Random Energy Models

Indian Statistical Institute, Bangalore

November 23-27, 2007 B.V.Rao

The aim of these lectures is to introduce random energy models, to understand some of the problems, their solutions and techniques. Neither are the results presented most general, nor are the paths geodesic.

1. How did we end up here:

The origins of the subject go back to our attempts to understand magnetism. While it was felt that the collective behaviour of atoms — more precisely the alignment of the spins — is the cause, the reasons for this collective behaviour were not clear. The discovery by Curie in 1895 of the Curie point — a certain temperature at which the magnetic property disappears — pressed for models to explain the phenomenon. As Bovier [17] put it, Wilhelm Lenz (1888-1957) had a beautiful and simple idea. Imagine atoms are at the sites of Z^d, the integer lattice. Assume the simplest spin variables ± 1. Assume that only neighbours interact and the interaction favours neighbouring atoms to take the same value. There is an external magnetic field which favours globally $+$ or globally $-$ spin. Bring in ideas from statistical mechanics and solve the problem. This was the problem he gave to his student Ernest Ising and ever since this is known as Ising model, sometimes called Lenz-Ising model (The Lenz Law in induced magnetism is due to Heinrich Lenz (1804-1865)). Mathematical formulation of the model is as follows. Since spins live on the integer lattice Z^d and take only ± 1 values, the configuration space is $\{-1,+1\}^{Z^d}$. If $\sigma = (\sigma_i; i \in Z^d)$ is a configuration, then the energy of the system in this configuration, called Hamiltonian, is $H(\sigma) = -\sum\limits_{|i-j|=1} J\sigma_i\sigma_j - h\sum\limits_i \sigma_i$. Here the numbers J and h are respectively the interaction between spins and strength of the external field. Of course $|i-j| = 1$ means that the two points i and j of Z^d differ by exactly one coordinate, that is, they are neighbours. Eventhough the sum is over neighbouring pairs, it is an infinite sum and does not converge. The best way to interpret is to consider N particle system and take limit as $N \to \infty$. To do this we first need some notation. For a set $\Lambda \subset Z^d$, put $\Omega_\Lambda = \{-1,+1\}^\Lambda$. For any finite set Λ and $\sigma \in \Omega_\Lambda$ and $\eta \in \Omega_{\Lambda^c}$ put $H_\Lambda(\sigma|\eta) = -\sum_\Lambda J\sigma_i\sigma_j - h\sum_{\imath\in\Lambda}\sigma_i$ where the first sum $\sum_\Lambda$ is over all pairs i and j with $|i-j| = 1$ and at least one of them is in Λ. In such a case, either both i and j are in Λ or one of them is in Λ and the other one in Λ^c; a boundary point of Λ. When this happens and $j \in \Lambda^c$, the corresponding spin σ_i is to be taken as η_i. This is the energy of the configuration σ in the finite volume Λ when the outside configuration is η. Of course, since only neighbours interact, this depends only on sites in Λ and their neighbours.

The key principle of statistical mechanics is the following (Feynman's [30]). If a system in equilibrium can be in one of N states then the probability of the system having energy E_n is $\frac{1}{Q}e^{-E_n/\kappa T}$ where $Q = \sum\limits_n e^{-E_n/\kappa T}$. Here κ is the Boltzman constant, T is temperature and Q is called the partition function. The expected value of an observable quantity f is $\frac{1}{Q}\sum\limits_i f(i)e^{-E_i/\kappa T}$. This fundamental law (formulated in the quantum set up) is the summit of statistical mechanics and the entire subject is either slide down from the summit as the principle is applied to various cases, or the climb up where the fundamental law is derived and the concepts of thermal equilibrium and temperature T clarified.

Returning to the Lenz-Ising model, given a finite set $\Lambda \subset Z^d$ we have the Hamiltonian $H_\Lambda(\sigma|\eta)$ defined above for $\sigma \in \Omega_\Lambda$ and $\eta \in \Omega_{\Lambda^c}$. The partition function is $Z_\Lambda^\eta = \sum\limits_{\sigma\in\Omega_\Lambda} e^{-\beta H_\Lambda(\sigma|\eta)}$. Here β is the inverse temperature and κ is taken as one. Let $\mathcal{F}_\Lambda$ denote the σ-field on Ω generated by the coordinate maps in Λ. When $\Lambda = Z^d$, we denote this σ-field by $\mathcal{F}$. For a finite set Λ we can think of the Gibbs distribution $\frac{1}{Z_\Lambda^\eta}\exp\{-\beta H_\Lambda(\sigma|\eta)\}$ as the conditional probability on $\mathcal{F}_\Lambda$ given $\mathcal{F}_{\Lambda^c}$. To describe the equilibrium state of the system, one looks for a probability μ on $(\Omega, \mathcal{F})$ such that under μ the conditional probability on $\mathcal{F}_\Lambda$ given $\mathcal{F}_{\Lambda^c}$ are as above for every finite subset $\Lambda \subset Z^d$. Whether there is any such probability at all, if so how many exist, if more than one exist then the nature of the set of all such probabilities and its extreme points etc are all classical. Moreover one considers general interaction functions (See Ruelle [48], Sinai [53], Preston [47], Georgii [32]).

The present story starts with efforts to understand magnetic properties of alloys. In the sixties there was much interest in the behaviour of isolated magnetic 'impurities' in non-magnetic hosts (like Mn in Cu or Fe in Au). The magnetic moments are not arranged in a lattice. The interaction between the magnetic moments is not always of the same sign. This calls for a change from the classical set up. One does see spins exhibiting a collective behaviour, a random freezing which exhibits several peculiarities. To explain the observed phenomena a model was proposed by Edwards and Anderson in 1975.

As Sherrington [52] says 'Edwards and Anderson produced a paper in 1975 that at one fell swoop recognized the importance of the combination of frustration and quenched disorder as the 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 (magnetic) susceptibility cusp. Edwards and Anderson's new approach was beautifully minimal, fascinating and attractive, their analysis was highly novel and sophisticated, involving radically new concepts and methods but also unusual and unproven ansatz, as well as several different approaches'. However, this model (see the Hamiltonian later in section 2) was not exactly solvable. Sherrington and Kirckpatrik proposed a mean-field model. The adjective 'mean-field' refers to the fact that each spin interacts with all the others. In 1981 and later in 1985, Bernard Derrida proposed two models called Random Energy Model and Generalized Random Energy model.

In its bare essentials then, for an N-particle system we have a configuration space Σ_N and for each σ in this space we have a random variable $H_N(\sigma)$, the Hamiltonian in this configuration. The partition function is $Z_N = \sum\limits_{\sigma} e^{-\beta H_N(\sigma)}$ where β is the inverse temperature. There are three main issues which lead to several problems. Firstly, does $\frac{1}{N} \log Z_N$ have a limit and if so is it non-random. This is the limiting energy, or just energy. It is important to have this limit to be non-random. Secondly, we have random Gibbs measures $G_N(\sigma) = e^{-\beta H_N(\sigma)}/Z_N$ on the space Σ_N. This is random because H_N is so. Can we say anything about their limit over N. Of course this needs careful formulation, since these Gibbs measures live on different spaces. Thirdly, the ground state in any system is the configuration having minimum energy and the ground state energy is the minimum possible energy. Thus we are looking at $\min\limits_{\sigma} H_N(\sigma)$ and the states σ where it is attained. Can we say anything about their asymptotics?

The first issue, loosely put, says that a random quantity (like $\frac{1}{N} \log Z_N$) which depends smoothly on a large number of independent variables, but not too much on any one of them is essentially constant. Formulated this way it leads to an extremely rich and powerful theory [52]. The third issue formulated as an optimization problem with random quantities, again belongs to a bigger picture, known as, combinatorial optimization theory [54].

2. Examples of Hamiltonians:

The N particle configuration space is denoted by Σ_N. The Hamiltonian is defined for σ in this space. We start with classical non-random Hamiltonians.

Curie-Weiss Ising Model:

We have two real numbers J and h, called coupling constant and strength of the external field, respectively.

$$\Sigma_N = \{-1,+1\}^N; \qquad H_N(\sigma) \;=\; \frac{J}{2}\left(\sum_1^N \sigma_i\right)^2 \;+\; h\sum_1^N \sigma_i$$

When $J > 0$, this model is called ferro-magnetic.

Curie-Weiss Potts model:

We have a real number J, an integer $q > 1$, $S = \{1, 2, \cdots, q\}$, and h is a map from S to R.

$$\Sigma_N = S^N; \qquad H_N(\sigma) \;=\; \frac{J}{2}\sum_{i,j=1}^N I_{\sigma_i=\sigma_j} \;+\; \sum_{i=1}^N h(\sigma_i)$$

Here $I_{\sigma_i=\sigma_j}$ is one or zero according as $\sigma_i = \sigma_j$ or not.

Curie-Weiss Clock model:

The numbers J and q are as above. Now S is the set of q-th roots of unity, h is a function from S to R.

$$\Sigma_N = S^N; \qquad H_N(\sigma) \;=\; \frac{J}{2}\sum_{i,j=1}^N \sigma_i\sigma_j \;+\; \sum_{i=1}^N h(\sigma_i)$$

Heisenberg Model:

As earlier J is a real number. Now S is the surface of the $(r+1)$-dimensional unit ball; h is a nice map from S to R.

$$\Sigma_N = S^N; \qquad H_N(\sigma) \;=\; \frac{J}{2}\sum_{i,j=1}^N \sigma_i\cdot\sigma_j \;+\; \sum_{i=1}^N h(\sigma_i)$$

(what if $r = 0$?) Here $\sigma_i \cdot \sigma_j$ is their inner product.

Derrida's Random Energy Model:

Here $\Sigma_N = \{-1,+1\}^N$ and $H_N(\sigma)$ are independent centered Gaussian random variables with variance N. This is also called Gaussian REM.

Derrida's Generalized Random Energy Model:

We fix an integer $n \geq 1$, called the level of the GREM. We have numbers a_i for $1 \leq i \leq n$, called weights. For each N, we have positive integers $k(i,N)$ for $1 \leq i \leq n$ adding upto N. Clearly, $\Sigma_N = \{-1,+1\}^N = \prod_i \{-1,+1\}^{k(i,N)}$. With such a representation, $\sigma \in \Sigma_N$ is thought of as $(\sigma_1\sigma_2\cdots\sigma_n)$. It is assumed that for each i the limit, $\lim_N k(i, N)/N$ exists and is strictly positive. For $\sigma_1 \in \{-1, +1\}^{k(1,N)}$ we have a random variable ξ_{σ_1}, for each $\sigma_1\sigma_2 \in \{-1, +1\}^{k(1,N)} \times \{-1, +1\}^{k(2,N)}$ we have a random variable $\xi_{\sigma_1\sigma_2}$, and so on, finally we have $\xi_{\sigma_1\sigma_2\cdots\sigma_n}$. All these are centered Gaussian with variance N. More important is that these are all independent.

$$\Sigma_N = \{-1, +1\}^N = \prod_i \{-1, +1\}^{k(i,N)}; \qquad H_N(\sigma_1\sigma_2\cdots\sigma_n) = \sum_{i=1}^{n} a_i \xi_{\sigma_1\sigma_2\cdots\sigma_i}$$

This is also called Gaussian GREM.

SK Model:

SK stands for Sherrington and Kirkpatrick. For $1 \le i < j \le N$ we have centered Gaussian variable J_{ij} with variance N. These are all independent.

$$\Sigma_N = \{-1, +1\}^N; \qquad H_N(\sigma) = \frac{1}{N} \sum_{1 \le i < j \le N} J_{ij}\sigma_i\sigma_j.$$

This is also called Gaussian SK model. Sometimes it is convenient to express $H_N = \frac{1}{\sqrt{N}} \sum_{1 \le i < j \le N} \xi_{ij}\sigma_i\sigma_j$ where the ξ are independent standard normals.

Hopfield Model:

We have an integer $p \ge 1$, we have independent random variables ξ_i^μ for $1 \le \mu \le p$ and $1 \le i \le N$. We put for $i \ne j$, $J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$.

$$\Sigma_N = \{-1, +1\}^N; \qquad H_N(\sigma) = -\frac{1}{2} \sum_{i \ne j} J_{ij}\sigma_i\sigma_j$$

Viana-Bray Model:

We have a number $\alpha > 0$, have a random variable ξ which is Poisson with parameter αN, have a sequence (J_μ) of i.i.d. symmetric random variables, have sequences i_μ and j_μ i.i.d random variables uniformly distributed over $\{1, 2, \cdots, N\}$. ALL these random variables are independent.

$$\Sigma_N = \{-1, +1\}^N; \qquad H_N(\sigma) = -\sum_{\mu=1}^{\xi} J_\mu \sigma_{i_\mu}\sigma_{j_\mu}.$$

Curie-Weiss Spin glass model:

This is just Hopfield model with $p = 1$.

There are several other models. Bethe lattice Model (for the N particle system) has the spins living on a random graph with connectivity constraints.

For example one can take the set of all graphs on N vertices with zN edges equipped with uniform probability; or graph on N vertices where i and j are connected with probability $z/(N-1)$. Here z is an appropriate parameter. The idea is that as $N \to \infty$ the distribution of the coordination number is Poisson(z), that is for large N, the total number of links is Nz. Hamiltonian is $\sum J_{ij}\sigma_i\sigma_j$ where the J are independent centered Gaussian with variance $z^{-1/2}$ and the sum is over all edges in the graph. Thus only spins connected by an edge interact. In Edwards-Anderson Model the spins live on d-dimensional integer lattice, only nearest neighbours interact and the interaction variables are centered Gaussian with variance $d^{-1/2}$. In Large Range Edwards-Anderson Model the spins live on d-dimensional integer lattice, only spins at distance less than R interact and the interaction variables are centered Gaussian with variance $R^{-d/2}$.

Remark: See Derrida [23, 24, 25], Ligget etal [42], Parisi [45], Guerra etal [35].

3. Free Energy:

3.1 REM

Definition: A sequence of probabilities (μ_n) on a polish space X is said to satisfy Large Deviation Principle with rate funcion I if

(i) $I : X \to [0, \infty]$ is not identically infinity and is such that for each $a < \infty$, the set $(x : I(x) \leq a)$ is a compact set ;

(ii) for every closed set F, $\limsup \frac{1}{n} \log \mu_n(F) \leq - \inf\limits_{x \in F} I(x)$ and for every open set G, $\liminf \frac{1}{n} \log \mu_n(G) \geq - \inf\limits_{x \in G} I(x)$. ∎

There are several reasons why this notion is important. Here is one.

Theorem 1 (Varadhan's Lemma):

If $(\mu_n) \sim LDP(I)$ and h is a bounded continuous function on X, then $\frac{1}{n} \log \int e^{-nh(x)} d\mu_n(x) \to - \inf\{h(x) + I(x)\}$. ∎

Here is one way to get the rate function.

Theorem 2 (Getting rates):

Let (μ_n) be a sequence of probabilities on a polish space X, all supported on a fixed compact set. Let $\mathcal{A}$ be a countable open base for the topology of X. Assume that for each $A \in \mathcal{A}$ the limit, $\lim \frac{1}{n} \log \mu_n(A)$ exists and equals, say $-L(A)$, where $0 \leq L(A) \leq \infty$. Set, $I(x) = \sup\{L(A) : x \in A \in \mathcal{A}\}$. Then $(\mu_n) \sim LDP(I)$. ∎

Why are we interested in this? Now let us return to REM. For each N, we have 2^N many i.i.d random variables $H_N(\sigma)$ Gaussian with mean zero and variance N and $Z_N = \sum_\sigma e^{-\beta H_N(\sigma)}$. Just to recap, here the summation is over all $\sigma \in 2^N$, space of sequences of +1 and -1 of length N, which is the configuration space of the N-particle system. We are interested in showing the existence of the limit, $\lim_N \frac{1}{N} \log Z_N$ and calculating it. For fixed ω define the map $\sigma \mapsto \frac{1}{N} H_N(\sigma)(\omega)$ from Σ_N to R. Let $\mu_N(\omega)$ be the induced probability on R when Σ_N has uniform probability.

Put $\Psi = \{x \in R : x^2 \leq 2 \log 2\}$. Define $I(x) = x^2/2$ for $x \in \Psi$ and ∞ for x not in Ψ.

Theorem 3 (REM large deviation rate):

For almost every sample point ω, the sequence of probabilities $(\mu_N(\omega))$ satisfies LDP(I). ∎

This combined with Varadhan's lemma leads immediately to the following.

Theorem 4 (REM energy):

Consider REM where for the N-particle system the Hamiltonians in different configurations are independent centered gaussian with variance N. Let $Z_N(\beta)$ be the partition function. Then $\lim \frac{1}{N} \log Z_N(\beta)$ exists almost surely and equals $\log 2 + \frac{1}{2}\beta^2$ for $0 \leq \beta \leq \sqrt{2 \log 2}$ and equals $\beta\sqrt{2 \log 2}$ for $\beta > \sqrt{2 \log 2}$. ∎

In situation as above, one says that there is phase transition at the critical value $\beta_c = \sqrt{2 \log 2}$. Note that for small β the limit is quadratic in β where as for large β it is linear. If $p(\beta)$ denotes the above limit, then a uniform integrability argument leads to the following.

Theorem 5 (REM Energy, L^p-convergence):

For any β, $\frac{1}{N} \log Z_N(\beta)$ converges almost surely as well as in L^P $(1 \leq p < \infty)$ to the nonrandom value $p(\beta)$ above. ∎

Proof of Theorem 1:

It is convenient to use the notation $I(A)$ for infimum of I over the set A.

First we claim that $\liminf \frac{1}{n} \log \int e^{-nh} d\mu \geq \sup\{-h(x) - I(x)\}$. So fix x_0 and $\epsilon > 0$. Take an open ball G around x_0 in which $h(x) < h(x_0) + \epsilon$. Since $\int_X \geq \int_G$, the required liminf is at least $-h(x_0) - \epsilon - I(G) \geq -h(x_0) - I(x_0) - \epsilon$. True for every $\epsilon > 0$ and every x_0, we are done.

Now we claim that $\limsup \frac{1}{n} \log \int e^{-nh} d\mu \leq \sup\{-h(x) - I(x)\} = \Lambda$, say. No loss to assume that $\Lambda < \infty$. Fix $M > 0$ so that $h \geq -M$. Fix $a > 0$ so that $M - a < \Lambda$. Fix $\epsilon > 0$. Define the compact set $K = \{x : I(x) \leq a\}$. Recall that h is continuous and for each b, the set $(I > b)$ is an open set. For each $x \in K$, fix an open ball G_x centered at x, through out which h is at least $h(x) - \epsilon$ and I is at least $I(x) - \epsilon$. Let B_x be the open ball centered at x and radius half that of G_x. These balls cover K. Take a finite subcover, say (B_i), with B_i centered at x_i. Set F_i to be closure of B_i and F to be the complement of $\cup B_i$. For any i,

$$\limsup \frac{1}{n} \log \int_{F_i} e^{-nh} d\mu_n \leq -h(x_i) + \epsilon - I(F_i) \leq -h(x_i) - I(x_i) + 2\epsilon \leq \Lambda + 2\epsilon$$

This is true for each F_i. Moreover F being a subset of K^c, $I(F) \geq a$ and hence

$$\limsup \frac{1}{n} \log \int_{F} e^{-nh} d\mu_n \leq M - I(F) \leq M - a \leq \Lambda$$

Observe that the integrand being non-negative, $\int_X \leq \sum \int_{F_i} + \int_F$. Now to complete the proof, use the fact that $\limsup\limits_n \frac{1}{n} \log \sum\limits_i a_{in} \leq \max\limits_i \; \limsup\limits_n \; \frac{1}{n} \log a_{in}$. This last fact can be proved as follows. Let $M_n = \max\limits_{1 \leq i \leq k} a_{in}$. Then $\frac{1}{n} \log M_n \leq \frac{1}{n} \log \sum\limits_i a_{in} \leq \frac{1}{n} [\log k + \log M_N]$. Since $\limsup\limits_n \frac{1}{n} \log M_n = \max\limits_i \; \limsup\limits_n \; \frac{1}{n} \log a_{in}$, we are done.

Proof of Theorem 2:

Clearly I maps X to $[0, \infty]$. If $I(x_0) > a$, then there is $A \in \mathcal{A}$ such that $x_0 \in A$ and $L(A) > a$. In particular, at all points of A, I value is larger than a. So the set $(I \leq a)$ is a closed set. If K is the compact set on which all the probabilities are supported, then clearly, $(I < \infty) \subset K$. This shows that for any $a < \infty$, the set $(I \leq a)$ is a compact set.

Let G be any open set. Need to show that $\liminf \frac{1}{n} \log \mu_n(G) \geq -I(G)$. Pick $x \in G$ and $A \in \mathcal{A}$ with $x \in A \subset G$. $\liminf \frac{1}{n} \log \mu_n(G) \geq \liminf \frac{1}{n} \log \mu_n(A) = -L(A) \geq -I(x)$. This being true for every $x \in G$ we have $\liminf \frac{1}{n} \log \mu_n(G) \geq \sup\limits_{x \in G} \{-I(x)\} = -I(G)$.

Let F be any closed set. Shall show $\limsup \frac{1}{n} \log \mu_n(F) \leq -I(F)$. If F does not intersect K, the compact set on which all the probabilities are supported, then both sides are $-\infty$. So need to consider the case when F has points of K. In this case clearly, the left side of the inequality remains unaltered when F is replaced by $F \cap K$. Right side also remains unaltered because I being infinity outside K, the infimum of I over F is same as that over $F \cap K$. Thus safely assume that F is a compact set.

Temporarily assume that for all $x \in F$, $I(x) < \infty$. Fix $\epsilon > 0$. For each $x \in F$, pick $A(x) \in \mathcal{A}$ such that $x \in A(x)$ and $I(x) \leq L(A(x)) + \epsilon$. These sets cover F, get a finite subcover $(A(x_i))$. Now $\limsup \frac{1}{n} \log \mu_n(F) \leq \limsup \frac{1}{n} \log \sum \mu_n(A(x_i)) \leq \max\limits_i \{-L(A(x_i))\} \leq \max\limits_i \{-I(x_i)\} + \epsilon \leq \sup\limits_{x \in F} \{-I(x)\} + \epsilon$ giving the desired result.

To remove the condition that I is finite on F, proceed as follows. Fix $M > 0$ and set I_M as minimum of I and M so that for any x, $I(x) \geq I_M(x)$. Proceed as above with I replacing I_M to conclude that $\limsup \frac{1}{n} \log \mu_n(F) \leq -I_M(F)$. This is true for every $M > 0$ and as M increases, $I_M(F)$ increases to $I(F)$ completing the proof.

Proof of Theorem 3:

Since, all the $H_N(\sigma)$ have the same distribution, we let H_N stand for a random variable with this distribution. Fix an interval $J = (a,b)$ and put $q_N = P(\frac{1}{N} H_N \in J)$. Let $m = \inf_J |x|$ and $M = \sup_J |x|$. Clearly $J \subset [-M,-m] \cup [m,M]$. If $M = \infty$ make these intervals open at $\mp M$.

$$q_N = \int_{\sqrt{N}a}^{\sqrt{N}b} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \leq \frac{2}{\sqrt{2\pi}} \int_{\sqrt{N}m}^{\infty} e^{-x^2/2} dx \leq \frac{1}{\sqrt{N}m} e^{-Nm^2/2}.$$

If $m = 0$, take this bound as one. If $0 < \delta < M - m$, then

$$q_N \geq \int_{\sqrt{N}m}^{\sqrt{N}(m+\delta)} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \geq \frac{1}{\sqrt{2\pi}} \sqrt{N} \delta e^{-N(m+\delta)^2/2}.$$

Claim 1: If $[a,b] \cap \Psi = \emptyset$, then for almost every sample point ω we have eventually $\mu_N(a,b) = 0$.

Indeed if m and M are the min and max (in modulus) of the interval $[a,b]$ then hypothesis and the fact that Ψ is a symmetric interval imply that $m^2 > 2\log 2$. In particular $m \neq 0$. The probability that one of the points H_N/N is in (a,b) is at most $2^N \frac{1}{\sqrt{N}\, m} e^{-Nm^2/2} = \frac{1}{\sqrt{N}\, m} e^{-\frac{N}{2}(m^2 - 2\log 2)}$ which is summable over N and Borel-Cantelli completes the proof.

Claim 2: If (a,b) contains a point x with $x^2 < 2\log 2$, then for almost every sample point ω we have eventually the following: for any $\epsilon > 0$; $(1-\epsilon) q_N \leq \mu_N(a,b) \leq (1+\epsilon) q_N$.

Pick x as in the hypothesis and m be as earlier. Since x and $-x$ are in Ψ and $m \leq |x|$, we conclude that $m^2 < 2\log 2$. Fix δ such that $(m+\delta)^2 < 2\log 2$ and $0 < \delta < M - m$. Recall that $\mu_N(a,b) = \frac{1}{2^N} \sum_\sigma I_{(a,b)}(\frac{H_N(\sigma)}{N})$. Denoting by E expectation w.r.t. ω, we have

$$E\{\mu_N(a,b)\} = q_N$$

$$Var\{\mu_N(a,b)\} = E \frac{1}{2^{2N}} \sum_{\sigma,\tau} I_{(a,b)}(\frac{H_N(\sigma)}{N}) I_{(a,b)}(\frac{H_N(\tau)}{N}) - q_N^2 \leq \frac{q_N}{2^N}$$

where the last inequality is arrived at after cancelling the $\sigma \neq \tau$ terms of the first sum with the second term. As a consequence, chebyshev gives

$$P\left[|\mu_N(a,b) - q_N| > \epsilon q_N\right] \;\leq\; \epsilon^{-2} 2^{-N} q_N^{-1} \;\leq\; \epsilon^{-2} \frac{2}{\delta} e^{-N[\log 2 - \frac{(m+\delta)^2}{2}]}.$$

Since the last expression is summable over N, Borel-Cantelli completes the proof of claim 2.

To complete the proof of the theorem, let $\mathcal{A}$ denote the collection of all open intervals (a,b) such that a and b are either rational or $\pm\infty$, but different from the two points $-\sqrt{2\log 2}$ and $+\sqrt{2\log 2}$. This family forms a base for the topology of the real line. If $A \in \mathcal{A}$ and closure of A is disjoint with Ψ then claim 1 shows that $\lim \frac{1}{N} \log \mu_N(A)$ exists and equals $-\infty$. If $|x| > \sqrt{2\log 2}$, then we clearly have sets A as above containing the point x and so for such points x, we have $\sup\limits_{x \in A \in \mathcal{A}} L(A) = \infty$. Now consider a point x with $x^2 \leq 2\log 2$. Take any set $A \in \mathcal{A}$ with $x \in A$. Then hypothesis of claim 2 holds for the interval A and hence, for any fixed $\epsilon > 0$, its conclusion holds. The bounds obtained for q_N at the beginning of the proof show that $\lim_N \frac{1}{N} \log q_N$ exists and equals $-m^2/2$. As a consequence, $\lim \frac{1}{N} \log \mu_N(A)$ exists and equals $-m^2/2$. It is now easy to see that, $\sup\limits_{x \in A \in \mathcal{A}} L(A) = x^2/2$. This is so for all points x with $x^2 \leq 2\log 2$. All this happens for almost every sample point ω (Remember μ_N is random) because $\mathcal{A}$ is a countable family. Theorem 2 applies completing the proof.

Proof of Theorem 4:

Use Varadhan's lemma for the sequence $(\mu_n(\omega))$ and the function $-\beta x$ along with the above theorem to conclude that for almost every sample point the required limit exists and equals $\log 2 + \sup\{\beta x - \frac{1}{2}x^2 : x^2 \leq 2\log 2\}$. We add $\log 2$ because Z_N is the sum of all the Gibbs factors, not their average. The required sup is same as sup of $\beta x - \frac{1}{2}x^2$ over $0 \leq x \leq \sqrt{2\log 2}$. If $\beta \leq \sqrt{2\log 2}$ then this sup is attained at $x = \beta$ and the value of the sup is $\frac{1}{2}\beta^2$. However if $\beta \geq \sqrt{2\log 2}$, then this sup is attained at $\sqrt{2\log 2}$ with value $\beta\sqrt{2\log 2} - \log 2$.

Proof of Theorem 5:

We show that the family $(\frac{1}{N} \log Z_N)$ is L^p bounded. First observe that if $-\xi_\sigma$ for $\sigma \in 2^N$ are i.i.d. standard normal, then the Hamiltonian $H_N(\sigma) = -\sqrt{N}\xi(\sigma)$. If M_N denotes the max of the 2^N standard normals ξ_σ, then clearly $\beta M_N/\sqrt{N} \leq \frac{1}{N} \log Z_N \leq \log 2 + \beta M_N/\sqrt{N}$. As a consequence we need to show that the sequence $(M_N/\sqrt{N})$ is L^p bounded. Fix $1 \leq p < \infty$.

$$E\left[\left(\frac{M_N}{\sqrt{N}}\right)^p I_{M_N > 0}\right] = \int_0^\infty P\left(\frac{M_N}{\sqrt{N}} > x^{1/p}\right) dx$$

$$P(M_N > \sqrt{N}x^{1/p}) \leq 2^N \int_{\sqrt{N}x^{1/p}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \leq C\ x^{-1/p}\ e^{-N[\frac{1}{2}x^{2/p} - \log 2]}$$

Let $a = (4\log 2)^{p/2}$ so that if $x > a$, then $\frac{1}{2}x^{2/p} - \log 2 > \frac{1}{4}x^{2/p}$. Thus

$$E\left[\left(\frac{M_N}{\sqrt{N}}\right)^p I_{M_N>0}\right] \leq a + C\int_a^{\infty} x^{-1/p} e^{-\frac{1}{4}x^{2/p}} dx$$

which is a finite number not depending on N. Also

$$\int_0^{\infty} P\left(M_N < -\sqrt{N}x^{1/p}\right) dx \leq \int_0^{\infty} P(\xi > \sqrt{N}x^{1/p}) dx$$

which is at most

$$1 + \int_1^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{N}x^{1/p}} e^{-\frac{1}{2}Nx^{2/p}} dx$$

showing that $E\left[\left(\frac{M_N}{\sqrt{N}}\right)^p I_{M_N<0}\right]$ is also bounded in N. This completes the proof of uniform integrability.

Remarks:

1. In the definition of rate function, usually one only demands that for each a, the set $\{x : I(x) \leq a\}$ is a closed set, in other words, that I is lower semi-continuous. Rate functions for which these sets are compact are called *good rate functions*.

2. If in the definition of LDP, the inequality required for all closed sets is demanded only for compact sets (the inequality for open sets is, of course, demanded for all open sets) then one says *weak LDP* holds.

3. In the definition of LDP, there are several indexings used, like n, ϵ, t, and normalizations $\frac{1}{n}$, a_n or a_ϵ etc. We have used n and $1/n$. It all depends on how our probabilities are indexed and the normalization needed.

4. There is no need to use the same function h for all n in Varadhan's lemma, we can have h_n but with some extra conditions.

5. There is no need to assume that all the μ_n are supported on a fixed compact set in Theorem 2. One can assume exponential tightness etc.

6. Discuss Theorem 2, when μ_n has mass $1/2$ at n and $1/2$ at zero.

7. We discussed REM only with Gaussian distributions, one could use other distributions too.

8. For LDP see Varadhan [60, 61] and Dembo etal [22]. Its use in statistical mechanics was systematically explored by Ellis [29], Eisele [28]. For the energy of REM and its fluctuations see Olivieri etal [43], Dorlas etal [27], Jana [36, 37], Galves etal [31], Talagrand [59], Derrida [23] and Ben Arous etal [2].

3.2 GREM:

In REM, 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, where the branches of the tree are in correspondence with the configuration space. We first recall GREM.

Fix an integer $n \geq 1$. This will be called the level of the GREM. 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 $\{-1, +1\}^N$, denoted as 2^N. Consider integers $k(i, N)$ for $1 \leq i \leq n$ such that 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. Imagine an n-level rooted tree. There are $2^{k(1,N)}$ nodes at the first level. These will be denoted as σ_1, for $\sigma_1 \in 2^{k(1,N)}$. 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 $\sigma_1\sigma_2$ for $\sigma_2 \in 2^{k(2,N)}$. 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-to-one correspondence with 2^N, the configuration space. At the edge ending with 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. centered Gaussian with variance N. We associate one weight for each level, say weight $a_i > 0$ for the i-th level. It is assumed tht $\sum_i a_i^2 = 1$. These are not random. In a configuration $\sigma = \sigma_1 \cdots \sigma_n$ the Hamiltonian is

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

For $\beta > 0$ the partition function is

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

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.

As in the case of REM, the random probabilities μ_N are defined on R^n by transporting the uniform distribution of $2^N = 2^{k(1,N)} \times \cdots \times 2^{k(n,N)}$ to R^n via the map

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

It is easy to see that $\mu_N \Rightarrow \delta_0$ a.s. as $N \to \infty$. Here δ_0 is the point mass at the zero vector of R^n. We assume from now on that $\frac{k(i,N)}{N} \to p_i > 0$ for $1 \le i \le n$. Let

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

We define the map $I : R^n \to R$, by $I(\tilde{x}) = \frac{1}{2} \sum\limits_{i=1}^{n} x_i^2$ if $\tilde{x} \in \Psi$ and $= \infty$ otherwise.

Theorem 6 (GREM large deviation rate):

For almost every sample point, the (random) sequence $\{\mu_N\}$ satisfies LDP with rate function I. ∎

Theorem 7 (GREM Energy):

There are numbers $0 < \beta_1 < \cdots < \beta_K < \infty = \beta_{K+1}$ and numbers $0 < r_1 < \cdots < r_K = n$ such that almost surely,

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

∎

The exact computation of the numbers β_j is in the proof of the Theorem. Two simple cases are worth mentioning. The numbers β_j mentioned below are same as the above, in these particular cases.

Theorem 8 (GREM Energy, Special Cases):

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.

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

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.

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

Note that in case i) above all the n levels of the GREM showed up in the formula for the energy where as in case ii) all the levels reduced to just one level, leading to REM situation. Let us denote by $p(\beta)$ the non-random limit in Theorem 7.

Theorem 9 (GREM Energy, L^p-convergence):

For any β, $\frac{1}{N}\log Z_N(\beta)$ converges almost surely as well as in L^P $(1 \leq p < \infty)$ to the nonrandom value $p(\beta)$ above. $\blacksquare$

Proof of Theorem 6:

The proof is similar to that of Theorem 3. Here are the main steps. Let $J = J_1 \times \cdots \times J_n$, where each J_i is an interval of R. Put $m_i = \inf_{x\in J_i}|x|$, $M_i = \sup_{x\in J_i}|x|$, and $q_{iN} = P(\frac{\xi}{N} \in J_i)$ we have

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

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

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

for any $0 < \delta < M_i - m_i$.

Firstly, one shows that if $\bar{J} \cap \Psi = \emptyset$, then a.s. eventually $\mu_N(J) = 0$. Moreover, the sequence $\{\mu_N\}$ is supported on a compact set.

Secondly, if $(\bar{J} \cap \Psi)^0 \neq \emptyset$, then for any $\epsilon > 0$ a.s. eventually

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

Finally, one shows that almost surely

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

This completes a sketch of a proof of the theorem.

Proof of theorem 7:

The free energy is given by

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

where Ψ^+ consists of points of Ψ with all coordinates non-negative.

Here is the general idea. Let $c_1, c_2, \cdots, c_n \geq 0$ with $c_1 > 0$. Let $\tilde{\alpha} = (\alpha_1, \cdots, \alpha_n) \in R^n$ with each $\alpha_i > 0$. Let $S \subset R^n$ be the set of all points $\tilde{x} = (x_1, \cdots, x_n) \in R^n$ with nonnegative coordinates and $\sum_1^i x_j^2 \leq \sum_1^i c_j$ for $i = 1, 2, \cdots, n$. 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 $\gamma = \min\limits_i \frac{c_1+\cdots+c_i}{\alpha_1^2+\cdots+\alpha_i^2}$. Let k be the largest index such that $\gamma = \frac{c_1+\cdots+c_k}{\alpha_1^2+\cdots+\alpha_k^2}$. Assume that $\frac{c_{k+1}+\cdots+c_i}{\alpha_{k+1}^2+\cdots+\alpha_i^2} \geq 1$, for $i > k$. Put $\tilde{\alpha}^* = (\alpha_1^*, \cdots, \alpha_n^*)$ 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

$$\sum_1^k (x_i-\alpha_i)^2 - \sum_1^k(\alpha_i^*-\alpha_i)^2 \geq \sum_1^k (x_i-\alpha_i^*)^2 \geq 0. \qquad (*)$$

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

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 the above analysis, let us put $B_{j,k} = \sqrt{\frac{(p_j + \cdots + p_k) 2 \log 2}{a_j^2 + \cdots + a_k^2}}$ for $1 \leq j \leq k \leq n$. Set

$$
\begin{array}{llll}
\beta_1 & = \min_k B_{1,k} & r_1 & = \max\{i : B_{1,i} = \beta_1\} \\
\beta_2 & = \min_{k > r_1} B_{r_1+1,k} & r_2 & = \max\{i > r_1 : B_{r_1+1,i} = \beta_2\}
\end{array}
$$

and in general

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

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

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

$$
\begin{array}{rcll}
x_i^* & = & \beta_l a_i & \text{if } i \in \{r_{l-1}+1, \cdots, r_l\} \text{ for some } l, 1 \leq l \leq j \\
 & = & \beta a_i & \text{if } i > r_j + 1.
\end{array}
$$

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

Proof of Theorem 8:

For the first part one only needs to see that the hypothesis implies that the constants β as defined above coincide with the values given in the statement of the theorem. The formula itself, for the energy, is just a special case of the one in the earlier theorem. For the second part one only needs to realize that in this case there is only one β_i.

Proof of Theorem 9:

The argument given for REM applies to show uniform integrability in this set-up as well.

Remarks:

1. It is possible to take, as in the case of REM, distributions other than Gaussian. In fact it is possible to take different distributions at different levels of the tree. However, explicit formulae for the energy appear to be difficult.

2. One can give a general tree formulation of the GREM and solve the resulting model. This allows one to consider random trees as well. For example, toss an n-faced die N times to decide what should be the numbers $k(i, N)$. However, no interesting trees that exhibit any behaviour other than expected have been found. May be there is some universality w.r.t. the trees.

3. An enirely new formulation in terms of rate functions of the driving distributions can be given to these models.

4. There are other models akin to GREM that are considered in the literature.

5. For the energy of GREM see Derrida [24], Derrida etal [25], Dorlas etal [26], Capocaccia etal [18], Contucci etal [21], Galves etal [31], Jana [37], and Jana etal [38, 39], Bolthausen etal [9].

3.3 SK Model:

For the SK model, to recap, we have

$$H_N(\sigma) = \frac{1}{\sqrt{N}} \sum_{1 \le i < j \le N} J_{ij}\sigma_i\sigma_j; \quad Z_N = \sum_{\sigma} e^{-\beta H_N(\sigma)}$$

To show the existence of the limit, $\lim_N \frac{1}{N} \log Z_N$ the method used is called the smart path method. Roughly speaking, if you have two quantities a and b computed in two systems and want to show that $a \le b$, one way is to find a path $\varphi(t)$ for $0 \le t \le 1$, (between the two systems that produced the values a and b respectively) such that φ is increasing; $\varphi(0) = a$ and $\varphi(1) = b$. Such a path is definitely a smart path! We need some preliminaries about Gaussian expectations.

Theorem 10 (integration by parts):

(i) Let X be centered Gaussian and $F : R \to R$ be a C^1 function such that both F and its derivative F' have exponential growth, that is, for some constants c and d, $|F(x)| \le ce^{d|x|}$ and similarly for F'. Then

$$E(X \cdot F(X)) = E(X^2) \cdot E(F'(X)).$$

(ii) Let $X, Y_1, Y_2, \cdots, Y_n$ be jointly Gaussian and centered. $F : R^n \to R$ be a C^1 function with exponential growth, that is, $|F(y)| \le ce^{d\Sigma|y_i|}$ and similarly for its first partial derivatives. Then denoting by F_i the derivative of F w.r.t. the i-th coordinate, and by Y the vector $(Y_1, \cdots, Y_n)$,

$$E(X \cdot F(Y)) = \sum_i E(X \cdot Y_i)E(F_i(Y)).$$

∎

Theorem 11 (Slepian's Lemma):

Let $F : R^M \to R$ be a C^2 function of exponential growth along with its first two derivatives. Assume that for $i \ne j$, $\frac{\partial^2}{\partial x_i \partial x_j} F \ge 0$. Let $U = (U_1, \cdots, U_M)$

and $V = (V_1, \cdots, V_M)$ be centered Gaussian with $E(U_i^2) = E(V_i^2)$ for all i, and $E(U_iU_j) \geq E(V_iV_j)$ for $i \neq j$. Then $E(F(U)) \geq E(F(V))$. ∎

Theorem 12 (Gaussian concentration inequality):

Let $F : R^M \to R$ be a C^2 function such that for some number $A > 0$, we have for all x and y; $|F(x) - F(y)| \leq A\ d(x, y)$. Let $X = (X_1, \cdots, X_M)$ where (X_i) are independent standard normal. Then for any $t > 0$,

$$P\{|F(X) - E[F(X)]| \ \geq\ t\} \quad \leq \quad 2e^{-t^2/4A^2}. \qquad ∎$$

Here d is the usual Euclidean distance. Results which assert that a random variable, with high probability, takes values near a number (mean, median) are called concentration inequalities.

Theorem 13 (Concentration inequality for SK model):

Consider the SK model with inverse temperature β. Let us denote $p_N(\beta) = \frac{1}{N} E \log Z_N(\beta)$. Then for any $t > 0$,

$$P\{|\frac{1}{N} \log Z_N - p_N| \ \geq\ t\} \quad \leq \quad 2e^{-Nt^2/2\beta^2}. \qquad ∎$$

Inequalities like the above are very useful, they allow us to deduce the almost sure convergence of the sequence $\frac{1}{N} \log Z_N$ from the convergence of their means.

Theorem 14 (SK model, Energy):

The sequence $(E[\log Z_N])$ is a superadditive sequence of numbers. $\frac{1}{N} E[\log Z_N]$ converges to a finite limit. The sequence $\frac{1}{N} \log Z_N$ converges almost surely. ∎

Neither the convergence nor the value of the limit seem to depend on the Gaussian nature of the environment (a term used to describe the randomness that enters the Hamiltonian). There is a universal behaviour. Anything reasonable seems to do.

Theorem 15 (Universality of Energy):

Consider any probability μ on R with mean zero, variance one and finite third moment. Consider the 'SK' Hamiltonian and partition function,

$$H_N(\sigma) = \sum_{1 \leq i < j \leq N} J_{ij} \sigma_i \sigma_j; \quad Z_N = \sum_{\sigma} e^{-\frac{\beta}{\sqrt{N}} H_N(\sigma)}$$

where J_{ij} are i.i.d. having distribution μ. Then $\frac{1}{N} \log Z_N$ converges almost surely, $\frac{1}{N} E(\log Z_N)$ converges. Moreover this limit is same as that for the Gaussian environment. ∎

Proof of Theorem 10:

Condition on F says required integrals below exist. First part follows by observing that if $EX^2 = \sigma^2$, then integration by parts gives

$$E(X \cdot F(X)) = \int \frac{1}{\sqrt{2\pi}\sigma} x e^{-x^2/2\sigma^2} F(x) dx = \sigma^2 \int \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} F'(x) dx.$$

For the second part put $Z_i = Y_i - c_i X$ where $c_i = E(Y_i X)/E(X^2)$. See that $Z = (Z_1, \cdots, Z_n)$ is independent of X. Apply earlier part to X and the function $\varphi(x) = F(\langle z_i + c_i x \rangle)$ for fixed numbers (z_i). We get (remember chain rule for derivative) $E(X\varphi(X)) = E(X^2)E(\varphi'(X)) = \sum E(Y_i X) E_X(F_i(\langle z_i + c_i X \rangle))$. Since Z is independent of X, now taking expectation w.r.t. Z we get $E(XF(Y)) = \sum E(Y_i X)\ E(F_i(Y))$.

Proof of Theorem 11:

There is no loss to assume that the families U and V are independent. Put $W(t) = \sqrt{t}U + \sqrt{1-t}V$ and $\varphi(t) = E[F(W(t))]$. Thus $W(t) = (W_1(t), \cdots, W_M(t))$ where $W_i(t) = \sqrt{t}U_i + \sqrt{1-t}V_i$. Under the assumptions of the theorem, we can interchange expectation with differentiation, yielding $\varphi'(t) = E[\sum W_i'(t) F_i(W(t))]$. Here F_i is the derivative of F w.r.t. its i-th coordinate and W_i' is the derivative w.r.t. t of the i-th coordinate of W. Thus $W_i'(t) = \frac{1}{2\sqrt{t}}U_i - \frac{1}{2\sqrt{1-t}}V_i$. Note that by independence of the Us and Vs, we have $E[W_i'(t) W_j(t)] = \frac{1}{2}[E(U_i U_j) - E(V_i V_j)]$. By previous theorem

$$E[W_i'(t)\ F_i(W(t))] = \sum_j E[W_i'(t)\ W_j(t)]\ E[F_{ij}(W(t))].$$

As a consequence

$$\varphi'(t) = \frac{1}{2} \sum [E(U_i U_j) - E(V_i V_j)] E[F_{ij}(W(t))]$$

By hypothesis, this quantity is positive. Thus $E(F(V)) = \varphi(0) \leq \varphi(1) = E(F(U))$.

Proof of Theorem 12:

Consider R^{2M} and think of a point $y = (y_1, \cdots, y_{2M})$ as (y^1, y^2) with $y^1 = (y_1, \cdots, y_M)$ and $y^2 = (y_{M+1}, \cdots, y_{2M})$. Define G on R^{2M} by $G(y) = e^{s[F(y^1) - F(y^2)]}$ for a fixed number $s \in R$. Let $(U_i)_{1 \leq i \leq 2M}$ be independent standard Gaussian. Fix independent standard gaussian $(V_i)_{1 \leq i \leq M}$ independent of the U family and also put $V_i = V_{i-M}$ for $M + 1 \leq i \leq 2M$. Note that $E(U_i U_j) - E(V_i V_j)$ is zero unless $|i - j| = M$, in which case it is -1. Set $W(t) = \sqrt{t}U + \sqrt{1-t}V$ where U and V are the $2M$ dimensional vectors (U_i)

and (V_i) respectively. Set $\varphi(t) = E[G(W(t))]$ so that as in the above theorem, we have

$$\varphi'(t) = \frac{1}{2}\sum_{i,j}[E(U_iU_j) - E(V_iV_j)]\, E[F_{ij}(W(t))] = -E\sum_{i\leq M} G_{i,i+M}(W(t))$$

where, as usual, the suffixes for F and G denote the corresponding partial derivatives. Observe that $G_{i,i+M}(y) = -s^2 F_i(y^1)F_i(y^2)G(y)$ so that

$$\varphi'(t) = s^2 E[\sum_{i\leq M} F_i(W^1(t))F_i(W^2(t))G(W(t)]$$

But for all $x \in R^M$ our hypothesis tells $\sum_{i\leq M}[F_i(x)]^2 \leq A^2$ So applying Cauchy-Schwarz for the F terms above, we get

$$\varphi'(t) \leq s^2 A^2 \varphi(t).$$

This combined with $\varphi(0) = 1$ yields $\varphi(1) \leq e^{s^2A^2}$. In other words, we have proved that for every $s \in R$, $Ee^{s[F(U^1)-F(U^2)]} \leq e^{s^2A^2}$. Writing this expectation E as $E^{U_1}E^{U_2}$ and performing E^{U_2} with the help of Jensen, we get

$$Ee^{s[F(U)-E(F(U))]} \leq e^{s^2A^2}$$

so that

$$P(F(U) - E[F(U)] \geq t) \leq e^{s^2A^2}/e^{st}.$$

Taking $s = t/(2A^2)$ we see that this last quantity is $e^{-t^2/(4A^2)}$. Apply this to $-F$ to complete the proof.

Proof of Theorem 13:

Take $M = N(N-1)/2$. For $\sigma \in \Sigma_N$, let $a(\sigma)$ be the vector of R^M given by $(-\frac{\beta}{\sqrt{N}}\sigma_i\sigma_j : 1 \leq i < j \leq N)$. Note that if J is a M dimensional standard normal vector, then the SK Hamiltonian is nothing but $H_N(\sigma) = J \cdot a(\sigma)$. If we consider the function $F(x) = \frac{1}{N}\log\sum_{\sigma} e^{a(\sigma)\cdot x}$ then $F(J) = \frac{1}{N}\log Z_N$. If we show that $|F(x)-F(y)| \leq \frac{\beta}{\sqrt{2N}}d(x,y)$, then the previous theorem completes the proof. But this is clear because, $\|a(\sigma)\| \leq \beta\sqrt{N/2}$ and hence $|a(\sigma)\cdot x - a(\sigma)\cdot y| \leq \beta\sqrt{N/2}d(x,y)$. Thus $a(\sigma)\cdot x \leq a(\sigma).y + \beta\sqrt{N/2}\, d(x,y)$ Take exponentials, add over σ and take log to get $F(x) \leq F(y) + d(x,y)\ \beta/\sqrt{2N}$. Interchange x and y to complete the proof.

Proof of Theorem 14:

Almost sure convergence of $\frac{1}{N}\log Z_N$ follows from the convergence of the numbers $\frac{1}{N}E(\log Z_N)$ by an application of Theorem 13. But the convergence of these numbers, in turn, follows from the super-additivity of the sequence of numbers $E(\log Z_N)$. This will be observed first.

Let (a_n) be a super-additive sequence of numbers, that is, $a_{m+n} \geq a_m + a_n$ for every m and n. Set $s = \sup_n(a_n/n)$. Clearly, $\limsup(a_n/n) \leq s$. We shall show that $\liminf(a_n/n) \geq s$, then it follows that (a_n/n) converges to s. First consider the case $s < \infty$. Fix $\epsilon > 0$. Fix k such that $(a_k/k) > s - \epsilon/2$. Fix $N > k$ such that for $1 \leq r \leq k$, $(a_r/N) > -\epsilon/2$. Now take any $n > N$. Let $n = kd + r$ with $r < k$. Then

$$a_n = a_{kd+r} \geq da_k + a_r \geq dk(s - \frac{\epsilon}{2}) + a_r$$

by super additivity and choice of k. Thus

$$\frac{a_n}{n} \;\geq\; \frac{dk}{dk+r}\left(s - \frac{\epsilon}{2}\right) + \frac{a_r}{n} \;\geq\; \left(s - \frac{\epsilon}{2}\right)\left(1 - \frac{r}{n}\right) - \frac{\epsilon}{2},$$

showing that $\liminf(a_n/n) \geq s - \epsilon$. If $s = \infty$, we fix $M > 0$ and need to show that $\liminf(a_n/n) > M$. Fix k as above but with $(a_k/k) > 2M$ and proceed.

Of course, in our case $s < \infty$. Indeed $\frac{1}{N}E(\log Z_N) \leq \log 2 + \frac{1}{2}\beta^2$. This is because by concavity of log function, Jensen yields $E(\log Z_N) \leq \log E(Z_N) = \log[2^N E(e^{-\beta X/\sqrt{N}})]$ where X is centered Gaussian with variance $N(N-1)/2$.

We now proceed to show that the sequence of numbers $\{E[\log Z_N]\}$ is indeed super-additive. Fix integers $N = N_1 + N_2$. Take independent standard standard normals (J_{ij}) for $1 \leq i < j \leq N$; (J'_{ij}) for $1 \leq i < j \leq N_1$; and (J''_{ij}) for $N_1 < i < j \leq N$. Set for $\sigma \in \{-1,+1\}^N$,

$$H_1(\sigma) = \frac{1}{\sqrt{N}} \sum_{1 \leq i < j \leq N} J_{ij}\sigma_i\sigma_j$$

$$H_2(\sigma) = \frac{1}{\sqrt{N_1}} \sum_{1 \leq i < j \leq N_1} J'_{ij}\sigma_i\sigma_j \;+\; \frac{1}{\sqrt{N_2}} \sum_{N_1+1 \leq i < j \leq N} J''_{ij}\sigma_i\sigma_j$$

and for $0 \leq t \leq 1$

$$H(t)(\sigma) = \sqrt{t}\, H_1(\sigma) + \sqrt{1-t}\, H_2(\sigma).$$

We use the notation of Slepian's lemma. We take $M = 2^N$; The family (U_σ) is the family $(H_1(\sigma))$; (V_σ) is the family $(H_2(\sigma))$; and the funcion $F(x) = \log \sum_\sigma e^{-\beta x_\sigma}$. Then (W_σ) is precisely the family $(H(t)(\sigma))$. As a result we have, denoting $\varphi(t) = E[F(W(t))]$,

$$\varphi'(t) = \frac{1}{2} \sum_{\sigma\eta} [E(U_\sigma U_\eta) - E(V_\sigma V_\eta)]\; E[F_{\sigma\eta}(W(t))].$$

Here $F_{\sigma\eta}$ is the derivative of F w.r.t. these variables. In the present case,

$$E(U_\sigma U_\eta) = \frac{1}{N}\sum_{i<j}\sigma_i\sigma_j\eta_i\eta_j = \frac{1}{2}[NR_{\sigma\eta}^2 - 1]$$

where $R_{\sigma\eta} = (\sum_i \sigma_i\eta_i)/N$, called the overlap between σ and η. Note that this quantity always lies between -1 and $+1$; it equals one if and only if the two configurations σ and η are the same. Similarly, keeping the independence of the families (J'_{ij}) and (J''_{ij}), we get

$$E(V_\sigma V_\eta) = \frac{1}{2}[N_1(R_{\sigma\eta}^*)^2 - 1] + \frac{1}{2}[N_2(R_{\sigma\eta}^{**})^2 - 1]$$

where $R_{\sigma\eta}^* = (\sum_{i\le N_1}\sigma_i\eta_i)/N_1$ and $R_{\sigma\eta}^{**} = (\sum_{N_1<i\le N}\sigma_i\eta_i)/N_2$.

Since $F(x) = \log\sum_\tau e^{-\beta x_\tau}$ we have, for its derivatives

$$F_\sigma(x) = -\beta e^{-\beta x_\sigma}/(\Sigma_\tau e^{-\beta x_\tau}) = -\beta G(\sigma, x)$$

where $G(\sigma, x)$ is the Gibbs measure of σ corresponding to the vector x, namely, $e^{-\beta x_\sigma}/Z$ where $Z = \sum_\tau e^{-\beta x_\tau}$.

$$F_{\sigma\eta}(x) \;=\; \frac{-\beta^2 e^{-\beta x_\sigma}e^{-\beta x_\eta}}{(\sum_\tau e^{-\beta x_\tau})^2} \;=\; -\beta^2 G(\sigma,x)G(\eta,x) \quad \text{for} \quad \sigma\neq\eta,$$

$$F_{\sigma\sigma}(x) \;=\; \frac{-\beta^2 e^{-2\beta x_\sigma}}{(\sum_\tau e^{-\beta x_\tau})^2} \;+\; \frac{\beta^2 e^{-\beta x_\sigma}}{\sum_\tau e^{-\beta x_\tau}} \;=\; -\beta^2 G(\sigma,x)G(\sigma,x) \;+\; \beta^2 G(\sigma,x)$$

substituting these value in the expression for φ' above, we get $\frac{4}{\beta^2}\,\varphi'$ equals

$$\sum_{\sigma\eta}[NR_{\sigma\eta}^2 - N_1R_{\sigma\eta}^{*2} - N_2R_{\sigma\eta}^{**2} + 1]E[-G(\sigma,W_t)G(\eta,W_t)] \;+\; \sum_\sigma E[G(\sigma,W_t)]$$

$$= -\frac{1}{N}E\left\{\sum_{\sigma\eta}G(\sigma,W_t)G(\eta,W_t)\left[R_{\sigma\eta}^2 - \frac{N_1}{N}R_{\sigma\eta}^{*2} - \frac{N_2}{N}R_{\sigma\eta}^{**2}\right]\right\}$$

Note that

$$R_{\sigma\eta} = \frac{N_1}{N}R_{\sigma\eta}^* + \frac{N_2}{N}R_{\sigma\eta}^{**}$$

The funcion $x\mapsto x^2$ being convex function, we have

$$R_{\sigma\eta}^2 - \frac{N_1}{N}R_{\sigma\eta}^{*2} - \frac{N_2}{N}R_{\sigma\eta}^{**2} \le 0$$

The Gs being positive we conclude from the above expression concerning φ' that $\varphi' \geq 0$. Thus $\varphi(0) \leq \varphi(1)$. In other words $E[\log Z_2] \leq E[\log Z_1]$ where the Z_i are the partition functions corresponding to the Hamiltonians H_i. It is not difficult to see that

$$E[\log Z_2] = E[\log Z_{N_1}] + E[\log Z_{N_2}] \quad \text{and} \qquad E[\log Z_1] = E[\log Z_N].$$

This completes the proof the super-additivity of the sequence of numbers $E[\log Z_n]$.

Proof of Theorem 15:

Let $H_{1N}(\sigma) = \sum\limits_{1\leq i<j\leq N} J_{ij}\sigma_i\sigma_j$ where (J_{ij}) are standard normal. Let $H_{2N}(\sigma) = \sum\limits_{1\leq i<j\leq N} J'_{ij}\sigma_i\sigma_j$ where (J'_{ij}) are i.i.d. mean zero, variance one, finite third moment variables. Set $Z_{1N} = \sum\limits_{\sigma} e^{-\beta H_{1N}(\sigma)/\sqrt{N}}$ and $Z_{2N} = \sum\limits_{\sigma} e^{-\beta H_{2N}(\sigma)/\sqrt{N}}$. We need to show that $E[\frac{1}{N}(\log Z_{1N} - \log Z_{2N})] \to 0$.

Some notational issues are to be addressed first. We set $M = N(N-1)/2$. A vector $x \in R^M$ is indexed by (ij) with $1 \leq i < j \leq N$. The indices (ij) are also 'arranged' in some order $1 \leq l \leq M$ when needed. For $\sigma \in \{-1,+1\}^N$, we let σ denote also the vector of R^M whose (ij)-th coordinate is $\sigma_i\sigma_j$. There will be no confusion with this dual notation for σ, it will be obvious from the context as to what is meant. Let J be the M-dimensional random vector whose (ij)-th coordinate is J_{ij}. Similarly J'. With this notation, $H_{1N} = J\cdot\sigma$ and $H_{2N} = J'\cdot\sigma$ where the dot denotes the inner product. For $0 \leq l \leq M$ let X^l be the M-dimensional random vector which has J_{ij} variables upto and including the coordinate l and J'_{ij} variables after the l-th coordinate. Thus X^M is just J and X^0 is J'. Let $F: R^M \to R$ be defined by $F(x) = \frac{1}{N}\log\sum\limits_{\sigma} e^{-\beta x\cdot\sigma/\sqrt{N}}$ where the dot in the exponent is the inner product. With all this notation, what we need to show amounts to

$$E[F(X^M) - F(X^0)] = \sum_{1}^{M} E[F(X^l) - F(X^{l-1})] \to 0$$

Set Y^l to be the M-dimensional random vector which is X^l in all its coordinates except the l-th coordinate which is set as zero. Thus replacing the l-th coordinate of Y^l by the l-th coordinate of J we get X^l, while replacing the l-th coorinate of Y^l by the l-th coordinate of J' we get X^{l-1}. Since our function F is smooth, we have

$$F(X^l) - F(X^{l-1}) = [F(X^l) - F(Y^l)] \; - \; [F(X^{l-1}) - F(Y^l)]$$

which equals, by Taylor expansion

$$F_l(Y^l)[J_l - J_l'] \ + \ \frac{1}{2}F_{ll}(Y^l)[J_l^2 - J_l'^2] \ + \ \frac{1}{6}[F_{lll}(?)J_l^3 - F_{lll}(??)J_l'^3]$$

Here the suffixes for F denote the partial derivative w.r.t. that coordinate. The question marks denote that F_{lll} is evaluated at an appropriate point in R^M, the actual point is not relevant (enough to note that we have random variables). Since Y^l, J_l, J_l' are independent and also that J_l and J_l' have zero means and unit second moments we get, on taking expectations of the above equation,

$$E[F(X^l) - F(X^{l-1})] \ = \ \frac{1}{6}E[F_{lll}(?)J_l^3 - F_{lll}(??)J_l'^3]$$

Now let us calculate the required derivatives. Let the coordinate l be (ij). Keep in mind that the partition function corresponding to x is $Z(x) = \sum\limits_{\sigma} e^{-\beta x\cdot\sigma/\sqrt{N}}$, the Gibbs measure corresponding to x on the σ-space is $G(\sigma, x) = \frac{1}{Z(x)}e^{-\beta x\cdot\sigma/\sqrt{N}}$ and for any function f on the σ-space its expectation w.r.t. this probability is denoted by $\langle f\rangle_x$.

$$F(x) = \frac{1}{N}\log\sum_{\sigma} e^{-\beta x\cdot\sigma/\sqrt{N}}$$

$$F_l(x) = \frac{-\beta}{N^{3/2}}\frac{1}{Z(x)}\sum_{\sigma} e^{-\beta x.\sigma/\sqrt{N}}\ \sigma_i\sigma_j \ = \ \frac{-\beta}{N^{3/2}}\langle\sigma_i\sigma_j\rangle_x$$

$$F_{ll}(x) = \frac{\beta^2}{N^2}\left[\frac{(\sum\limits_{\sigma} e^{-\beta x.\sigma/\sqrt{N}}\ \sigma_i\sigma_j)^2}{Z^2(x)} - \frac{\sum\limits_{\sigma} e^{-\beta x.\sigma/\sqrt{N}}\ \sigma_i^2\sigma_j^2}{Z(x)}\right]$$

$$= \frac{\beta^2}{N^2}[(\langle\sigma_i\sigma_j\rangle_x)^2 - 1]$$

$$F_{lll}(x) = \frac{\beta^2}{N^{5/2}}\left[\frac{2\sum\limits_{\sigma} e^{-\beta x\cdot\sigma/\sqrt{N}}\sigma_i\sigma_j.\sum\limits_{\sigma} e^{-\beta x\cdot\sigma/\sqrt{N}}}{Z^2(x)} - \frac{2(\sum\limits_{\sigma} e^{-\beta x\cdot\sigma/\sqrt{N}}\sigma_i\sigma_j)^3}{Z^3(x)}\right]$$

$$= \frac{2\beta^2}{N^{5/2}}\left[\langle\sigma_i\sigma_j\rangle_x^3 - \langle\sigma_i\sigma_j\rangle_x\right] \leq C\,N^{-5/2}$$

As a consequence $E[F(X^l) - F(X^{l-1})] \ \leq \ C\,N^{-5/2}$ and finally

$$E[F(X^M) - F(X^0)] \ \leq \ MC\,N^{-5/2}$$

Since $M = N(N-1)/2$ this last expression converges to zero as $n \to \infty$, completing the proof.

Remarks:

1. The overlap that appeared in the proof of theorem 14 plays a prominent role in understanding SK model.

2. See Talagrand [59] for SK model. Guerra and Toninelli [34] proved the existence of the limiting energy for the SK model. The identification of the limit with a conjecture of Parisi was achieved by Guerra [33] and Talagrand [58].

3. The universality was noted in Carmona etal [19] and Chatterjee [20], see also Talagrand [56]. The proof is an imitation of the Lindeberg proof of the Central Limit Theorem as given, for example, in Billingsley [7].

4. For $\beta < 1$, the limiting energy is $\log 2 + \frac{1}{4}\beta^2$. See Talagrand [59].

5. We have not included the external field term, but that can also be included in proving the existence of the limit.

4. Gibbs Distribution:

Let us now turn our attention to another problem, namely, the fate of the Gibbs distributions. For each N, we have a random probability on $\{-1,+1\}^N$, namely the Gibbs distribution, which is given by $G(\sigma) \propto e^{-\beta H_N(\sigma)}$. More precisely, it is the probability which puts mass $e^{-\beta H_N(\sigma)}/Z_N$ at the point σ. This is a random probability, because H_N are random. The question is to find out if there is any limiting object for these. Of course, one has to formulate carefully, because the space on which these probabilities live changes with N.

Theorem 16 (REM High temperature Gibbs measures):

Consider the Gaussian REM with $\beta < \sqrt{2\log 2}$. Then almost surely the Gibbs measures converge to the uniform distribution on the space$\{-1,+1\}^\infty$ in the following sense. Fix any k. For every $N \geq k$ and ω, the Gibbs distribution on $\{-1,+1\}^N$ be denoted by $G_N(\omega, .)$. Let its marginal on the first k coordinate space be denoted by $G_N^k(\omega, .)$. Then for almost every ω, as $n \to \infty$, these converge to the uniform distribution on $\{-1,+1\}^k$. ∎

To discuss the fate of Gibbs measures at low temperatures, we need some notation. For $0 < m < 1$, let P_m be the Poisson Point Process on $(0,\infty)$ with intensity $x^{-m-1}dx$. This means the following. Say that a subset $D \subset (0,\infty)$ is locally finite if for any two numbers a and b with $0 < a < b < \infty$, the set $D \cap [a,b]$ is a finite set. Clearly, a locally finite set is countable. Let Ω be the collection of all locally finite sets. For a Borel set $B \subset (0,\infty)$ let N_B be the 'number map' defined on Ω by $D \mapsto |B \cap D|$, where $|A|$ is the number of points in the set A. This map takes non-negative integer values, including possibly infinity. Let $\mathcal{F}$ be the smallest σ-field on Ω making all these maps N_B — as B

varies over Borel sets — measurable. There is a unique probability P_m on $\mathcal{F}$ such that for each Borel $B \subset (0,\infty)$ the P_m distribution of N_B is Poisson with parameter $\nu(B) = \int\limits_B x^{-m-1}dx$ and moreover for disjoint Borel sets $B_i \subset (0,\infty)$ the random variables $N_{B_i}; 1 \leq i \leq k$ are independent. Poisson distribution with parameter ∞ means mass one at infinity and parameter 0 means mass one at zero.

Since for each $a > 0$ we have $\int\limits_a^\infty x^{-m-1}dx < \infty$, we conclude that almost every D (under P_m) has only finitely many points larger than a. Thus we can arrange D as a decreasing sequence. Since $\int\limits_0^1 x^{-m-1}dx = \infty$ we conclude that D has infinitely many points and hence, when arranged as decreasing sequence, it converges to zero. Since $\int\limits_0^1 x.x^{-m-1}dx < \infty$ the sum of points of D which are smaller than one is finite. But since there are only finitely many points of D larger than one, we deduce that the sequence D is summable. Let $S \subset [0,1]^\infty$ be the set of all seqeunces $(x_1, x_2, \cdots)$ which are decreasing with sum at most one. Clearly S is a compact set (usual topology, as a subset of the product space). We define a map from Ω to S as $D \mapsto (\frac{d_1}{d}, \frac{d_2}{d}, \cdots)$ where (d_i) is decreasing enumeration of D and d is sum of points of D. Let Λ_m be the probability induced on S by P_m via this map. This is called the Poisson-Dirichlet distribution with parameter m.

Let us now consider REM with $\beta > \sqrt{2\log 2}$. Put $m = \beta/\sqrt{2\log 2}$, so that $0 < m < 1$. Arrange the 2^N numbers $H_N(\sigma)/Z_N$ in decreasing order and continue with zeros to get a random point of S. Let its distribution be denoted by μ_N. Thus μ_N is the distribution of the Gibbs measures.

Theorem 17 (REM Low temperature Gibbs measures):

Consider the Gaussian REM with $\beta > \sqrt{2\log 2}$. Denote $m = \beta/\sqrt{2\log 2}$ and μ_N be the distribution of Gibbs measures. Then $\mu_N \to \Lambda_m$ weakly on S. ∎

Proof of Theorem 16:

If $\beta^2 < \log 2$ then the Gibbs distributions converge almost surely to the uniform probability on $\{-1,+1\}^\infty$ hereafter denoted 2^ω. Let us put $Z_n = \sum\limits_\sigma e^{\beta\sqrt{N}\xi_\sigma}$ where ξ_σ are independent standard normal. Then

$$\alpha_n = EZ_N = 2^N e^{N\beta^2/2}$$

$$var(Z_N) = E\sum_{\sigma,\eta} e^{\beta\sqrt{N}\xi_\sigma} e^{\beta\sqrt{N}\xi_\eta} - \sum_{\sigma,\eta} E[e^{\beta\sqrt{N}\xi_\sigma}]E[e^{\beta\sqrt{N}\xi_\eta}].$$

Cancelling the $\sigma \neq \eta$ terms and ignoring the remaining positive terms from the second sum we get

$$var(Z_N) \leq \sum_\sigma E[e^{2\beta\sqrt{N}\xi_\sigma}] = 2^N e^{2N\beta^2}$$

As a consequence

$$P\left(|\frac{Z_N}{\alpha_N} - 1| > \epsilon\right) \leq \frac{1}{\epsilon^2} var(Z_N/\alpha_N) \leq \frac{1}{\epsilon^2} e^{-N(\log 2 - \beta^2)}$$

Since the last terms form a convergent series, Borel-Cantelli shows that $\frac{Z_N}{\alpha_N} \to 1$ almost surely. Fix $\tau \in 2^k$. Then for $N > k$ the projection of the Gibbs measure G_N on 2^k puts mass $G_N^k(\tau)$ at τ which is described as follows. For $\eta \in 2^{N-k}$, let us denote the concatenation of τ followed by η as $\langle \tau\eta \rangle \in 2^N$. Put $\widetilde{Z}_N = \sum_\eta e^{\beta\sqrt{N}\xi_{\langle\tau\eta\rangle}}$. Then $G_N^k(\tau) = \widetilde{Z}_N/Z_N$. One can repeat the above argument to show that $E[\widetilde{Z}_N] = 2^{-k}\alpha_N$ and the ratio $(\widetilde{Z}_N)/(2^{-k}\alpha_N)$ converges to one almost surely. Combining this result with the earlier one for Z_N one gets that, $G_N^k(\tau) = \widetilde{Z}_N/Z_N \to 2^{-k}$ almost surely to complete the proof.

Now let us assume that $\log 2 \leq \beta^2 < 2\log 2$. The proof is similar to the above, just that on needs to truncate the Hamiltonians. Recall that $H_N(\sigma) = \sqrt{N}\xi_\sigma$ where ξs are independent standard normal. We shall fix a δ such that $\sqrt{2\log 2} < \delta < 2\beta$. Put $Z'_N = \sum_\sigma e^{\beta H_N(\sigma)} I_{H_N(\sigma) \leq N\delta}$. Then

$$P(Z_N \neq Z'_N) \leq 2^N P(H_N > N\delta) \leq 2^N \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{N}\delta} e^{-\delta^2 N/2} \leq \frac{1}{\sqrt{N}\delta} e^{-N[\frac{\delta^2}{2} - \log 2]}$$

which is summable, so that almost surely eventually, Z_N and Z'_N are equal.

Further

$$E(Z'_N) = 2^N \int_{-\infty}^{N\delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2N}(x-\beta N)^2} dx \; e^{N\beta^2/2} > \frac{1}{2} 2^N e^{N\beta^2/2}$$

where we used $N\delta > N\beta$. Regarding variance, note that $var(Z'_N)$ equals

$$E[\sum_{\sigma,\eta} e^{\beta H_N(\sigma)} \;\; I_{(H_N(\sigma) \leq N\delta)} e^{\beta H_N(\eta)} I_{(H_N(\eta) \leq N\delta)}]$$

$$- \sum_{\sigma,\eta} E[e^{\beta H_N(\sigma)} I_{(H_N(\sigma)\leq N\delta)}] \;\; E[e^{\beta H_N(\eta)} I_{(H_N(\eta)\leq N\delta)}].$$

Cancel the $\sigma \neq \eta$ terms and ignore the others from the second expression, to get $var(Z'_N)$ is at most

$$2^N \int_{-\infty}^{N\delta} \frac{1}{\sqrt{2\pi}} e^{-(x-2\beta N)^2/(2N)} dx \; e^{2N\beta^2} = 2^N e^{2N\beta^2} \int_{N(2\beta-\delta)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

which after usual estimate and simplification, yields

$$var(Z'_N) \leq \frac{1}{N(2\beta-\delta)} 2^N e^{-N(\frac{\delta^2}{2}-2\beta\delta)}$$

This leads to

$$P[|Z'_N - E(Z'_N)| \;\geq\; \epsilon E(Z'_N)] \;\leq\; \frac{4}{N\epsilon^2(2\beta-\delta)}\; e^{-N[\log 2\,+\,\frac{\delta^2}{2}\,-\,2\beta\delta\,+\,\beta^2]}.$$

If $\log 2 + \frac{\delta^2}{2} - 2\beta\delta + \beta^2 > 0$, then the last expression will be summable over N. We now show that such a choice of δ is possible. Since $(2x-y)^2 > 2x^2 - y^2$, we have $(2\beta - \sqrt{2\log 2})^2 > 2\beta^2 - 2\log 2$ so that $2\beta - \sqrt{2\log 2} > \sqrt{2(\beta^2 - \log 2)}$. Thus $2\beta - \sqrt{2(\beta^2-\log 2)} > \sqrt{2\log 2}$. Choose δ between these two quantities.

All this shows, via Borel-Cantelli, that almost surely eventually

$$(1-\epsilon)E(Z'_N) \;\leq\; Z'_N \;\leq\; (1+\epsilon)E(Z'_N)$$

or

$$(1-\epsilon)E[e^{\beta H_N} I_{(H_N \leq N\delta)}] \;\leq\; \frac{Z_N}{2^N} \;\leq\; (1+\epsilon)E[e^{\beta H_N} I_{(H_N \leq N\delta)}].$$

Similarly, fixing a $\tau \in 2^k$ and defining $\widetilde{Z}_N$ as in the earlier para, we see

$$(1-\epsilon)E[e^{\beta H_N} I_{(H_N \leq N\delta)}] \;\leq\; \frac{\widetilde{Z}_N}{2^{N-k}} \;\leq\; (1+\epsilon)E[e^{\beta H_N} I_{(H_N \leq N\delta)}].$$

Hence almost surely eventually, $\widetilde{Z}_N/Z_N$ lies between $\frac{1-\epsilon}{1+\epsilon}2^{-k}$ and $\frac{1+\epsilon}{1-\epsilon}2^{-k}$, showing that $G_N^k(\tau) \to 2^{-k}$.

Proof of Theorem 17:

The proof needs a series of steps.

(**1°**). Suppose that ξ_N is a centered Gaussian variable with variance N. Put

$$\alpha_N^2 = \frac{2}{N}\log\frac{2^N}{\sqrt{N}} = 2\log 2 - \frac{\log N}{N}.$$

Put $\eta_N = \xi_N + N\alpha_N$. Let $-\infty < a < b \leq \infty$Then

$$2^N P(a < -\eta_N < b) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-t(\alpha_N + \frac{t}{2N})}\,dt \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-t\sqrt{2\log 2}}\,dt.$$

This is because,

$$
\begin{aligned}
P(a < -\eta_N < b) &= \int\limits_{a+N\alpha_N}^{b+N\alpha_N} \frac{1}{\sqrt{2\pi N}} e^{-t^2/(2N)} dt \\
&= \int\limits_a^b \frac{1}{\sqrt{2\pi N}} e^{-(t+N\alpha_N)^2/(2N)} dt \\
&= \int\limits_a^b \frac{1}{\sqrt{2\pi N}} e^{-\frac{t^2}{2N} - t\alpha_N} \; e^{-N\alpha_N^2/2} \\
&= 2^{-N}\sqrt{N} \frac{1}{\sqrt{2\pi N}} \int\limits_a^b e^{-\frac{t^2}{2N} - t\alpha_N} dt \\
&= 2^{-N} \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-t(\alpha_N + \frac{t}{2N})} \; dt.
\end{aligned}
$$

Since

$$
\alpha_N + \frac{t}{2N} = \sqrt{2\log 2 - \frac{\log N}{N}} + \frac{t}{2N} \longrightarrow \sqrt{2\log 2}
$$

as $N \to \infty$, which proves the claim. This is true even if $b = \infty$, we only need $a > -\infty$.

(**2°**). Consider the REM Hamiltonians $H_N(\sigma)$ which are i.i.d. centered Gaussian with variance N. Set α_N as above and $H'_N(\sigma) = H_N(\sigma) + N\alpha_N$. Note that the Gibbs measures are same whether we consider H_N or H'_N. Consider disjoint intervals $I_j = (a_j, b_j)$ and non-negative integers n_j for $1 \le j \le k$. then

$$
P[-H'_N(\sigma) \in (a_j, b_j) \text{ for exactly } n_j \text{ values of } \sigma,\, 1 \le j \le k] \longrightarrow \prod_{j=1}^{k} \frac{e^{-\lambda_j}\lambda_j^{n_j}}{n_j!}
$$

where $\lambda_j = \int\limits_{a_j}^{b_j} \frac{1}{\sqrt{2\pi}} e^{-t\sqrt{2\log 2}}\, dt$.

This follows from the above and usual multinomial convergence to Poisson. This essentially says that the point process $\{-H'_N(\sigma)\}$ converges to the Poisson Point process with intensity $\frac{1}{\sqrt{2\pi}} e^{-t\sqrt{2\log 2}} dt$. Of course, we have not defined convergence of point processes. From what was said above it also follows that if $b > 0$ and $k \ge 1$, given that there are k points of the point process $\{-H'_N(\sigma)\}$ in (b, ∞) the conditional distribution of their positions $(X_1^N, \cdots X_k^N)$ converges in law to the product of k i.i.d. variables with density proportional to $\frac{1}{\sqrt{2\pi}} e^{-t\sqrt{2\log 2}} dt\; 1_{(b,\infty)}(t)$.

(**3°**). There is a number $c > 0$ such that the following happens. Set the Gibbs factor $g_N(\sigma) = ce^{-\beta H'_N(\sigma)}$ and $m = \sqrt{2\log 2}/\beta$. Note that the Gibbs measures are same whether we consider the Gibbs factors $e^{-\beta H_N(\sigma)}$ or $e^{-\beta H'_N(\sigma)}$ or $g_N(\sigma)$.

Consider disjoint intervals $I_j = (a_j, b_j) \subset (0, \infty)$ and non-negative integers n_j for $1 \le j \le k$. then

$$P(g_N(\sigma) \in (a_j, b_j) \text{ for exactly } n_j \text{ values of } \sigma,\ 1 \le j \le k) \longrightarrow \prod_{j=1}^{k} \frac{e^{-\lambda_j} \lambda_j^{n_j}}{n_j!}$$

where $\lambda_j = \int\limits_{a_j}^{b_j} t^{-m-1} dt$.

This essentially says that the point process $\{g_N(\sigma)\}$ converges to the Poisson Point process with intensity $t^{-m-1}dt$.

This follows from the above calculations because,

$$P(a < g_N < b) = P(\frac{1}{\beta} \log \frac{a}{c} < -H'_N < \frac{1}{\beta} \log \frac{b}{c}) = \int\limits_a^b \frac{1}{\sqrt{2\pi}} \frac{1}{\beta} c^m t^{-m-1} dt.$$

and choose c so that $c^m = \beta\sqrt{2\pi}$.

t is worth noting that if we have a Poisson point process on R with intensity $\frac{1}{\sqrt{2\pi}} e^{-t\sqrt{2\log 2}} dt$ then the transformation $x \mapsto ce^{\beta x}$ gives us a poisson point process on $(0, \infty)$ with intensity $t^{-m-1}dt$ where $m = \sqrt{2\log 2}/\beta$.

($\mathbf{4^o}$). For any real number b, let the truncated Gibbs factors $g_N^b(\sigma)$ be defined as $e^{-\beta H'_N(\sigma)}$ if $-H'_N(\sigma) \ge b$ and zero otherwise. The truncated partition function Z_N^b is the sum of all the truncated Gibbs factors. Arrange the numbers $g_N^b(\sigma)/Z_N^b$ in decreasing order to get a random point of S and let μ_N^b be its distribution.

Consider a Poisson point process with intensity $\frac{1}{\sqrt{2\pi}} e^{-t\sqrt{2\log 2}} dt$ on R. For a Poisson point $D = (d_n)$, let the truncated point be defined as D^b namely, all points of D which are $\ge b$. Let z^b be the sum of all numbers $e^{\beta d}$ over $d \in D^b$ and arrange the numbers $e^{\beta d}/z^b$ (for $d \in D^b$) in decreasing order to get a random point of S and let its distribution be λ^b.

claim: $\mu_N^b \Rightarrow \lambda^b$ on S.

To see this, take any continuous function on S. If, for $k \ge 0$, we denote by S_k the set of points of S with the exactly the first k coordinates non-zero, then (2^o) tells us that

$$\mu_N^b(S_k) \longrightarrow \lambda^b(S_k) \quad k \ge 0$$

$$\frac{1}{\mu_N^b(S_k)} \int\limits_{S_k} f d\mu_N^b \longrightarrow \frac{1}{\lambda^b(S_k)} \int\limits_{S_k} f d\lambda^b \quad k \ge 1,$$

From these we deduce that

$$\int\limits_S f d\mu_N^b = \sum_k \int\limits_{S_k} f d\mu_N^b \longrightarrow \sum_k \int\limits_{S_k} f d\lambda^b = \int\limits_S f d\lambda^b.$$

($\mathbf{5^o}$). Recall that for any real number b, the truncated Gibbs factors $g_N^b(\sigma)$ were defined as $e^{-\beta H_N'(\sigma)}$ if $-H_N'(\sigma) \geq b$ and zero otherwise. The truncated partition function Z_N^b is the sum of all the truncated Gibbs factors.

Fix $\epsilon > 0$. We can get a number $b \in R$ and non-negative integer N_0 such that for all $N \geq N_0$, $P[Z_N^b \geq (1-\epsilon)Z_N] \geq 1-\epsilon$.

This is seen as follows. First fix x such that $\exp\{-\int\limits_x^\infty e^{-t\sqrt{2\log 2}}\} < \epsilon/2$. Since $P(Z_N \leq e^{\beta x}) \leq P[|\sigma : -H_N'(\sigma) \geq x| = 0]$ and this later quantity converges to $\exp\{-\int\limits_x^\infty e^{-t\sqrt{2\log 2}}\}$ we conclude that for all large N, $P(Z_N \leq e^{\beta x}) < \epsilon/2$. Denoting by f the density function of $-H_N'$ and F its distribution function, we have for any b,

$$\begin{aligned} E[Z_N - Z_N^b] &= 2^N \int\limits_{-\infty}^{b} e^{\beta x} f(x)dx \\ &= 2^N[F(b)e^{\beta b} - \int\limits_{-\infty}^{b} \beta e^{\beta x} F(x)dx] \\ &= \beta 2^N \int\limits_{-\infty}^{b} [F(b) - F(x)]e^{\beta x}dx \\ &\leq \beta 2^N \int\limits_{-\infty}^{b} [1 - F(x)]e^{\beta x} f(x)dx \\ &\leq \beta \tfrac{1}{\sqrt{2\pi}\alpha_N} \int\limits_{-\infty}^{b} e^{x(\beta - \alpha_N)}dx \end{aligned}$$

the last inequality follows from

$$2^N[1-F(x)] = \frac{1}{\sqrt{2\pi}} \int\limits_x^\infty e^{-t(\alpha_N + \frac{t}{2N})}\, dt \leq \frac{1}{\sqrt{2\pi}} \int\limits_x^\infty e^{-t\alpha_N}\, dt = \frac{1}{\sqrt{2\pi}\alpha_N} e^{-\alpha_N x}$$

Since $\beta - \alpha_N > \beta - \sqrt{2\log 2} > 0$ we can find a number b (negative) such that $E[Z_N - Z_N^b] < \epsilon^2 e^{\beta x}/2$ for all large N. Then $P(Z_N - Z_N^b > \epsilon e^{\beta x}) \leq \epsilon/2$ by markov inequality. Thus we have a number b and N_0 such that for all $N \geq N_0$ the following hold: $P(Z_N \leq e^{\beta x}) < \epsilon/2$ and also $P(Z_N - Z_N^b \geq \epsilon e^{\beta x}) < \epsilon/2$. These two give $P(Z_N^b \geq (1-\epsilon)Z_N) \geq 1-\epsilon$ as desired.

($\mathbf{6^o}$). Let z^b be as in ($\mathbf{4^o}$) and z be the untruncated sum of all the factors $e^{\beta d}$ for $d \in D$. Given $\epsilon > 0$, there is b such that $P(z^b \geq (1-\epsilon)z) \geq 1-\epsilon$. This is clear since $z^b \uparrow z$ as $b \downarrow -\infty$.

(**7°**).

$$\sum_\sigma |g_N(\sigma) - g_N^b(\sigma)| = 2\frac{Z_N - Z_N^b}{Z_N}.$$

To see this, break the sum into two parts, one over $-H_N'(\sigma) \geq b$ and the second over the remaining part. The first sum is

$$\sum^* \frac{e^{-\beta H_N'(\sigma)}}{Z_N} - \frac{e^{-\beta H_N'(\sigma)}}{Z_N^b} = \sum^* e^{-\beta H_N'(\sigma)} \left| \frac{1}{Z_N} - \frac{1}{Z_N^b} \right| = \frac{Z_N - Z_N^b}{Z_N}.$$

The remaining sum is treated similarly.

(**8°**). Finally, to prove the theorem let f be a continuous function on S. Shall show that $|\int f d\mu_N - \int f d\lambda| \to 0$. First fix a $\delta > 0$ such that $x, y \in S$ and $\sum |x_n - y_n| < \delta/2$ implies $|f(x) - f(y)| < \epsilon$. Fix b satisfying (5^o) and (7^o).

Firstly, $|\int f d\mu_N - \int f d\mu_N^b| = E|f(G) - f(G^b)|$ with the obvious notation. This expectation over the set $|Z - Z^b| < \delta/2$ is at most ϵ by (7^o) and choice of δ and the expectation over the other part is at most $2||f||\delta$ by choice of b. This is so for all $n \geq N_0$

Secondly, consider $|\int f d\lambda - \int f d\lambda^b|$ with the obvious notation. This expectation over the set $|z - z^b| < \delta/2$ is at most ϵ by (7^o) and choice of δ and the expectation over the other part is at most $2||f||\delta$ by choice of b.

Thirdly, $|\int f d\mu_N^b - \int f d\lambda^b|$ goes to zero as $N \to \infty$ by (4^o).

These three observations complete the proof.

Remarks:

1. For the asymptotics of Gibbs distributions, see Bovier [17], Jana [36], Talagrand [59] and Galves etal [31].

2. Concerning Theorem 16, since for fixed k we are dealing with probabilities on a finite set, we could also say convergence holds in the total variation norm. There are estimates for the total variation distance between the measures G_N^k and uniform distribution. See Talagrand [59].

3. For the SK model also there are results concerning th convergence of Gibbs measures at high temperature, that is, for small values of β. However for large values of β there seem to be none.

4. For GREM also one can find the limiting object of the Gibbs distributions. Ruelle [49] postulates the limiting object. For proof that these are limiting objects see Galves etal [31], Bovier etal [16]. The limiting object turns out to be the *Ruelle Cascade*. To describe this fix numbers $0 < x_1 < x_2 < \cdots < x_k < 1$. Let N^k denote the set of all k-tuples $(n_1, \cdots, n_k)$ of strictly positive integers. We consider lots (and lots) of independent Poisson processes, all on

$(0,\ \infty)$, in what follows. First take one process ξ_i^1 with intensity $t^{-x_1-1}dt$. Its points arranged in decreasing order are denoted by $(\xi_{n_1}^1)$. Take for each integer n, one process ξ_n^2 with intensity $t^{-x_2-1}dt$ and denote its points in decreasing order by $\xi_{n,m}^2$ and so on. Finally for each $(k-1)$-tuple $(n_1, \cdots n_{k-1})$ a process $\xi_{n_1,\cdots n_{k-1}}^k$ with intensity $t^{-x_k-1}dt$ and its points denoted by $\xi_{n_1,\cdots n_{k-1},n_k}^k$. Let X be the point process $(\xi_{n_1}^1, \xi_{n_1,n_2}^2 \cdots, \xi_{n_1,\cdots,n_k}^k)$. This is called a Ruelle Cascade. If we have a Ruelle cascade, we first define for any $(n_1, n_2, \cdots, n_k)$ the random point $\eta_{(n_1,n_2,\cdots,n_k)} = \prod_i \xi_{n_1,\cdots,n_i}^i$. These numbers are summable and the probability proportional to these numbers gives us a random probability on N^k. Consider the distribution of this random probability. This is denoted by $\lambda(x_1, \cdots, x_k)$. Actually we can define an η-cascade also going backwards as follows. $\eta_{(n_1,n_2,\cdots,n_k)}$ was defined already. Put $\eta_{(n_1,n_2,\cdots,n_{k-1})}$ as sum of all the $\eta_{(n_1,n_2,\cdots,n_k)}$ over n_k and so on. Then we get an η-cascade.

Properly formulated, $\lambda(x_1, \cdots, x_k)$ turns out to be the limiting object for the law of the Gibbs distributions. For instance if the set-up is as in Theorem 8(i), with the change in notation that we are considering a k-level grem so that we have k numbers β_i. Then we should consider $\beta > \beta_k$ and the parameters of the limiting object λ turn out to be $x_i = \beta_i/\beta$ which are increasing in i. Of course if the GREM is not in reduced form, more care is needed.

5. What next:

We have only discussed the elementary part of the theory, in a sense, the soft part. More interesting material comes next. There are several directions in which one can proceed. But, one should first read Talagrand's monograph.

1. As mentioned earlier, the overlap is very important and its asymptotics are thoroughly discussed in Talagrand [59] including several concentration inequalities. The recent works of Aizenmann etal [1] and Panchenko etal [65] should be well understood. See also the papers of Talagrand and Bovier in [8].

2. Koukiou [40] provides a random covering interpretation of the phase transition in REM and GREM, using an analysis of Shepp[62].

Suppose that μ is a measure on $(0, \infty)$ which is finite for compact sets $[a, b]$ with $0 < a < b < \infty$. Let λ be the lebesgue measure on the real line. Consider a Poisson point process on $R \times R^+$ with intensity $\lambda \times \mu$. Given a Poisson point $p = \{(x_i, l_i)\}$ let $C_p = \cup(x_i, x_i + l_i)$. Say that the real line is covered by the point p if $C_p = R$. A theorem of Shepp is that either almost every poisson point covers the real line or almost no point covers. The first case occurs if

$\int_0^1 \left[\exp \int_x^\infty (y-x)\mu(dy)\right] dx = \infty$ and the second case occurs if this integral is finite.

Say that μ gives a low frequency covering if μ restricted to $[1,\infty)$ covers R. That is, the above process with μ replaced by $\mu'(B) = \mu(B \cap [1,\infty))$ covers R. Say that μ gives a high frequency covering of R if μ restricted to $(0,1)$ covers. Since the energy levels of REM, properly normalized, converge to a Poisson point process with intensity $mt^{-m-1}dt$, one can postulate that REM is a Poisson point process on $(0,\infty)$ with intensity $mt^{-m-1}dt$ where $m > 0$. it is not difficult to see that whether the Poisson process with intensity $mt^{-m-1}dt$ gives a low frequency covering or high frequency covering depends on $m > 1$ or $m < 1$. According to the identification of the parameters m corresponds to $\sqrt{2\log 2}/\beta$. Thus $m > 1$ and $m < 1$ correspond to β below or above the critical value.

For GREM also there is a similar interpretation appears in Koukiou etal [41].

3. Liggett etal [42] consider the following interesting problem. Suppose we fix an N and consider the N-particle system with configuration space $2^N = \{-1,+1\}^N$ and a Hamiltonian H_N. Consider the Gibbs probability $G_N(\sigma)$ on 2^N. Can it be extended as an exchangeable probability on 2^∞? Recall that a probability on 2^∞ is exchangeable, if it is invariant under permutation of finitely many coordinates. References to earlier work can be found in this paper. Of course, they consider non-random Hamiltonians.

4. Comets and Neveu [63] and Comets [64] discuss SK model using stochastic calculus. The idea is to think of β as 'time parameter' and replace exponentials with exponential martingales and log by logarithmic martingales. Consider $(B_{ij}(t))_{t\geq 0}$ independent standard Brownian motions for $1 \leq i < j \leq N$. The Hamiltonian is $H_N(t,\sigma) = \frac{1}{\sqrt{N}} \sum_{1\leq i<j\leq N} B_{ij}(t)\sigma_i\sigma_j$. For each σ this is a martingale and the Gibbs factors are the exponential martingales $e_N(t,\sigma) = \exp\{H_N(t,\sigma) - \frac{1}{4}t(N-1)\}$ and partition function is the martingale $Z_N(t) = 2^{-N}\sum_\sigma e_N(t,\sigma)$. Finally, $\log Z_N(t)$ is the martingale $M_N(t) = \int_0^t \frac{1}{Z_N(s)} dZ_N(s)$. This treatment allows not only to find new results but also one can discuss convergence at the process level.

4. Our entire discussion did not concern any dynamics. It is possible to bring in dynamics and discuss a phenomenon called aging. See [3, 4] and references there in. Essentially one considers the nearest neighbour walk on $\{-1,+1\}^N$.

From σ you move to η that differs from σ in one coordinates or you stay at σ. The transition probabilities depend on the Hamiltonian.

5. In REM the limiting distribution of energies is Poisson. This appears to be a general phenomenon, leading to some universality conjectures. see [6, 14, 15] and references there in. One could also consider Gibbs distributions restricted to a window. More precisely, consider a set $S_N \subset \{-1,+1\}^N$ and look at only the energies corresponding to $\sigma \in S_N$. One could ask for the limit of this point processes, of course, after proper normalization.

6. A list of references is given below, which is by no means exhaustive, but will be useful to scan the literature further.

References:

1. Aizenman Michael, Sims Robert and Starr Shannon L [2006]: Meanfield spin glass models from the Cavity-ROST perspective. arXiv:math-ph/0607060 v1.
2. Ben Arous, G., Bogachev L. V. and Molchanov, S. A. [2005]. Limit theorems for sums of random exponentials. PTRF v.132, p.579–612.
3. Ben Arous Gerard, Bovier Anton and Gayrand Veronique [2001]: Glauber dynamics of the REM: 1. Meta Stable motion on extreme states; 2. Aging below the critical temperature. WIAS preprint 690.
4. Ben Arous Gerard and Cerny Jiri [2005]: Bouchaud's model exhibits two different aging regimes in dimension one. AAP v.15, p.1161-1192.
5. Ben Arous Gerard, Bovler Anton and Cerny Jiri [2007]: Universality of the REM for dynamics of mean-field spin glasses. arXiv: 0706.2135v1 [math.PR].
6. Ben Arous G, Gayrard V and Kuptsov A [2006]: A new REM conjecture. arXiv:math.PR/0612373v1.
7. Billingsley P [1968]: *Convergence of Probability measures.* John Wiley.
8. Bolthausen Erwin and Bovier Anton (editors)[2007]: *Spin Glasses.* Springer LNM 1900.
9. Bolthausen E and Kistler N [2006]: On a non-hierarchical version of the generalized random energy model. AAP v.16 p. 1-14.
10. Boutet de Monvel, L. Pastur and M. Shcherbina [1995]: On the Statistical Mechanics Approach in the Random Matrix Theory: Integrated density of States. JSP v.79, p.585-611.
11. Bovier Anton and Picco Pierre [1998]: Mathematical Aspects of spin Glasses and Neural Networks (edited volume) Progress in mathematics Birkhauser.
12. Bovier, A., Gayrard, V. and Picco, P. [1995]. Gibbs states of the Hopfield model with extensively many patterns. JSP. v.79, p.395–414.
13. Bovier Anton and Kurkova Irina [2002]: rigorous results on some simple spin glass models.

14. Bovier Anton and Kurkova Irina [2006]: Local Energy Statistics in Disordered systems: A proof of the local REM conjecture. CMP v.263 p. 513-533.

15. Bovier Anton and Kurkova, Irina [2006]: A tomography of the GREM: Beyond the REM conjecture. CMP v.263, p.535-552.

16. Bovier, A. and Kurkova, I. [2004]. 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.and 2. Models with Continuous Hierarchies. Ann. Inst. H. Poincare. Prob. et Statisques (B) Prob. Stat., 40, 480–485.

17. Bovier Anton [2006]: *Statistical Mechanics of Disordered Systems.* Cambridge Univ. Press.

18. Capocaccia D, Cassandro M and Picco P [1987]: On the Existence of Thermodynamics for the Generalized Random Energy Model. JSP v.46 p. 493-505.

19. Carmona Philippe and Hu Yueyun [2004]: Universality in Sherrington-Kirkpatrick's Spin Glass Model. arXiv:math PR/0403559 v2.

20. Chatterjee Sourav [2005]: A Simple Invariance Theorem. arXiv:math.PR/0508213v1.

21. Contucci, P., Esposti, M.D., Giardinà, C. and Graffi, S. [2003]. Thermodynamical Limit for Corelated Gaussian Random Energy Models. CMP, v. 236, p. 55–63.

22. Dembo, A. and Zeitouni, O. [1998]. *Large Deviation: Techniques and Applications.* Springer-Verlag.

23. Derrida, B. [1981]. Random Energy Model: An Exactly Solvable Model of Disordered Systems. Phys. Rev. B, v.46, no. 5, p.401–407.

24. Derrida, B. [1985]. A Generalization of The Random Energy Model Which Includes Correlations Between Energies. J. Phys. Lett. v.46, p.401–407.

25. Derrida, B. and Gardner, E. [1986]. Solution of the Generalized Random Energy Model. J. Phys. C, v.19, p.2253–2274.

26. Dorlas T C and Dukes W M B [2002]: Large deviation approach to the generalized random energy model. J. Phys. A: Math. gen v.35 p.4385-4394.

27. Dorlas, T C. and Wedagedera, J.R. [2001]. Large Deviations and The Random Energy Model. Int. J. of Mod. Phy. B, v. 15, p. 1-15.

28. Eisele Theodor [1983]: On a Third-Order Phase Transition. CMP v.90 p. 125-159.

29. Ellis, R. S. [1985]: *Entropy, Large Deviations and Statistical Mechanics.* Springer-Verlag.

30. Feynman R P [1972]: *Statistical Mechanics.* W. A. Benjamin Inc.

31. Galves Antonio, Martinez Servet and Pierre Picco [1989]: Fluctuations in Derrida's random Energy and Generalized Random Energy Models. JSP v.54 p. 515-529.

32. Georgii H. O. [1979]: *Canonical Gibbs measures.* Springer LNM 760.

33. Guerra Francesco [2003]: Broken Replica Symmetric Bounds in the Mean Field Spin Glass Model. CMP v.233 p.1-12.

34. Guerra Francesco and Toninelli Fabio Lucio [2002]: The thermodynamic limit in mean field spin glass models. CMP v.230 p. 71-79.

35. Guerra Francesco and Toninelli Fabio Lucio [2004]: The high temperature region of the Viana-Bray Spin Glass Model. JSP v.115, p.531-555.

36. Jana N.K [2006]: Exponential Random Energy Model. arXiv:math/0602670v1[math.PR]

37. Jana N.K [2007]: *Contributions to Random Energy Models.* Thesis submitted to the Indian Statistical Institute. http://arxiv.org/abs/0711.1249.

38. Jana N. K. and Rao B.V. [2006]: generalized Random Energy Model. JSP v.123, p. 1033-1058.

39. Jana N. K. and Rao B. V. [2007]: Generalized Random Energy Model II. JSP v.127, p.841-850.

40. Koukiou F [1990]: A Random Covering interpretation for the phase transition of the Random Energy Model. JSP v.60, p. 669-674.

41. Koukiou F and Picco [1991]: Poisson point Processes, Cascades and Random Coverings of R^n. JSP v62, p.481-489.

42. Liggett Thomas, Steif Jeffrey E and Balint Toth [2006]: Statistical Mechanical Systems on complete graphs, infinite exchangeability, finite extensions and a discrete finite moment problem. arXiv.math.PR/0512191v3.

43. Olivieri Enzo and Pierre Picco [1984]: On the Existence of Thermodynamics for the Random Energy Model. CMP v. 96 p.125-144.

44. Panchenko Dmitry and Michel Talagrand [2007]: Guerra's interpolation using Derrida-Ruelle Cascades. arXiv:0708.3641v1[math.PR].

45. Parisi Giorgio [2007]: Mean field theory of spin glasses: statics and dynamics. arXiv: 0706.0094v1[cond-mat.dis-nn]

46. Pastur L. a. and M V Shcherbina [1991]: Absence of the self averaging of the Order Parameter in the Sherrington-Kirkpatrick Model. JSP v.62 p.1-19.

47. Preston C. J. [1974]: *Gibbs states on countable sets.* Cambridge Univ Press.

48. Ruelle David [1979]: *Statistical Mechanics: Rigorous results.* W. A. Benjamin Inc, Third edition.

49. Ruelle David [1987]: A Mathematical Reformulation of Derrida's REM and GREM. CMP v.108 p.225-239.

50. Ruzmaikina Anastasia and Aizenman Michael [2005]: Characterization of invariant measures at the leading edge for competing particle systems. Ann.Prob, v33, p.82-113.

51. Shcherbina M and Tirozzi B [1993]: The Free Energy of a class of Hopfield Models. JSP v.72 p.113-125.

52. Sherrington David [2007]: Spin Glasses: A Perspective. See in [8] above.

53. Sinai Ya. G.[1982]: *Theory of Phase Transitions: Rigorous results.* Pergamon Press.

54. Steele J. M. [1997]: *Probability theory and combinatorial optimization.* SIAM publications.

55. Talagrand Michel [1996]: A new look at independence. Ann. Prob. v.24 p.1-34.

56. Talagrand Michel [2002]: Gaussian Averages, Bernoulli Averages and Gibbs' Measures. AIHP(?) p.197-204.

57. Talagrand Michel [2003]: The generalized Parisi formula. C.R. Acad Sci. Paris Ser I, v.337 p.111-114.

58. Talagrand Michel [2006]: The Parisi formula. Ann. Math. v.163. p. 221-263.

59. Talagrand Michel [2003]: *Spin Glasses: A Challenge for Mathematicians.* Springer Verlag.

60. Varadhan S. R. S. [1966]: Asymptotic probabilities and differential equations. Comm. pure. app.math. v.19. p. 261-286.

61. Varadhan S. R. S. [1984]: *Large Deviations and Applications.* CBMS-NSF conf. app.math. vol 46, SIAM.

62. Shepp L. A. [1972]: Covering the line with random intervals. Zeitsch. Wahr. verw. Geb. v.23 p. 163-170.

63. Comets F and Neveu J [1995]: The Sherrington-Kirkpatrick model of spin glasses and stochastic calculus; the high temperature case. CMP v.166 p.549-564.

64. Comets F [1998]: The martingale method for mean-field disordered systems at high temperature. see [11] above, p.91-113.

65. Panchenko Dmitry and Talagrand Michel [2007]: Guerra's interpolation using Derrida-Ruelle cascades. arXiv:0708.3641v1[math.PR].