\beta h)$ so that

$$\inf_{0\leq y\leq 1} \varphi(y) = -\frac{1}{2}\beta^2 a^2 - \log\cosh(\beta h).$$

On the other hand, if $\beta \leq \frac{1}{a}\sqrt{2\log 2}$ and $y_0 > c_\beta$ then $c_\beta < y_0 < \tanh(\beta h)$ and the function φ is decreasing up to $y = y_0$ and then increasing. In such case, the $\inf_{0\leq y\leq 1} \varphi(y)$ will occur at $y = y_0$ so that

$$\inf_{0\leq y\leq 1} \varphi(y) = \log 2 - \beta a x_0 - \beta h y_0,$$

where $x_0 = \sqrt{2[\log 2 - \mathcal{I}_0(y)]} = \frac{a \tanh^{-1} y_0}{h}$.

Finally, if $\beta > \frac{1}{a}\sqrt{2\log 2}$ then the function φ is decreasing up to $y = y_0$ and then increasing. Hence in this case, the $\inf_{0 \le y \le 1} \varphi(y)$ will occur at $y = y_0$ so that

$$\inf_{0 \le y \le 1} \varphi(y) = \log 2 - \beta a x_0 - \beta h y_0,$$

where $x_0 = \sqrt{2[\log 2 - \mathcal{I}_0(y)]} = \frac{a \tanh^{-1} y_0}{h}$.

We can summarize the above discussion in the following:

Theorem 4.4.1. *In the Gaussian REM with external field, the limiting free energy exists almost surely and given by*

$$\mathcal{E}(\beta, h) = \begin{cases} \log 2 + \frac{\beta^2 a^2}{2} + \log\cosh(\beta h) & \text{if } \beta \le \frac{1}{a}\sqrt{2\log 2} \text{ and } y_0 \le c_\beta \\ \beta\,(a x_0 + h y_0) & \text{otherwise,} \end{cases}$$

where y_0 be the non-negative solution of $\frac{a \tanh^{-1}(y)}{\sqrt{2[\log 2 - \mathcal{I}_0(y)]}} = h$, $\mathcal{I}_0(y) = y \tanh^{-1} y - \log\cosh(\tanh^{-1} y)$, c_β is the solution of $\mathcal{I}_0(c_\beta) = \log 2 - \frac{1}{2}\beta^2 a^2$ and $x_0 = \frac{a \tanh^{-1} y_0}{h}$.

Note that the case 'otherwise' in the theorem above consists of $\beta \le \frac{1}{a}\sqrt{2\log 2}$ and $y_0 > c_\beta$ or if $\beta > \frac{1}{a}\sqrt{2\log 2}$.

Theorem 4.4.1 provides yet another justification for the phase diagram (FIG. 3) in [16] of Derrida.

Bibliography

[1] M. Aizenman, J. L. Lebowitz, and D. Ruelle, *Some rigorous results on the Sherrington-Kirkpatrick spin glass model*, Comm. Math. Phys. **112** (1987), no. 1, 3–20.

[2] F. Antonelli and M. Isopi, *Limit behaviour of the partition function of spin glasses via stochastic calculus*, citeseer.ist.psu.edu/55348.html, 1997.

[3] E. Bolthausen and N. Kistler, *On a nonhierarchical version of the generalized random energy model*, Ann. Appl. Probab. **16** (2006), no. 1, 1–14.

[4] A. Bovier, *Statistical mechanics of disordered systems*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2006, A mathematical perspective.

[5] A. Bovier and I. Kurkova, *Derrida's generalised random energy models. I. Models with finitely many hierarchies*, Ann. Inst. H. Poincaré Probab. Statist. **40** (2004), no. 4, 439–480.

[6] ______, *Derrida's generalized random energy models. II. Models with continuous hierarchies*, Ann. Inst. H. Poincaré Probab. Statist. **40** (2004), no. 4, 481–495.

[7] A. Bovier, I. Kurkova, and M. Löwe, *Fluctuations of the free energy in the REM and the p-spin SK models*, Ann. Probab. **30** (2002), no. 2, 605–651.

[8] D. Capocaccia, M. Cassandro, and P. Picco, *On the existence of thermodynamics for the generalized random energy model*, J. Statist. Phys. **46** (1987), no. 3-4, 493–505.

[9] P. Carmona and Y. Hu, *Universality in Sherrington-Kirkpatrick's spin glass model*, arXiv:math.PR/0403359 v2, 2004.

[10] S. Chatterjee, *A simple invariance theorem*, arXiv:math.PR/0508213 v1, 2005.

[11] F. Comets, *The martingale method for mean-field disordered systems at high temperature*, Mathematical aspects of spin glasses and neural networks, Progr. Probab., vol. 41, Birkhäuser Boston, Boston, MA, 1998, pp. 91–113.

[12] F. Comets and J. Neveu, *The Sherrington-Kirkpatrick model of spin glasses and stochastic calculus: the high temperature case*, Comm. Math. Phys. **166** (1995), no. 3, 549–564.

[13] P. Contucci, M. Degli Esposti, C. Giardinà, and S. Graffi, *Thermodynamical limit for correlated Gaussian random energy models*, Comm. Math. Phys. **236** (2003), no. 1, 55–63.

[14] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, second ed., Applications of Mathematics (New York), vol. 38, Springer-Verlag, New York, 1998.

[15] B. Derrida, *Random-energy model: limit of a family of disordered models*, Phys. Rev. Lett. **45** (1980), no. 2, 79–82.

[16] ———, *Random-energy model: an exactly solvable model of disordered systems*, Phys. Rev. B (3) **24** (1981), no. 5, 2613–2626.

[17] B. Derrida, *A generalization of the random energy model which includes correlations between energies*, Journal de Physique Lettres **46** (1985), 401–407.

[18] B. Derrida and E. Gardner, *Solution of the generalized random energy model*, J. Phys. C **19** (1986), 2253–2274.

[19] T. C. Dorlas and W. M. B. Dukes, *Large deviation approach to the generalized random energy model*, J. Phys. A **35** (2002), no. 20, 4385–4394.

[20] T. C. Dorlas and J. R. Wedagedera, *Large deviations and the random energy model*, Internat. J. Modern Phys. B **15** (2001), no. 1, 1–15.

[21] S. F. Edwards and P. W. Anderson, *Theory of spin glasses*, J. Phys. F **5** (1975), 965.

[22] T. Eisele, *On a third-order phase transition*, Comm. Math. Phys. **90** (1983), no. 1, 125–159.

[23] J. Fröhlich and B. Zegarliński, *Some comments on the Sherrington-Kirkpatrick model of spin glasses*, Comm. Math. Phys. **112** (1987), no. 4, 553–566.

[24] A. Galves, S. Martínez, and P. Picco, *Fluctuations in Derrida's random energy and generalized random energy models*, J. Statist. Phys. **54** (1989), no. 1-2, 515–529.

[25] S. Ghirlanda and F. Guerra, *General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity*, J. Phys. A **31** (1998), no. 46, 9149–9155.

[26] F. Guerra and F. L. Toninelli, *The thermodynamic limit in mean field spin glass models*, Comm. Math. Phys. **230** (2002), no. 1, 71–79.

[27] T C Hu, F. Móricz, and R. L. Taylor, *Strong laws of large numbers for arrays of rowwise independent random variables*, Acta Math. Hungar. **54** (1989), no. 1-2, 153–162.

[28] N. K. Jana, *Exponential random energy model*, arXiv:math.PR/0602670, 2005.

[29] N. K. Jana and B. V. Rao, *Generalized random energy model*, J. Stat. Phys. **123** (2006), no. 5, 1033–1058.

[30] ______, *Generalized random energy model. II*, J. Stat. Phys. **127** (2007), no. 4, 841–850.

[31] F. Koukiou, *A random covering interpretation for the phase transition of the random energy model*, J. Statist. Phys. **60** (1990), no. 5-6, 669–674.

[32] F. Koukiou and P. Picco, *Poisson point processes, cascades, and random coverings of* $\mathbf{R}^n$, J. Statist. Phys. **62** (1991), no. 1-2, 481–489.

[33] M. Mézard, G. Parisi, and M. Angel Virasoro, *Spin glass theory and beyond*, World Scientific Lecture Notes in Physics, vol. 9, World Scientific Publishing Co. Inc., Teaneck, NJ, 1987.

[34] E. Olivieri and P. Picco, *On the existence of thermodynamics for the random energy model*, Comm. Math. Phys. **96** (1984), no. 1, 125–144.

[35] D. Ruelle, *A mathematical reformulation of Derrida's REM and GREM*, Comm. Math. Phys. **108** (1987), no. 2, 225–239.

[36] A. Ruzmaikina and M. Aizenman, *Characterization of invariant measures at the leading edge for competing particle systems*, Ann. Probab. **33** (2005), no. 1, 82–113.

[37] D. Sherrington, *Spin glasses: a perspective*, arXiv:cond-mat/0512425v2, 2006.

[38] D. Sherrington and S. Kirkpatrick, *Solvable model of a spin-glass*, Phys. Rev. Lett. **35** (1975), 1792.

[39] M. Talagrand, *Replica symmetry breaking and exponential inequalities for the Sherrington-Kirkpatrick model*, Ann. Probab. **28** (2000), no. 3, 1018–1062.

[40] ______, *On the high temperature region of the Sherrington-Kirkpatrick model*, C. R. Acad. Sci. Paris Sér. I Math. **332** (2001), no. 2, 177–182.

[41] ______, *On the high temperature phase of the Sherrington-Kirkpatrick model*, Ann. Probab. **30** (2002), no. 1, 364–381.

[42] ______, *Spin glasses: a challenge for mathematicians*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics], vol. 46, Springer-Verlag, Berlin, 2003, Cavity and mean field models.

[43] S. R. S. Varadhan, *Asymptotic probabilities and differential equations*, Comm. Pure Appl. Math. **19** (1966), 261–286.

[44] ———, *Large Deviations and Applications*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 46, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1984.