Journal of Statistical Physics, Vol. 123, No. 5, June 2006 (© 2006)
DOI: 10.1007/s10955-006-9043-9

Generalized Random Energy Model

N. K. Jana[1] and B. V. Rao[2]

Received July 12, 2005; accepted February 15, 2006
Published Online: May 14, 2006

This paper applies the large deviation technique to a general tree formulation of Generalized Random Energy Model. The similarities and dissimilarities between the exponential and the Gaussian model are discussed. Formulae for limiting free energies, characterization of energy functions are obtained. Limiting case of the tree structure and randomization of the trees are also investigated.

KEY WORDS: Spin glasses, Large Deviation Principle, Random Tree, Free Energy.

1. INTRODUCTION

In the random energy model (REM)[(14)] 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 structure of the Hamiltonians. Of course, very little can be achieved by assuming an arbitrary covariance matrix. An n-level tree structure was suggested by Derrida,[(7)] where the branches of the tree are in correspondence with the configuration space. In this paper we discuss a reformulation of this model. There are two essential differences from what is usually considered in the literature. First, we provide a general framework of trees. Second, we split the number of particles N into n groups rather than splitting the number of spins (or 'factorizing' 2 as considered in the literature). We have independent identically distributed random variables, one with each node of the tree corresponding to the N particle system. We read along a branch and associate the resulting n-tuple (here, n is the height of the tree) of numbers with the branch. The uniform probability on the set of branches is transported to $\mathbb{R}^n$. We first establish certain basic inequalities for these random probabilities. The frame work

[1] Department of Mathematics, Bijoy Krishna Girls' College, Howrah and Stat-Math Unit, Indian Statistical Institute, Kolkata, India; e-mail: nabin_r@isical.ac.in
[2] Stat-Math Unit, Indian Statistical Institute, Kolkata, India; e-mail: bvrao@isical.ac.in

of trees allows us to introduce a further randomization at the tree level, like Poisson trees and multinomial trees. The above inequalities carry over to these random trees as well.

Then we specialize to the exponential and Gaussian driving distributions. The basic inequalities lead to the large deviation principle (LDP) for the random probabilities mentioned above and gives an explicit formula for the free energy. Though large deviation techniques are well known in statistical mechanics,[10] in the context of the present systems they are of recent origin.[8,9] Our treatment, hopefully, is elegant and notationally less cumbersome than in ref. 8. For the exponential GREM, the driving distribution does not depend on the number of particles. This does not make it less interesting. The present treatment clearly brings out the similarities between the two cases. In fact, the Gaussian case is no more complicated than the exponential case. 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. 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,[5] while it is not true here. In this work, we have not considered the limiting Gibbs' measures. They are worked out in detail by Bovier and Kurkova[1,2] for the Gaussian GREM. For Gaussian REM it is well understood (see for example[14]) and for exponential REM it is studied in ref. 13.

The organization of the paper is as follows: We formulate the framework in Sec. 2. The exponential and Gaussian GREMs are discussed in Secs. 3 and 4 respectively. Section 5 contains several illuminating remarks.

We thank Professor S. R. S. Varadhan for drawing our attention to ref. 8 at the early stage of this work. We thank the referees for helpful comments.

2. THE SETUP

We formulate GREM as follows. 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_{iN}, $1 \leq i \leq n$ with each $k_{iN} \geq 1$ and $\sum_i k_{iN} = N$. The configuration space 2^N, naturally splits into product, $\prod 2^{k_{iN}}$ and $\sigma \in 2^N$ can be written as $\sigma_1\sigma_2 \ldots \sigma_n$ with $\sigma_i \in 2^{k_{iN}}$. An obvious tree structure can be brought in the configuration space. Imagine an n-level tree. There are $2^{k_{1N}}$ nodes at the first level. These will be denoted as σ_1, for

$\sigma_1 \in 2^{k_{1N}}$. Below each of the first level nodes there are $2^{k_{2N}}$ nodes at the second level. The second level nodes below σ_1 of the first level will be denoted by $\sigma_1\sigma_2$ for $\sigma_2 \in 2^{k_{2N}}$. In general, below a node $\sigma_1\sigma_2 \ldots \sigma_{i-1}$ of the $(i-1)$-th level there are $2^{k_{iN}}$ nodes at the i-th level denoted by $\sigma_1\sigma_2 \ldots \sigma_{i-1}\sigma_i$ for $\sigma_i \in 2^{k_{iN}}$. Thus a typical branch of the tree reads like $\sigma_1\sigma_2 \ldots \sigma_n$. Obviously the branches are in one one correspondence with 2^N, the configuration space. At the node $\sigma_1 \ldots \sigma_i$, we place a random variable $\xi_{\sigma_1} \ldots \sigma_i$. 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 $\sigma = \sigma_1 \ldots \sigma_n$ the Hamiltonian is

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

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. 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.

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. We have not found any special trees that give rise to interesting phenomena, but it appears possible.

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 $\sigma_1\sigma_2 \ldots \sigma_{i-1}$ of the $(i-1)$-th level, $\sigma_1\sigma_2 \ldots \sigma_{i-1}\sigma_i$ is a typical node at the i-th level. We shall now define some useful quantities associated with the tree. Let $\sum_N$ be the set of all branches $\sigma_1\sigma_2 \ldots \sigma_n$ 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 $\sigma_1\sigma_2 \ldots \sigma_i$ of the i-th level, let $e(\sigma_1\sigma_2 \ldots \sigma_i)$ denote the number of nodes at the n-th level below the node $\sigma_1\sigma_2 \ldots \sigma_i$. Equivalently, $e(\sigma_1\sigma_2 \ldots \sigma_i)$ is the total number of branches extending $\sigma_1\sigma_2 \ldots \sigma_i$. Clearly, $\sum_{\sigma_1,\ldots,\sigma_i} e(\sigma_1, \ldots, \sigma_i) = B_N$ for any i. Let $s_{iN}^2 = \sum_{\sigma_1,\ldots,\sigma_i} e^2(\sigma_1, \ldots, \sigma_i)$.

Assume that $\xi_{\sigma_1,\ldots,\sigma_i}$ is a symmetric random variable associated with node $\sigma_1\sigma_2 \ldots \sigma_i$. 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 $\sum_N \to \mathbb{R}^n$ defined by

$$\sigma \mapsto \xi_\sigma = (\xi_{\sigma_1}, \xi_{\sigma_1\sigma_2}, \ \ldots\ , \xi_{\sigma_1\ldots\sigma_n}).$$

Let μ_N be the induced probabilit on $\mathbb{R}^n$ when $\sum_N$ has uniform distribution, that is, each $\sigma \in \sum_N$ 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 $\Delta = \Delta_1 \times \cdots \times \Delta_n$ with each $\Delta_t \subseteq \mathbb{R}$ then

$$\mu_N(\Delta) = \frac{1}{B_N} \sum_{<\sigma_1\ldots\sigma_n>} \prod_{i=1}^{n} \mathbf{1}_{\Delta_i}(\xi_{\sigma_1\sigma_2\ldots\sigma_i}).$$

Here now is the basic result.

Theorem 1. *Let $\Delta = \Delta_1 \times \cdots \times \Delta_n \subseteq \mathbb{R}^n$. Denote $q_{iN} = P(\xi \in \Delta_i)$ for $1 \leq i \leq n$.*

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

$$(1-\epsilon)\mathrm{E}\mu_N(\Delta) \leq \mu_N(\Delta) \leq (1+\epsilon)\mathrm{E}\mu_N(\Delta).$$

b) *If $\sum_{N\geq 1} B_{iN} q_{1N} \ldots q_{iN} < \infty$ for some i, $1 \leq i \leq n$ then a.s. eventually, $\mu_N(\Delta) = 0$.*

Since all the $\xi_{\sigma_1\ldots\sigma_i}$ (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.

Proof.

a) We follow Capocaccia *et al.*[3]

$$\begin{aligned}
&var(\mu_N(\Delta)) \\
&= \mathbf{E}(\mu_N(\Delta))^2 - (\mathbf{E}\mu_N(\Delta))^2 \\
&= \frac{1}{B_N^2} \sum_{\substack{\sigma_1\ldots\sigma_n \\ \tau_1\ldots\tau_n}} \left[\mathbf{E} \prod_{i=1}^{n} \mathbf{1}_{\Delta_i}(\xi_{\sigma_1\ldots\sigma_i}) \mathbf{1}_{\Delta_i}(\xi_{\tau_1\ldots\tau_i}) - q_{1N}^2 \cdots q_{nN}^2\right]
\end{aligned}$$

$$\leq \frac{1}{B_N^2}\sum_{j=1}^{n}\sum_{\sigma_1\ldots\sigma_j}\sum_{\substack{\sigma_{j+1}\ldots\sigma_n\\ \tau_{j+1}\ldots\tau_n\\ \sigma_{j+1}\neq\tau_{j+1}}} \mathbf{E}\prod_{i=1}^{j}\mathbf{1}_{\Delta_i}(\xi_{\sigma_1\ldots\sigma_i})\prod_{i=j+1}^{n}$$

$$\times\ \mathbf{1}_{\Delta_i}(\xi_{\sigma_1\ldots\sigma_i})\mathbf{1}_{\Delta_i}(\xi_{\tau_1\ldots\tau_i})$$

$$\leq \frac{1}{B_N^2}\sum_{j=1}^{n} q_{1N}\ldots q_{jN}q_{(j+1)N}^2\ldots q_{nN}^2\sum_{\sigma_1\ldots\sigma_j} e^2(\sigma_1\ldots\sigma_j)$$

$$= \frac{1}{B_N^2}\sum_{j=1}^{n} q_{1N}\ldots q_{jN}q_{(j+1)N}^2\ldots q_{nN}^2 s_{jN}^2$$

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

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

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

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

b) We follow Dorlas and Dukes.[8] Let j_0 be such that $\sum_{N\geq 1} B_{j_0 N}\, q_{1N}\ldots q_{j_0 N} < \infty$. Then

$$\mu_N(\Delta) = \frac{1}{B_N}\sum_{\sigma_1\ldots\sigma_n}\prod_{i=1}^{n}\mathbf{1}_{\Delta_i}(\xi_{\sigma_1\ldots\sigma_i})$$

$$\leq \frac{1}{B_N}\sum_{\sigma_1\ldots\sigma_{j_0}}\prod_{i=1}^{j_0}\mathbf{1}_{\Delta_i}(\xi_{\sigma_1\ldots\sigma_i})e(\sigma_1\ldots\sigma_{j_0})$$

$$= G_N, \text{(say)}.$$

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

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

Now by Chebyshev's inequality,

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

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

For GREM type regular trees the condition above will simplify as follows. This result is in ref. 8 though not explicitly stated.

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

a) If $\sum_{N \ge 1} 2^{-(k_{1N}+\cdots+k_{iN})} q_{1N}^{-1} \cdots q_{iN}^{-1} < \infty$ for each $i = 1, \cdots, n$, then for any $\epsilon > 0$ a.s. eventually,

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

b) If $\sum_{N \ge 1} 2^{k_{1N}+\cdots+k_{iN}} q_{1N} \ldots q_{iN} < \infty$ for some i, $1 \le i \le n$ then a.s. eventually, $\mu_N(\Delta) = 0$.

In Corollary 1, we fixed integers k_{iN}, $1 \le i \le n$ such that $\sum_i k_{iN} = N$. Then we considered a deterministic tree which has $2^{k_{iN}}$ nodes at the i-th level below each node of the $(i-1)$-th level. This can be called GREM setup with parameter $\tilde{k}$, where $\tilde{k}$ is the sequence of vectors $\{(k_{iN} : 1 \le i \le n), N \ge n\}$. It is not necessary to have this exactly satisfied, it can hold either approximately or on an average. These can be interpreted in several ways. For instance, "on an average" could mean any of the interpretations below.

Let us introduce another randomness at the tree level which is independent of the randomness of the Hamiltonians. Consider, for each N, independent random variables $L_{1N}, \ldots, L_{nN}$ where $L_{iN} \sim P(2^{k_{iN}})$, i.e. a Poisson random variable with parameter $2^{k_{iN}}$. Consider a random tree with $(1 + L_{iN})$ nodes at the i-th level below each node of the $(i-1)$-th level. Here 1 is added to L_{iN} to take care of the situation $L_{iN} = 0$. Of course this is also a regular tree, but random, and could be called regular Poisson tree. The corresponding GREM model can be called a *regular Poisson tree GREM* with parameter $\tilde{k}$. The next result says that if the same conditions as in Corollary 1 hold then even with randomization of tree, the conclusion holds for almost every tree sequence.

Corollary 2. *Consider a regular Poisson tree GREM with parameter $\tilde{k}$.*

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

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

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

Proof. a) 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 1(a) holds for almost every tree sequence.

We prove a stronger statement, namely, $\mathbf{E}_T \sum_{N \geq n} \frac{s_{iN}^2}{B_N^2 q_{1N} \ldots q_{iN}} < \infty$ for each 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 sufficies to show

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

But $s_{iN}^2 = \prod_{j=1}^{i} (1 + L_{jN}) \prod_{j=i+1}^{n} (1 + L_{jN})^2$ and $B_N^2 = \prod_{j=1}^{n} (1 + L_{jN})^2$. Using independence of the random variables $(L_{jN}, 1 \leq j \leq n)$, we get

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

Since $L_{jN} \sim P(2^{k_{jN}})$, a simple calculation shows

$$\mathbf{E} \left(\frac{1}{1 + L_{jN}} \right) \leq 2^{-k_{jN}} \tag{3}$$

Substituting (3) in (2) we get (1).

b) is simpler. □

The reader might be wondering why the above is called regular Poisson tree model. It is conceivable to use independent Poisson variables at each of the nodes. As in Corollary 2, let $\{k_{1N}, \ldots, k_{nN}\}$ be a partition of N. Unlike in that corollary, now consider an n-level tree with $P(2^{k_{iN}}) + 1$ many nodes below each of the nodes at the $(i - 1)$-th level. Here all the Poisson random variables are independent. The corresponding model can be called a true *Poisson tree GREM* with parameter $\tilde{k}$. Even for this model the same conclusions as above hold.

Corollary 3. *Consider a Poisson tree GREM with parameter $\tilde{k}$. Then (a) and (b) of Corollary 2 hold.*

Proof. The proof is routine and involves too much notation. We only give the outline. First observe the following.

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

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

and

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

To prove (a), as in Corollary 2 it suffices to show that for each i,

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

For this we need some notation to describe the random tree for the N-particle system. Let $L_0 \sim P(2^{k_{1N}})$. For $1 \le \sigma_1 \le L_0+1$, let $L_{\sigma_1} \sim P(2^{k_{2N}})$. In general, for $\sigma_1\sigma_2\cdots\sigma_i$, with $1 \le \sigma_1 \le L_0+1$, $1 \le \sigma_2 \le L_{\sigma_1}+1, \cdots, 1 \le \sigma_i \le L_{\sigma_1\cdots\sigma_{i-1}}^{+1}$ let $L_{\sigma_1\cdots\sigma_i} \sim P(2^{k_{(i+1)N}})$. With this notation

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

Now to estimate its expectation, first condition on all random variables up to σ_{n-2} level and use the first inequality above. Continue this process, noting that from level σ_i onward, the second inequality takes over giving the required result.

Part (b) is again straight forward. □

This leads to the same conclusion as in this set-up as well. This tree is regular only with a very small probability.

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. There are several choices, but here we deal with only one such. We fix $p_i > 0$ for $1 \le i \le n$ with $\sum_1^n p_i = 1$. Consider an n-faced die with p_i being the chance of face i appearing in a throw. Roll the die N times and let K_{iN} be the number of times face i appears. Clearly, $k_{iN} \ge 0$ and $\sum_{i=1}^n K_{iN} = N$. We can consider GREM with parameter $\tilde{K}$. This can be called a *multinomial tree GREM* with parameter $\tilde{p} = (p_1, \ldots, p_n)$. For this randomization also we have a result similar to the above.

Corollary 4. *Consider a multinomial tree GREM with parameter $\tilde{p}$.*

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

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

b) If $\sum_{N\ge1}^{\frac{N}{2\log 2}(p_1+\cdots+p_i)} q_{1N}\cdots q_{iN} < \infty$ for some i, $1 \le i \le n$ then for a.e. tree sequence, a.s. eventually, $\mu_N(\Delta) = 0$.

Proof. a) As in Corollary 2, it suffice to show that

$$\mathbf{E}_T\left(\frac{s_{iN}^2}{B_N^2}\right) \leq 2^{-\frac{N}{2\log 2}(p_1+\cdots+p_i)}$$

for each i. Since $s_{iN}^2 = 2^{k_{1N}+\ldots+k_{iN}} 2^{2(K_{(i+1)N}+\ldots+k_{nN})}$ and $B_N^2 = 2^{2\sum_{j=1}^n k_{jN}}$, we need to show

$$\mathbf{E}2^{-\sum_1^i K_{jN}} \leq 2^{-\frac{N}{2\log 2}(p_1+\cdots+p_i)}.$$

Using the fact that $\sum_1^i K_{jN}$ is binomial with parameters N and $\sum_1^i p_j$ a simple calculation shows

$$\mathbf{E}2^{-\sum_1^i K_{jN}} = \left(1 - \frac{1}{2}\sum_1^i p_j\right)^N \leq e^{-\frac{N}{2}\sum_1^i p_j} = 2^{-\frac{N}{2\log 2}\sum_1^i p_j}.$$

b) is simpler. □

Remark 2.1. *The reader would have noticed the difference in the hypothesis of (a) and (b) in the above corollary. More specifically, there is a factor 1/2 extra in the exponent of 2 in part (a). But, of course, it should also be pointed out that we have not assumed any relation between the distribution of N-th and (N + 1)-th trees either. For instance one could imagine a sequence of independent throws of the die and take $\tilde{K}_N$ as the outcome of the first N throws. We shall return to this in Remark 5.3*

Remark 2.2. *Going back to Theorem 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 1(a) holds for (T_N) iff it holds for $(\tilde{T}_N)$. Accordingly, the conclusion of Theorem 1(a) holds for (T_N) iff it holds for $(\tilde{T}_N)$. Same remark applies for Theorem 1(b). This is what we meant when we said earlier that the hypothesis of Corollary 1 need not hold exactly, enough if it holds approximately.*

Remark 2.3. *Under suitable conditions – for instance, when $\sum_N e^{k_{iN}} < \infty$ for all i – it is possible to show that almost every tree sequence ceases to be regular after some stage in the Poisson tree model.*

To proceed further we need two well known results on LDP. The second result is a variant of Varadhan's lemma. See Dembo and Zeitouni[6] for details.

Proposition 1. *Let S be a polish space with an open base $\mathcal{A}$, and $\{\mu_N\}$ be a sequence of probabilities on S. Suppose that, for each $A \in \mathcal{A}$, $\lim_{N\to\infty} \frac{1}{N} \log \mu_N(A)$ exists and equals, $-L_A$ (say). Define $I(x) = \sup\{L_A : x \in A \in \mathcal{A}\}$. Assume that I* *is supported on a compact set, that is, $I(x) = \infty$ outside a compact set. Then the sequence $\{\mu_N\}$ satisfies LDP with rate function I.*

Proposition 2. *Suppose $\{\mu_N\}$ is a sequence of probabilities on a polish space S satisfying the LDP with rate function I. Assume that $\{\mu_N\}$ is eventually supported on a compact set C. Let $f : S \to \mathbb{R}$ be a continuous function. Then denoting E_N for expectation under μ_N,*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathrm{E}_N e^{-Nf} = -\inf_{x\in C}\{f(x) + I(x)\}.$$

3. EXPONENTIAL GREM

In this section, we consider GREM where each ξ is a double exponential variable with parameter 1. In other words, ξ has density

$$\phi(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < \infty.$$

Note that the density does not depend on N. Let $N \geq n$ and let $k_{1N}, \ldots, k_{nN}$ be integers ≥ 1 adding up to N. The random probabilities μ_N are defined on $\mathbb{R}^n$ by transporting the uniform distribution of $2^N = 2^{k_{1N}} \times \cdots \times 2^{k_{nN}}$ to $\mathbb{R}^n$ via the map

$$\sigma \mapsto \left(\frac{\xi_{\sigma_1}(\omega)}{N}, \frac{\xi_{\sigma_1\sigma_2}(\omega)}{N}, \ldots, \frac{\xi_{\sigma_1\ldots\sigma_n}(\omega)}{N}\right).$$

In evaluating the free energy, we will be applying Varadhan's lemma (Proposition 2 above). This explains the factor $\frac{1}{N}$ in the above map, which was not present in the general framework of Theorem 1.

Proposition 3. *$\mu_N \Rightarrow \delta_0$ a.s. as $N \to \infty$.*

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

$$\mathbf{P}(\mu_N(\Delta^c(\epsilon)) > \epsilon) < \frac{1}{\epsilon}\mathbf{E}\mu_N(\Delta^c(\epsilon)) < \frac{n}{\epsilon}\mathbf{P}(|\xi| > \epsilon N) = \frac{n}{\epsilon}e^{-\epsilon N}.$$

The proposition now follows from Borel-Cantelli lemma. □

From now on we assume that $\frac{k_{iN}}{N} \to p_i > 0$ for $1 \leq i \leq n$. Clearly, $\sum p_i = 1$. Let

$$\Psi = \{\tilde{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n : \sum_{i=1}^{k} |x_i| \leq \sum_{i=1}^{k} p_i \log 2, 1 \leq k \leq n\}.$$

Let $\Delta = \Delta_1 \times \cdots \times \Delta_n$ be a non-empty open rectangle of $\mathbb{R}^n$. For such Δ and $1 \leq i \leq n$ define $m_i = \inf_{x \in \Delta_i} |x|$ and $M_i = \sup_{x \in \Delta_i} |x|$. Clearly, $m_i < \infty$ for all i. Observe that in case $m_i > 0$ then $\Delta_i \subseteq (-M_i, -m_i) \cup (m_i, M_i)$ and in case $m_i = 0$ then $\Delta_i \subseteq (-M_i, M_i)$. In any case $\Delta_i \subseteq (-M_i, -m_i] \cup [m_i, M_i)$ for each i. Let $\tilde{m} = (m_1, \ldots, m_n)$. Also define $q_{iN} = P(\frac{\xi}{N} \in \Delta_i)$, for $1 \leq i \leq n$.

We have the following two observations:

$$q_{iN} \leq \int_{Nm_i}^{NM_i} e^{-x} dx \leq \int_{Nm_i}^{\infty} e^{-x} dx = e^{-Nm_i} \tag{4}$$

and

$$q_{iN} \geq \frac{1}{2} \int_{Nm_i}^{NM_i} e^{-x} dx > \frac{1}{2} \int_{Nm_i}^{N(m_i+\delta)} e^{-x} dx > \frac{N\delta}{2} e^{-N(m_i+\delta)}, \tag{5}$$

for any $0 < \delta < M_i - m_i$. Both (4) and (5) remain true even if $m_i = 0$.

Proposition 4. *If $\bar{\Delta} \cap \Psi = \phi$ then a.s. eventually $\mu_N(\Delta) = 0$. Moreover, the sequence $\{\mu_N\}$ is supported on a compact set.*

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

$$\sum_{i=1}^{j} m_i > \sum_{i=1}^{j} p_i \log 2. \tag{6}$$

By (4), (6) and the fact that, $\frac{k_{iN}}{N} \to p_i$,

$$\sum_{N \geq 1} 2^{k_{1N}+\cdots+k_{jN}} q_{1N} \ldots q_{jN} \leq \sum_{N \geq 1} e^{-N\sum_{i=1}^{j}\left(m_i - \frac{k_{iN}}{N}\log 2\right)} < \infty.$$

Hence by Corollary 1, a.s. eventually $\mu_N(\Delta) = 0$.

To see the last statement of the Proposition, fix any $\delta > 0$. Let J be the compact set $[-\delta - \log 2, \delta + \log 2]^n$. 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 5. *If $(\bar{\Delta} \cap \Psi)^0 \neq \phi$, then for any $\epsilon > 0$ a.s. eventually*

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

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

We are going to show that the hypothesis of Corollary 1(a) holds. Fix i, $1 \leq i \leq n$. For sufficiently small δ (we choose a specific δ later), we have from (5),

$$2^{-(k_{1N}+\cdots+k_{iN})} q_{1N}^{-1} \ldots q_{iN}^{-1} < \left(\frac{2}{N\delta}\right)^i e^{-N\left[\sum_{j=1}^{i}\left(\frac{k_{jN}}{N}\log 2 - m_j\right) - i\delta\right]}.$$

Since $\tilde{m}$ is an interior point of Ψ, there is an $\alpha > 0$ such that $\sum_{j=1}^{i} p_j \log 2 - \sum_{j=1}^{i} m_j > \alpha$. Now use the fact that $\frac{k_{jN}}{N} \to p_j$ to deduce that eventually $\sum_{j=1}^{i}(\frac{k_{jN}}{N}\log 2 - m_j) > \alpha$. Choose $\delta > 0$ so that eventually $\sum_{j=1}^{i}(\frac{k_{jN}}{N}\log 2 - m_j) - i\delta > \alpha$. With such a choice of δ, the above inequalities imply

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

Hence by Corollary 1, the proposition follows. □

Remark 3.1. $(\bar{\Delta} \cap \Psi)^0 \neq \phi$ *implies in particular, that* $p_1 > 0$. *In fact,* $\Psi^0 \neq \phi$ *iff* $p_1 > 0$.

Now, we have the following,

Proposition 6. *For a.e. sample point* ω,

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

Proof. When $\bar{\Delta} \cap \Psi = \phi$, the result is immediate from Proposition 4.

Assume that $(\bar{\Delta} \cap \Psi)^0 \neq \phi$. Use Proposition 5 with any fixed ϵ, $0 < \epsilon < 1$, take logarithms and use (4) and (5) to see $\lim_{N\to\infty} \frac{1}{N} \log q_{iN} = -m_i$ for each i. □

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

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

Theorem 2. *Almost surely, the sequence* $\{\mu_N\}$ *satisfies LDP with rate function I.*

Proof. Let $\mathcal{A}$ be the collection of all rectangles $\Delta = \Delta_1 \times \cdots \times \Delta_n \subseteq \mathbb{R}^n$ such that each Δ_i is an interval with rational endpoints and either $\bar{\Delta} \cap \Psi = \phi$ or $(\bar{\Delta} \cap \Psi)^0 \neq \phi$.

It is easy to check that $\mathcal{A}$ forms a base for the usual topology of $\mathbb{R}^n$. For $\Delta \in \mathcal{A}$, by Proposition 6, the limit $-\lim_{N\to\infty} \frac{1}{N} \log \mu_N(\Delta)$ exists almost surely. Denote this by L_Δ. Since $\mathcal{A}$ is a countable family, out side a null set, these limits are well defined for all $\Delta \in \mathcal{A}$.

In view of Proposition 1, to complete the proof, we show that for $\tilde{x} \in \mathbb{R}^n$,

$$I(\tilde{x}) = \sup_{\tilde{x}\in\Delta\in\mathcal{A}} L_\Delta. \tag{7}$$

If $\tilde{x} \notin \Psi$, clearly $\sup_{\tilde{x}\in\Delta\in\mathcal{A}} L_\Delta = \infty = I(\tilde{x})$. Now consider, $\tilde{x} = (x_1, \ldots, x_n) \in \Psi$. Suppose $\tilde{x} \in \Delta \in \mathcal{A}$. If $\Delta = \Delta_1 \times \cdots \times \Delta_n$ with $m_i = \inf_{y\in\Delta_i} |y|$, then $m_i \leq |x_i|$. Therefore, by Proposition 6, $L_\Delta = \sum_{i=1}^n m_i \leq \sum_{i=1}^n |x_i|$. Thus

$$\sup_{\tilde{x}\in\Delta\in\mathcal{A}} L_\Delta \leq I(\tilde{x}) \tag{8}$$

On the other hand, consider $\epsilon > 0$ so that $\epsilon < |x_i|$ for any i with $x_i \neq 0$. Let Δ be the box with sides $\Delta_i = (x_i - \epsilon, x_i + \epsilon)$. By choice of ϵ, $m_i = \inf_{y\in\Delta_i} |y|$ 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 Δ so constructed, we have, $L_\Delta = \sum_{\{i:x_i\neq 0\}} |x_i \pm \epsilon|$. This being true for all sufficiently small ϵ, we conclude that

$$\sup_{\tilde{x}\in\Delta\in\mathcal{A}} L_\Delta \geq \sum_{i=1}^n |x_i| = I(\tilde{x}) \tag{9}$$

(8) and (9) complete the proof of (7) thus completing the proof of the theorem. □

We shall now proceed to evaluate the free energy. Denoting $f(\tilde{x}) = \sum_{i=1}^n \beta a_i x_i$,

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

by Proposition 2. This last infimum equals $\inf_{\tilde{x}\in\Psi} \sum_{i=1}^n [1 + \beta a_i \operatorname{sgn}(x_i)]|x_i|$. 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

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

To calculate this last infimum, here is the general idea. Though simple, it helps in a better understanding of a quadratic problem in the next section. Let $c_1, c_2, \ldots, c_n \geq 0$ with $c_1 > 0$ and $(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$.

Let $S \subset \mathbb{R}^n$ be the set of all points $\tilde{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ with nonnegative coordinates and $\sum_1^i x_j \leq \sum_1^i c_j$ for $i = 1, 2, \ldots, n$. Here then is the formula for $l = \inf_{\tilde{x} \in S} \sum_1^n \alpha_i x_i$.

i) If $\alpha_i \geq 0$ for all i then clearly $l = 0$.

ii) If $\alpha_k = \min_i \alpha_i < 0$ and $\alpha_j \geq 0$ for $j > k$ then the infimum, $l = \alpha_k \sum_1^k c_i$ attained at the vector $\tilde{x}^* \in S$ with k-th coordinate $\sum_1^k c_i$ and others zero. Indeed, for any $\tilde{x} \in S$,

$$\begin{aligned}\sum_1^n \alpha_i x_i &\geq \sum_1^k \alpha_i x_i \quad \text{because } \alpha_j \geq 0 \text{ for } j > k\\ &\geq \alpha_k \sum_1^k x_i \quad \text{by choice of } k\\ &\geq \alpha_k \sum_1^k c_i \quad \text{since } \alpha_k < 0\end{aligned}$$

iii) Suppose that $\alpha_{k_1} = \min\{\alpha_i : 1 \leq i \leq n\} < 0, \alpha_{k_2} = \min\{\alpha_i : k_1 < i \leq n\} < 0$, and $\alpha_j \geq 0$ for $j > k_2$. We choose the largest k_1 in case there are two such indices. In this case, the infimum $l = \alpha_{k_1} \sum_1^{k_1} c_i + \alpha_{k_2} \sum_{k_{1+1}}^{k_2} c_i$ attained at the vector $\tilde{x}^* \in S$ with k_1-th coordinate $\sum_1^{k_1} c_i$ and k_2-th coordinates $\sum_{k_{1+1}}^{k_2} c_i$ and other coordinates zero. Indeed, for any $\tilde{x} \in S$

$$\begin{aligned}\sum_1^n \alpha_i x_i &\geq \sum_1^{k_2} \alpha_i x_i \quad \text{because } \alpha_j \geq 0 \text{ for } j > k_2\\ &\geq \alpha_{k_1} \sum_1^{k_1} x_i + \alpha_{k_2} \sum_{k_{1+1}}^{k_2} x_i \quad \text{by choice of } k_1, k_2\\ &\geq \alpha_{k_1} \sum_1^{k_1} c_i + \alpha_{k_2} \sum_{k_{1+1}}^{k_2} c_i \quad \text{since } \alpha_{k_1}, \alpha_{k_2} < 0\end{aligned}$$

the last inequality follows from the fact that $\sum_1^{k_1} x_i \leq \sum_1^{k_1} c_i$ and when equality holds, $\sum_{k_{1+1}}^{k_2} x_i \leq \sum_{k_{1+1}}^{k_2} c_i$.

It is possible to give a general formula either by proceeding as above or by appealing to the simplex method. Since it involves notation, we shall not continue with the generalities. Instead, we shall explain this in our situation, namely, when $S = \Psi^+$, $\alpha_i = (1 - \beta a_i)$ and $c_i = p_i \log 2$. Here then is the needed notations.

For $k = 1, 2, \ldots$ define β_k, r_k as follows:

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

and for $k > 1$,

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

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 $r_0 = 0$ and $\beta_{K+1} = \infty$. For example, if $a_1 > a_2 > \cdots > a_n$ then $\beta_k = \frac{1}{a_k}$, $r_k = k$ for $k = 1, 2, \ldots, n$, 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 $K = 1$.

Here then is the formula for the free energy

Theorem 3. *Almost surely*

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

Two special cases are worth mentioning.

Corollary 5. *i) Let $a_1 > a_2 > \cdots > a_n$. Then a.s.*

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

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

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta) \quad = \log 2 \qquad \text{if } \beta < \frac{1}{a_n}$$
$$= \beta a_n \log 2 \quad \text{if } \beta \ge \frac{1}{a_n}.$$

Remark 3.2. *It is interesting to note that exponential GREM with parameters $(p_1, \ldots, p_n, a_1, \ldots, a_n)$ is equivalent to GREM with parameters $(p'_1, \ldots, p'_K, a'_1, \ldots, a'_K)$ where $p'_1 = \sum_1^{r_1} p_j$, $p'_2 = \sum_{r_1+1}^{r_2} !p_j, \ldots, p'_K = \sum_{r_{(K-1)}+1}^{n} !p_j$ and $a'_1 = a_{r_1}, a'_2 = a_{r_2}, \ldots, a'_K = a_{r_K}$. This is evident from Theorem 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.*

4. GAUSSIAN GREM

In this section, we consider GREM with each ξ centered Gaussian. More specifically, for the N particle system, ξ are i.i.d. centered Gaussian with variance N. Thus it has density

$$\phi(x) = \frac{1}{\sqrt{2N\pi}} e^{-\frac{x^2}{2N}}, \quad -\infty < x < \infty.$$

Note that the density now depends on N.

As earlier, let $N \ge n$ and $k_{1N}, \ldots k_{nN}$ be integers ≥ 1 adding up to N. The random probabilities μ_N are defined on $\mathbb{R}^n$ by transporting the uniform distribution of $2^N = 2^{k_{1N}} \times \cdots \times 2^{k_{nN}}$ to $\mathbb{R}^n$ via the map

$$\sigma \mapsto \left(\frac{\xi_{\sigma_1}(\omega)}{N}, \frac{\xi_{\sigma_1\sigma_2}(\omega)}{N}, \ldots, \frac{\xi_{\sigma_1\ldots\sigma_n}(\omega)}{N} \right).$$

All propositions of the previous section have parallel versions with similar proofs.

Proposition 3′. *$\mu_N \Rightarrow \delta_0$ a.s. as $N \to \infty$.*

As earlier, we assume that $\frac{k_{iN}}{N} \to p_i > 0$ for $1 \le i \le n$. Let

$$\Psi = \left\{ \tilde{x} \in \mathbb{R}^n : \sum_{i=1}^{k} x_i^2 \le \sum_{i=1}^{k} 2p_i \log 2, \quad 1 \le k \le n \right\}.$$

We caution the reader that this set Ψ is different from that of the previous section. With the same notation as earlier, namely, $\Delta = \Delta_1 \times \cdots \times \Delta_n, m_i = \inf_{x \in \Delta_i} |x_i|$, $M_i = \sup_{x \in \Delta_i} |x|$, and $q_{iN} = P(\frac{\xi}{N} \in \Delta_i)$ we have

$$q_{iN} \leq \frac{2}{\sqrt{2\pi}} \int_{\sqrt{N}m_i}^{\sqrt{N}M_i} e^{-\frac{x^2}{2}} dx < \int_{\sqrt{N}m_i}^{\infty} e^{-\frac{x^2}{2}} dx \leq \frac{1}{\sqrt{N}m_i} e^{-\frac{Nm_i^2}{2}}, \tag{10}$$

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

$$q_{iN} \geq \frac{1}{\sqrt{2\pi}} \int_{\sqrt{N}m_i}^{\sqrt{N}M_i} e^{-\frac{x^2}{2}} dx > \frac{1}{2} \int_{\sqrt{N}m_i}^{\sqrt{N}(m_i+\delta)} e^{-\frac{x^2}{2}} dx > \frac{\sqrt{N}\delta}{2} e^{-\frac{N}{2}(m_i+\delta)^2}, \tag{11}$$

for any $0 < \delta < M_i - m_i$. As earlier we have,

Proposition 4′. *If $\bar{\Delta} \cap \Psi = \phi$ then a.s. eventually $\mu_N(\Delta) = 0$. Moreover, the sequence $\{\mu_N\}$ is supported on a compact set.*

Proposition 5′. *If $(\bar{\Delta} \cap \Psi)^0 \neq \phi$ then for any $\epsilon > 0$, a.s. eventually*

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

Proposition 6′. *Almost surely*

$$\begin{aligned} \lim_{N \to \infty} \frac{1}{N} \log \mu_N(\Delta) &= -\tfrac{1}{2} \sum_{i=1}^{n} m_i^2 \qquad \text{if } (\bar{\Delta} \cap \Psi)^0 \neq \phi \\ &= -\infty \qquad \text{if } \bar{\Delta} \cap \Psi = \phi. \end{aligned}$$

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

$$\begin{aligned} I(\tilde{x}) &= \frac{1}{2} \sum_{i=1}^{n} x_i^2 \quad \text{if } \tilde{x} \in \Psi \\ &= \infty \quad \text{otherwise.} \end{aligned}$$

Theorem 4. *The sequence $\{\mu_N\}$ satisfies LDP with rate function I.*

Thus proceeding as in the earlier section, the free energy is given by

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

a formula first derived in ref. 3 by different methods and later in ref. 8 by large deviation technique.

We now proceed to explicitly evaluate the infimum that occurs above. The formula was already given in ref. 8. Our purpose is to bring out the similarities between the exponential case and the Gaussian case.

Here is the general idea. Let $c_1, c_2 \dots, c_n \geq 0$ with $c_1 > 0$. Let $\tilde{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ with each $\alpha_i > 0$. Let $S \subset \mathbb{R}^n$ be the set of all points $\tilde{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ with nonnegative coordinates and $\sum_1^i x_j^2 \leq \sum_1^i c_j$ for $i = 1, 2, \dots, n$. Here then is the formula for $l = \inf_{\tilde{x} \in S} \sum_1^n (x_i - \alpha_i)^2$.

i) If $\frac{c_1 + \cdots + c_i}{\alpha_1^2 + \cdots + \alpha_i^2} \geq 1$ for all i then clearly $\tilde{\alpha} \in S$ and $l = 0$.

ii) Let $\gamma = \min \frac{0_1 + \cdots + c_i}{\alpha_1^2 + \cdots + \alpha_i^2}$. Let k be the largest index such that $\gamma = \frac{c_1 + \cdots + c_k}{\alpha_1^2 + \cdots + \alpha_k^2}$. Assume that $\frac{c_{k+1} + \ldots + c_i}{\alpha_{k+1}^2 + \ldots + \alpha_i^2} \geq 1$, for $i > k$. Put $\tilde{\alpha}^* = (\alpha_1^*, \dots, \alpha_n^*)$ where

$$\begin{aligned} \alpha_i^* &= \sqrt{\gamma}\alpha_i \quad \text{for } i \leq k \\ &= \alpha_i \qquad \text{for } i > k. \end{aligned}$$

Clearly $\tilde{\alpha}^* \in S$. Moreover the infimum, $l = \sum_1^k (\alpha_i^* - \alpha_i)^2 = (1 - \sqrt{\gamma})^2 \sum_1^k \alpha_i^2$. To see this, consider any $\tilde{x} \in S$. By Cauchy-Schwarz, $\sum_1^k \alpha_i^* x_i \leq \sum_1^k \alpha_i^{*2}$ and hence $\sum_1^k \alpha_i^* (\alpha_i^* - x_i) \geq 0$. Since $\gamma < 1$, $\sum_1^k \alpha_i^* (\alpha_i^* - x_i) \leq \sum_1^k \alpha_i (\alpha_i^* - x_i)$. A simple algebra shows

$$\sum_1^k (x_i - \alpha_i)^2 - \sum_1^k (\alpha_i^* - \alpha_i)^2 \geq \sum_1^k (x_i - \alpha_i^*)^2 \geq 0. \tag{12}$$

iii) Let γ and k be as above. Suppose $\frac{c_{k+1} + \ldots + c_i}{\alpha_{k+1}^2 + \ldots + \alpha_i^2} < 1$, for some $i > k$. Put $\eta = \min_{i > k} \frac{c_{k+1} + \ldots + c_i}{\alpha_{k+1}^2 + \ldots + \alpha_i^2}$, so that $\eta < 1$. Let m be the largest index when this ratio equals η. Clearly $m > k$. Assume that $\frac{c_{m+1} + \ldots + c_i}{\alpha_{m+1}^2 \ldots + \alpha_i^2} \geq 1$, for $i > m$. Put

$$\begin{aligned} \alpha_i^* &= \sqrt{\gamma}\alpha_i \quad \text{for } i \leq k \\ &= \sqrt{\eta}\alpha_i \quad \text{for } k+1 \leq i \leq m \\ &= \alpha_i \qquad \text{for } i > m. \end{aligned}$$

Clearly $\tilde{\alpha}^* \in S$. Further, the infimum, $l = \sum_1^m (\alpha_i^* - \alpha_i)^2 = (1 - \sqrt{\gamma})^2 \sum_1^k \alpha_i^2 + (1 - \sqrt{\eta})^2 \sum_{k+1}^m \alpha_i^2$. To see this, consider any point $\tilde{x} \in S$. It is enough to show (12) with k replaced by m. As earlier $\sum_1^k \alpha_i^* (\alpha_i^* - x_i) \geq 0$ and $\sum_1^m \alpha_i^* (\alpha_i^* - x_i) \geq 0$. Using $\gamma < \eta < 1$, we have $\sum_1^m \alpha_i^* (\alpha_i^* - x_i) \leq \frac{1}{\sqrt{\eta}} \sum_1^m \alpha_i^* (\alpha_i^* - x_i) \leq \frac{1}{\sqrt{\eta}} \sum_1^m \alpha_i^* (\alpha_i^* - x_i) + (\frac{1}{\sqrt{\gamma}} - \frac{1}{\sqrt{\eta}}) \sum_1^k \alpha_i^* (\alpha_i^* - x_i)$. In other words, $\sum_1^m \alpha_i (\alpha_i^* - x_i) \geq \sum_1^m \alpha_i^* (\alpha_i^* - x_i)$. A simple algebra completes proof of (12) with k

replaced by m. Incidentally the above inequality says that the angle between the vectors $\tilde{\alpha} - \tilde{\alpha}^*$ and $\tilde{\alpha}^* - \tilde{x}$ is at most $\frac{\pi}{2}$.

We shall not continue with the generalities, instead we explain this in our situation, namely, $S = \Psi^+$, $\alpha_i = \beta a_i$ and $c_i = p_i \log 2$.

Following,[8] let us put $B_{jk} = \sqrt{\frac{(p_j+\cdots+p_k)2\log 2}{a_j^2+\cdots+a_k^2}}$ for $1 \le j \le k \le n$. Set

$$\beta_1 = \min_k B_{1,k} \quad r_1 = \max\{i : B_{1,i} = \beta_1\}$$

$$\beta_2 = \min_{k>r_1} B_{r_1+1,k} \quad r_2 = \max\{i > r_1 : B_{r_1+1,i} = \beta_2\}$$

and in general

$$\beta_{m+1} = \min_{k>r_m} B_{r_m+1,k} \quad r_{m+1} = \max\{i > r_m : B_{r_m+1,i} = \beta_{m+1}\}.$$

Clearly, for some K with $1 \le K \le n$, we have $r_K = n$. Put $r_0 = \beta_0 = 0$ and $\beta_{K+1} = \infty$. Note that $\beta_0 < \beta_1 < \beta_2 \ldots < \beta_K < \beta_{K+1} = \infty$.

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

$$\begin{aligned} x_i^* &= \beta_l a_i \quad \textit{if } i \in \{r_{l-1}+1, \ldots, r_l\} \text{ for some } l,\ 1 \le l \le j \\ &= \beta a_i \quad \textit{if } i > r_j + 1. \end{aligned}$$

Then $\inf_{\tilde{x}\in\Psi} \sum_{i=1}^n (x_i - \beta a_i)^2$ occurs at $\tilde{x}^*$. This immediately leads to the following

Theorem 5. *Almost surely.*

$$\begin{aligned} \lim_N \frac{1}{N} \log Z_N(\beta) &= \log 2 + \frac{\beta^2}{2} \sum_{i=1}^n a_i^2 \quad \textit{if} \quad \beta < \beta_1 \\ &= \log 2 + \frac{\beta^2}{2} \sum_{l=1}^n a_i^2 - \frac{1}{2} \sum_{l=1}^j (\beta_l - \beta)^2 \sum_{r_{l-1}+1}^{r_l} a_i^2 \\ &\qquad \textit{if} \quad \beta_j \le \beta < \beta_{j+1}. \end{aligned}$$

With proper identification of parameters this is essentially the same formula as in ref. 3,8. In defining the β_i, Capocaccia *et al.* use a variant in ref. 3 Sec. 3.2. In defining r_i, Dorlas and Dukes[8] consider the least index. This makes no difference because *'nothing happens'* until the maximum index is captured. Their weights a_i are incorporated in the density, there was no need to assume $\sum a_i = 1$, their parameter J can be incorporated in the weights.

Here also two simple cases are worth mentioning. The number β_j mentioned below are same as the above, in these particular cases.

Corollary 6.

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* $j = 1, \ldots, n$. *Then a.s.*

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \log 2 + \frac{\beta^2}{2} \sum_1^n a_i^2 \quad \textit{if} \quad \beta < \beta_1,$$

$$= \sum_{j+1}^n p_i \log 2 + \sum_1^j \beta a_i \sqrt{2p_i \log 2} + \frac{\beta^2}{2} \sum_{j+1}^n a_i^2$$

$$\textit{if} \quad \beta_j \le \beta < \beta_{j+1} \textit{ for } 1 \le j < n,$$

$$= \beta \sum_1^n a_i \sqrt{2p_i \log 2} \quad \textit{if} \, \beta \ge \beta_n.$$

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

$$\lim_N \frac{1}{N} \log Z_N(\beta) = \log 2 + \frac{\beta^2}{2} \sum_1^n a_i^2 \quad \textit{if} \quad \beta < \sqrt{\frac{2\log 2}{\sum a_i^2}}$$

$$= \beta \sqrt{2 \log 2 \sum a_i^2} \quad \textit{if} \quad \beta \ge \sqrt{\frac{2\log 2}{\sum a_i^2}}$$

A phenomenon similar to the exponential GREM can be observed in the present case also. An n-level Gaussian GREM with parameters $(p_1, \ldots p_n; a_1, \ldots, a_n)$ is equivalent to a K level (K is as defined earlier) Gaussian GREM with parameters $(p'_1, \ldots, p'_K; a'_1, \ldots a'_K)$ where $p'_{l+1} = \sum_{r_l+1}^{r_{l+1}} p_i$ and $a'_{l+1} = \sqrt{\sum_{r_{l+1}}^{r_{l+1}} a_i^2}$. Thus in order that an n-level Gaussian GREM does not collapse to lower level it is necessary and sufficient that $\frac{p_i}{a_i^2}$ be strictly increasing. Thus unlike in the exponential case, the condition now depends on both the parameters (p_i) and (a_i).

5. REMARKS

So far we assumed that the vector $\tilde{a} = (a_1, \ldots, a_n)$ and $\tilde{p} = (p_1, \ldots, p_n)$ have strictly positive entries. It is natural to ask: what happens if some of these quantities are zero? The following remark answers this question.

Remark 5.1. **(A)** Let us assume $p_i > 0$ for all i. Suppose $a_n > 0$ and some of the a_i are possibly zero. Let $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$ be those which are not zero with $1 \leq i_1 < i_2 < \cdots < i_m = n$. Then GREM with parameters $(\tilde{p}; \tilde{a})$ is same as GREM with parameters $(p'_1, \ldots, p'_m; a'_1, \ldots, a'_m)$ where $p'_{l+1} = \sum_{j=i_l+1}^{i_{l+1}} p_j$ (take $i_0 = 0$) and $a'_l = a_{i_l}$. However if $a_n = 0$ the calculation differs. Suppose for some $k < n$, $a_k \neq 0$ and $a_{k+1} = \ldots = a_n = 0$. Then either by using the partition function or by recalculating the necessary infimum, one can see that the limiting energy equals $p^* \log 2 + (1 - p^*)e$ where $p^* = \sum_{k+1}^{n} p_i$ and e is the energy of the GREM with parameters as follows. In case of Gaussian GREM the parameters are $(\frac{p_1}{1-p^*}, \ldots, \frac{p_k}{1-p^*}; \frac{a_1}{\sqrt{1-p^*}}, \ldots, \frac{a_k}{\sqrt{1-p^*}})$. And in case of exponential GREM the parameters are $(\frac{p_1}{1-p^*}, \ldots, \frac{p_k}{1-p^*}; a_1, \ldots, a_k)$. Observe that there are no multipliers for a_i in the last case.

The above results are only expected because, if an intermediate level of the tree gets zero weights, it always amounts to passing to the next level whereas if the last level gets zero weights it amounts to multiplying the partition function of the first $(n - 1)$ level tree with the number of furcations at the last level.

(B) Let us assume $a_i > 0$ for all i, but some p_i are allowed to be zero. To start with, suppose $p_1 > 0$. Notice that as long as $p_1 > 0$ the set Ψ has non-empty interior and the argument regarding rate function goes through. If $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ are the indices of nonzero p values, then the $(\tilde{p}; \tilde{a})$ GREM is equivalent to a k level GREM with parameters as follows. In the Gaussian case the parameters are $(p'_1, \ldots p'_k; a'_1 \ldots, a'_k)$ where $p'_j = p'_{i_j}$ and $a'_j = \sqrt{\sum_{l=i_j}^{i_{j+1}-1} a_l^2}$. In the exponential case the parameters are $(p'_1, \ldots p'_k; a'_1 \ldots, a'_k)$ where $p'_j = p'_{i_j}$ and $a'_j = \max\{a_i : i_j \leq i \leq i_{j+1} - 1\}$.

Let us now consider the case $p_1 = 0$. Let us assume, for simplicity, that $p_i = 0$ for $i \leq j$ and $p_i > 0$ for $i > j$. In this case, the $(\tilde{p}; \tilde{a})$ GREM is equivalent to the $(n - j)$ level $(p_{j+1}, \ldots, p_n; a_{j+1}, \ldots a_n)$ GREM. This can be seen either by modifying the LDP argument or by using properties of maxima of random variables involved. We explain these for the exponential case, same holds for the Gaussian case as well.

To see the changes needed for the LDP argument, let us return to Sec. 3. Proposition 4 holds as stated. Regarding Proposition 5, note that, in this case, first j coordinates of any point in Ψ are zero. In the hypothesis of Proposition 5 we need to take cubes Δ such that $0 \in \Delta_i$ for $i \leq j$ and $[\overline{\Delta(j)} \cap \Psi(j)]^0 \neq \phi$ where $\Delta(j) = \Delta_{j+1} \times \cdots \times \Delta_n$ and $\Psi(j)$ is the projection of Ψ to the last $(n - j)$ space. Naturally, interior is in $\mathbb{R}^{n-j}$. The conclusion will now read that for any $\epsilon > 0$, a.s. eventually $(1 - \epsilon) \prod_{i \geq j+1} q_{iN} \leq \mu_N(\Delta) \leq (1 + \epsilon) \prod_{i \geq j+1} q_{iN}$. For the proof of Proposition 5 stated like this, first observe that, by Proposition 4, $\mu_N(\Delta)$ is eventually same as its marginal on $\Delta(j)$. This being so, one can consider the tree from (j+1)-th level onwards (thus at the first level this tree has $2^{\sum_{i \leq j+1} k_{iN}}$ nodes) and apply Corollary 1. Proposition 6 is to be modified in a similar manner. Theorem 2 stands as stated.

Alternatively, denote $\xi_i^* = \max_{\sigma_1\ldots\sigma_i} |\xi_{\sigma_1\ldots\sigma_i}|$, for $1 \le i \le j, \sigma'_{j+1} = \sigma_1\sigma_2\cdots\sigma_{j+1}$, $Z_N^*(\beta) = \sum_{\sigma'_{j+1},\sigma_{j+2},\ldots\sigma_n} e^{-\beta\sum_{i\ge j+1} a_i\xi_{\sigma'_{j+1}\sigma_{j+2}\ldots\sigma_n}}$ and observe

$$-\frac{\beta}{N}\sum_{i\le j} a_i\xi_i^* + \frac{1}{N}\log Z_N^*(\beta) \le \frac{1}{N}\log Z_N(\beta) \le \frac{\beta}{N}\sum_{i\le j} a_i\xi_i^* + \frac{1}{N}\log Z_N^*(\beta).$$

It now suffice to show that for each $i \le j$, $\frac{\xi_i^*}{N} \to 0$ a.s. Fix $i \le j$, put $l_N = \sum_{m\le i} k_{mN}$. For any $\epsilon > 1$, observe

$$\mathbf{P}\left(\frac{\xi_i^*}{N} > \epsilon(\epsilon-1)\right) \le 2^{l_N}\mathbf{P}\left(\frac{|\xi_{\sigma_1}|}{N} > \epsilon(\epsilon-1)\right) \le 2^{l_N}e^{-(\epsilon-1)N}\mathbf{E}e^{|\xi_{\sigma_i}|}$$

which is summable over N. Thus for any $\epsilon > 1$ a.s. eventually $\frac{\xi_i^*}{N} \le \epsilon(\epsilon-1)$. For the Gaussian case one has to consider $\epsilon(\epsilon-2)$ instead of $\epsilon(\epsilon-1)$ with $\epsilon > 2$.

(C) We refrain from a full discussion of all the cases that may arise when one allows some of the quantities p_i or a_i to be zero. We mention only two examples.

Example 1. $n = 2$, $p_1 = 1$, $p_2 = 0$ while $a_1 = 0$ and $a_2 = 1$. Then this system is nothing but REM. In fact, $a_1 = 1$ and $a_2 = 0$ also corresponds to REM.

Example 2. $n = 2$, $p_1 = 0$, $p_2 = 1$ while $a_1 = 1$ and $a_2 = 0$. Then limiting energy is log 2, no matter what β is. When $a_2 = 1$ this is again REM.

As mentioned in the introduction, the trees considered in the literature are in a sense obtained by factoring 2, the number of spins. The next remark compares our formulation with this.

Remark 5.2. It is the practice in literature to consider GREM as follows. Fix constants $\alpha_1, \ldots \alpha_n$, each greater than 1 with $\prod \alpha_i = 2$. For the N particle system one considers the regular tree with $[\alpha_i^N]$ furcations at the i-th level. Here $[x]$ is the largest integer $\le x$. Since we are using $[x]$, the tree may not have 2^N branches to exactly correspond to the configuration space. However, this does not pose any problem in view of Remark 2.2. To be precise, if $l_{iN} \le [\alpha_i^N] < l_{iN} + 1$ then the tree considered in the literature, for the N particle system, has l_{iN} furcations of each node of the $(i-1)$-th level. Let k_{iN} be such that $2^{k_{iN}} \le l_{iN} < 2^{k_{iN+1}}$. Consider, in the notation of the present paper, GREM with parameters $\tilde{k} = \{(k_{iN} : 1 \le i \le n), N \ge n\}$ Observe that $\frac{k_{iN}}{N} \to \frac{\log\alpha_i}{\log 2}$ and $1 \le \frac{l_{iN}}{2^{k_{iN}}} \le 2$ for $1 \le i \le n$. A simple calculation shows that $N - 2n \le \sum_i k_{iN} \le N$. As a consequence the fact that $\sum_i k_{iN}$ is not exactly equal to N does not matter, in the sense, any short fall can be absorbed at any one of the levels.

As is clear from the above, the relation of the $\tilde{\alpha}$- GREM with $\tilde{k}$- GREM is that $p_i = \frac{\log \alpha_i}{\log 2}$. Apart from bringing in the tree structure, we have found no intuitive explanation for considering the numbers α_i. (In a sense, when we consider the $\tilde{\alpha}$ setup, the limits for the proportions $\frac{k_{iN}}{N}$ are already taken.) However in the present formulation the N particles are divided into n groups with k_{iN} particles in the i-th group. Moreover the present formulation leads to new problems as seen in Remark 5.1.

The following remark concerns the randomization of the tree.

Remark 5.3. Going back to Corollary 2, it must be clear by now, that even if we take a regular Poisson tree as in there, for almost every tree the free energy exists and corresponds to the usual one. Regarding the multinomial tree GREM, there is a discrepancy in the two conditions of Corollary 4. However instead of taking general multinomial trees at each level, we proceed as follows. Consider a die with n faces with chance of heads for face i being p_i. Consider a sequence of independent throws of the die and let $K_N = (K_{1N}, \ldots, K_{nN})$ be the outcome of the first N throws. By the strong law of large numbers, we have, $\frac{k_{iN}}{N} \to p_i$ almost surely. As a consequence, in this model also, for almost every tree the free energy exists and corresponds to the usual one.

The next remark concerns the treatment of Contucci *et al.*[(4)] We thank the authors for discussions regarding their setup.

Remark 5.4. Contucci *et al.*[(4)] generalized the powerful convexity argument of Guerra and Toninelli.[(12)] To briefly recall, the convexity hypothesis is the following: For $N = N_1 + N_2$, and for $\sigma, \tau \in 2^N$ with projections $\pi_1(\sigma), \pi_1(\tau)$ on 2^{N_1} and $\pi_2(\sigma), \pi_2(\tau)$ on 2^{N_2} the following inequality

$$C_N(\sigma, \tau) \leq \frac{N_1}{N} C_{N_1}(\pi_1(\sigma), \pi_1(\tau)) + \frac{N_2}{N} C_{N_2}(\pi_2(\sigma), \pi_2(\tau)),$$

should hold, where $C_N(\sigma, \tau) = \mathrm{E}\,(H_N(\sigma) H_N(\tau))$.

Here now is the setup of Contucci *et al.*. Suppose we have two trees S and T with n layers each. Say S has m_1 nodes $\alpha_1, \ldots, \alpha_{m_1}$, at the first level. Node α_i furcates to m_2 nodes α_{ij}; $1 \leq j \leq m_2$ etc. This goes on till the n-th level. Similarly the tree T has M_1 nodes $\beta_1, \ldots, \beta_{M_1}$, at the first level. Node β_i furcates to M_2 nodes β_{ij}; $1 \leq j \leq M_2$ etc. till the n-th level. Then the product tree $S \oplus T$ is the tree which has $m_1 M_1$ nodes $\left\{(\alpha_i, \beta_i') : 1 \leq i \leq m_1; 1 \leq i' \leq M_1\right\}$ at the first level. Node (α_i, β_i') furcates into $m_2 M_2$ nodes $\left\{(\alpha_{ij}, \beta_{i'j'}) : 1 \leq j \leq m_2; 1 \leq j' \leq M_2\right\}$ and so on. In such a case we say that the tres S and T are complementary in the tree $S \oplus T$. We denote $T \oplus T$ by T^2.

Now fix an n-tuple of integers $(b_1, b_2, \ldots, b_n)$. Let $b = \sum b_i$ Consider a tree T with 2^{b_1} nodes at the first level, with each node furcating to 2^{b_2} nodes at the second level etc. Thus the tree has 2^b many leaves. Now for each integer $l \geq 1$, consider the tree T^l, the l-fold product of the tree T. The tree T^l corresponds to bl-particle system. They have proved the existence of the free energy limit along such a sequence of trees. This is covered by the present treatment as well.

However, if one considers an arbitrary sequence of trees, as we did, the convexity hypothesis may not hold. Here is a simple example. Consider $n = 2, a_1 > 0, a_2 > 0$. $N = 5$ with $k_{1N} = 2$ and $k_{2N} = 3$; $N_1 = 3$ with $k_1 N_1 = 2$ and $k_{2N_1} = 1$; $N_2 = 2$ with $k_{1N_2} = k_{2N_2} = 1$. Take $\sigma = (+1, +1, +1, +1, +1)$, $\tau = (+1, +1, -1, -1, -1)$ so that $\pi_1(\sigma) = (+1, +1, +1)$, $\pi_1(\tau) = (+1, +1, -1)$; $\pi_2(\sigma) = (+1, +1)$ and $\pi_2(\tau) = (-1, -1)$. It is easy to see that $C_N(\sigma, \tau) = C_{N_1}(\pi_1(\sigma), \pi_1(\tau)) = a_1$ while $C_{N_2}(\pi_2(\sigma), \pi_2(\tau)) = 0$. Thus the condition of convexity amounts to $a_1 \leq \frac{N_1}{N} a_1$, which is clearly not true.

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 GREM.

Remark 5.5. As observed in Remark 3.2, an n level GREM may reduce to a k level GREM for some $k < n$. In such a case, some weights a_i either do not appear in the formula for free energy as in the exponential case (Theorem 3), or some weights occur in groups and get added up as in the Gaussian case (Theorem 5). 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.

(A) 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 Corollary 5, 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 $0 < x_1 < \ldots < x_n$ 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

$$0 = c_0 < c_1 < \cdots < c_n \quad \text{and} \quad \sum_{i=1}^{n} x_i(c_i - c_{i-1}) = \log 2. \tag{13}$$

(B) Since a Gaussian GREM is in reduced from if and only if all the p_i, a_i are non zero and $\frac{p_1}{a_1^2} < \frac{p_2}{a_2^2} < \cdots < \frac{p_n}{a_n^2}$, let us assume this to be the case. From Corollary 6, it follows that $\mathcal{E}(\beta)$ is piecewise quadratic continuous function with $\mathcal{E}(0) = \log 2$. It has a continuous derivative $\mathcal{E}'(\beta)$ with $\mathcal{E}'(0) = 0$. Further, $\mathcal{E}''(\beta) = \sum_1^n a_i^2$ in $(0, \frac{\sqrt{2p_1 \log 2}}{a_1})$; $= \sum_{k+1}^n a_i^2$ in $(\frac{\sqrt{2p_k \log 2}}{a_k}, \frac{\sqrt{2p_{k+1} \log 2}}{a_{k+1}})$; and $= 0$ for $\beta > \frac{\sqrt{2p_n \log 2}}{a_n}$.

Here also the energy function $\mathcal{E}(\beta)$ identifies the parameters. Let $0 < x_1 < \ldots < x_n$ be the points where the left and right derivatives of $\mathcal{E}'(\beta)$ are unequal and $\mathcal{E}''(\beta) = c_i^2$ in (x_i, x_{i+1}) with $x_0 = 0$ and $x_{n+1} = \infty$. Then $a_i = \sqrt{c_i^2 - c_{i+1}^2}$ and $p_i = \frac{x_i^2(c_i^2 - c_{i+1}^2)}{2 \log 2}$, for $1 \leq i \leq n$, note that $c_{n+1} = 0$.

The energy functions can be characterized in this case also. Let f be a C^1 function on $[0, \infty)$ with $f(0) = \log 2$ and $f'(0) = 0$. Further suppose that there are finitely many points $0 < \cdots < x_n$ where the left and right derivatives of f' are unequal and f'' is a positive constant, say, c_i^2 in (x_{i-1}, x_i). Here $x_0 = 0$ and $x_{n+1} = \infty$. Then f is the energy function for some Gaussian GREM iff

$$c_1^2 > \cdots > c_n^2 > c_{n+1}^2 = 0 \quad \text{and} \quad \frac{1}{2} \sum_{i=1}^{n} x_i^2 \left(c_i^2 - c_{i+1}^2\right) = \log 2. \tag{14}$$

It is interesting to compare (13) and (14).

Remark 5.6. In ref. 13 one of the authors considered REM with driving density for the N-particle system being

$$\phi(x) = \frac{1}{2\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{\alpha}{N}\right)^{\frac{\alpha-1}{\alpha}} e^{-\frac{|x|^\alpha}{\alpha N^{\alpha-1}}} \qquad -\infty < x < \infty$$

Here $\alpha \geq 1$. One can consider GREM in this environment and obtain easily a rate function. A referee pointed out to us that distributions with exponentially decaying tails, as in, ref. 11 could also be treated by the present method. Since the solution of the variational problem and explicit formula for the free energy eludes us, we refrain from giving the details.

6. CONCLUSIONS

A reformulation of GREM is considered which allows general tree structure. This makes randomization of the tree possible. A general result on trees is proved in Sec. 2 and applied to these models. It is demonstrated that large deviation technique is best suited for these problems. In several respects, as seen in Secs. 2 and 4, Gaussian GREM is no more complicated than exponential GREM, contrary to popular belief. There are some dissimilarities as seen in Sec. 5. It is observed that the energy function identifies the parameters of the models.

REFERENCES

1. A. Bovier and I. Kurkova, Derrida's generalized random energy models. 1. Models with finitely many hierarchies. *Ann. Inst. H. Poincare. Prob. et Statisques (B) Prob. Stat.* **40**:439–480 (2004).
2. A. Bovier and I. Kurkova, Derrida's generalized random energy models. 2. Models with continuous hierarchies. *Ann. Inst. H. Poincare. Prob. et Statisques (B) Prob. Stat.* **40**:480–485 (2004).
3. D. Capocaccia, M. Cassandro and P. Pico, On the existance of thermodynamics for the generalized random energy model. *J. Stat. Phys.* **46**:493–505 (1987).
4. P. Contucci, M. D. Esposti, C. Giardinà and S. Graffi, Thermodynamical limit for correlated Gaussian random energy models. *Commun. Math. Phys.* **236**:55–63 (2003).
5. P. Carmona, and Y. Hu, Universality in Sherrington-Kirkpatrick's spin glass model. *arXiv:math.PR/0403359 v2.* (2004).
6. A. Dembo and O. Zeitouni, *LARGE DEVIATION: Techniques and Applications*, 2nd Ed. (Springer-Verlag, New York, 1998).
7. B. Derrida, A generalization of the random energy model which includes correlations between energies. *J. Phys. Lett.* **46**:401–407 (1985).
8. T. C. Dorlas and W. M. B. Dukes, Large deviation approach to the generalized random energy model. *J. Phys. A: Math. Gen.* **35**:4385–4394 (2002).
9. T. C. Dorlas and J. R. Wedagedera, Large deviations and the random energy model. *Int. J. Mod. Phy. B* **15**(1):1–15 (2001).
10. R. S. Ellis, *Entropy, Large Deviations and Statistical Mechanics* (Spring-Verlag, New York, 1985).
11. T. Eisele, On a third-order phase transition. *Commun. Math. Phys.* **90**:125–159 (1983).
12. F. Guerra and F. L. Toninelli, The thermodynamic limit in mean field spin glass models. *Commun. Math. Phys.* **230**:71–79 (2002).
13. N. K. Jana, Exponential random energy model. *arXiv: math.PR/0602670.* (2005).
14. M. Talagrand, *SPIN GLASSES: A Challenge for Mathematicians* (Spring-Verlag, New York, 2003).