Gibbs Measures of Derrida's Generalized Random Energy Models and the Genealogy of Neveu's Continuous State Branching Process

Anton Bovier[1,2]

Weierstraß–Institut
für Angewandte Analysis und Stochastik
Mohrenstrasse 39, D-10117 Berlin, Germany

Irina Kurkova[3]

Laboratoire de Probabilités et Modèles Aléatoires
Université Paris 6
4, place Jussieu, B.C. 188
75252 Paris, Cedex 5, France

Abstract: In this paper we conclude our analysis of Derrida's Generalized Random Energy Model (GREM) by describing the geometry of its Gibbs measure in the thermodynamic limit in terms of genealogies of a continuous state branching process introduced by Neveu.

Keywords: Gaussian processes, Generalized Random Energy Model, continuous state branching process, coalescent processes, genealogy, Ghirlanda-Guerra identities.

AMS Subject Classification: 82B44, 60G70,60K35

[1] e-mail: bovier@wias-berlin.de
[2] Research supported in part by the DFG in the Dutch-German Bilateral Research Group "Mathematics of Random Spatial Models from Physics and Biology".
[3] e-mail: kourkova@ccr.jussieu.fr

1. Introduction.

In a series of papers [BK1,BK2] we have recently taken up the analysis of a class of mean field spin glass models introduced by Derrida and Gardner in the 1980's [D,DG1,DG2]. In purely mathematical terms, these models can be described as follows. Consider the N dimensional hypercube $\mathcal{S}_N \equiv \{-1,1\}^N$ endowed with the (normalized) ultrametric distance $1-d_N$, where $d_N(\sigma,\tau) \equiv N^{-1}(\min(i : \sigma_i \neq \tau_i) - 1)$. Define a normal Gaussian processes X_σ indexed by $\mathcal{S}_N$ with covariance function

$$\mathbb{E} X_\sigma X_\tau = A(d_N(\sigma,\tau)) \tag{1.1}$$

for some nondecreasing function $A : [0,1] \to [0,1]$. The principal object of interest is the analysis of the asymptotic behavior of the Gibbs measures

$$\mu_{\beta,N}(\sigma) \equiv \frac{e^{\beta\sqrt{N}X_\sigma}}{Z_{\beta,N}} \tag{1.2}$$

where the partition function $Z_{\beta,N}$ assures that $\mu_{\beta,N}$ is a probability measure.

1.1. History of the models. Let us briefly dwell on the history of this problem. The model was introduced and analyzed by Derrida and Gardner [D,DG1,DG2] in the case when A is a step function with finitely many steps (the corresponding models are called GREMs, or Derrida's GREMs) in the sense that the limit of the free energy

$$F_{\beta,N} \equiv -\frac{1}{\beta N} \ln Z_{\beta,N} \tag{1.3}$$

and some further thermodynamics functions were computed. The computation of the free energy was later done rigorously in [CCP]. Derrida and Gardner then also considered limits of their results as the number of steps tended to infinity, and interpreted these results as corresponding to continuous functions A [DG1]. These results were then also compared to those of the more commonly studied (and more difficult) class of Sherrington-Kirkpatrick models (which essentially differs from the class studied here in that the covariance is a function of the Hamming distance rather than our hierarchical distance).

While there were very few further rigorous results on these models (but see [GMP]), Ruelle in a seminal paper of 1988 [Ru] introduced a new class of models based on *Poisson cascades* (to which we will henceforth refer to as "Ruelle's GREM") which he apparently understood to be the appropriate asymptotic models to describe the limiting Gibbs measures of Derrida's GREMs. Ruelle noted a number of remarkable features of these models, and in particular observed that it was possible to construct limits as the number of steps went to infinity in terms of projective limits. Surprisingly, his paper at no point contains a precise hint on how his models are to be related to the original spin glass models of Derrida.

Shortly after that, Neveu [Ne] noted a connection between Ruelle's models and continuous state branching processes. This paper also outlined a proof of the convergence of the rescaled partition function of the REM and GREM to a functional of the Poisson process, respectively Poisson cascades of Ruelle. Unfortunately, these observations are only contained in an internal report that was never published and that contains these ideas only in a somewhat embryonic form. Following a much later paper by Bolthausen and Sznitman [BoS], where it was explained how the results of replica theory of spin glasses can be interpreted in terms of a *coalescent process* (now known as the *Bolthausen-Sznitman coalescent*), Bertoin and Le Gall [BeLG] finally gave a precise and complete form of the relation between continuous state branching processes, the Ruelle's GREM and the Bolthausen-Sznitman coalescent.

Around the time when these fascinating results appeared, we began to investigate more closely the link to the original spin glass models with Ruelle's models. In the REM, this connection was then made in a paper with M. Löwe [BKL] which was elaborated on in the lecture notes by one of us [B] (see also [T3,T4, BoS2]). These results were extended to the GREMs in the paper [BK1], using essentially elementary methods (see also [BoS2], Ch. 9 for related results). We observed, however, that the use of the so-called Ghirlanda-Guerra identities [GG] allowed a different approach that circumvented parts of these explicit computations (this fact was first observed in the REM by Talagrand [T3] who also exploited these identities heavily in his work on the p-spin SK models [T1,T2,T3,T4]). In fact, these identities impose structural constraints on any limit point that allow to prove convergence of the Gibbs measure (in a suitable sense) only on the basis of the convergence of the free energy, and that, moreover, allow to characterize the limit. These observations allowed us in [BK2] to extend our convergence results to the general class of models defined above (which we call CREM in case the function A is not a step function).

In the present paper we want to conclude this investigation by linking our result up to the continuous state branching model, i.e. by identifying the limit proven to exist in [BK2] explicitly in terms of Neveu's branching process. This requires in fact little more than combining our results from [BK2] with those of Bertoin and Le Gall [BeLG], but we feel that the emerging complete picture is well worth to be put in evidence.

1.2. Geometry of Gibbs measures. Let us recall the central problem one is faced with when analyzing mean field spin glasses. What we want to do is to describe the geometric structure of a random probability measure on a set $\mathcal{S}_N$. One expects that this measure will concentrate (at low temperatures) on a relatively very small subset with rather complicated structure. Since due to randomness and symmetries there are no external references, we need a way to describe the structural geometric properties of such measures in an intrinsic, reference-free way. On the other hand, we need to allow sufficient compactness for limits to exist.

To resolve this problem, we introduced in [BK2] what we called the *empirical distance distribution function*, i.e. the random measure

$$\mathcal{K}_{\beta,N} \equiv \sum_{\sigma\in\mathcal{S}_N} \mu_{\beta,N}(\sigma)\delta_{m_\sigma(\cdot)} \tag{1.4}$$

where, for $t \in [0,1]$,

$$m_\sigma(t) \equiv \mu_{\beta,N}\left(\sigma' : d_N(\sigma,\sigma') > t\right). \tag{1.5}$$

This object describes the probability of a mass distribution around a randomly (according to the Gibbs measure) drawn point on $\mathcal{S}_N$. A key object is the mean first moment of this random measure,

$$\int \mathcal{K}_{\beta,N}(dm)m(t) \equiv 1 - f_{\beta,N}(t) \tag{1.6}$$

which is nothing but the probability that two configurations, σ, σ', drawn independently from the Gibbs sample satisfy $d_N(\sigma,\sigma') > t$. The function

$$f_{\beta,N}(t) \equiv \mu_{\beta,N}^{\otimes 2}\left(d_N(\sigma,\sigma') \leq t\right) \tag{1.7}$$

is now the analogue of Parisi's order parameter[4]. In [BK3] we proved that

$$\mathbb{E}f_{\beta,N}(t) \to \mathbb{E}f_\beta(t) = \begin{cases} \frac{\sqrt{2\ln 2}}{\beta\sqrt{\bar{a}(t)}}, & \text{if } t < t_\beta \\ 1, & \text{if } t \geq t_\beta \end{cases} \tag{1.8}$$

where $\bar{a}$ is the right-derivative of the concave hull of the function A, $t_\beta = \sup(t : \frac{\sqrt{2\ln 2}}{\beta\sqrt{\bar{a}(t)}} < 1)$. We also showed that

$$\mathcal{K}_{\beta,N} \xrightarrow{\mathcal{D}} \mathcal{K}_\beta. \tag{1.9}$$

[4] In the context of the SK models, this function is usually defined with d_N replaced by the "overlap" parameter $R_N(\sigma,\sigma') \equiv N^{-1}\sum_i \sigma_i\sigma'_i$. In [BK2] we have shown that in the GREM, the choice of the distance used in the definition of $f_{\beta,N}$ does not affect the result in the limit $N \uparrow \infty$.

The limit is uniquely determined by Ghirlanda-Guerra relations, which give recursive formulas to compute all moments of $\mathcal{K}_\beta$ starting from the function $\mathbb{E}f_\beta$ that determines the second moment.

In fact, while the random measures $\mathcal{K}_{\beta,N}$ may look somewhat unfamiliar, their moments are closely linked and even equivalent to the more conventional n-replica distance distribution $\mathbb{Q}^{(n)}_{\beta,N}$. These are measures on the space $[0,1]^{n(n-1)/2}$

$$\mathbb{Q}^{(n)}_{\beta,N}(\mathcal{A}) \equiv \mathbb{E}\mu^{\otimes n}_{\beta,N}\left(\left(d_N(\sigma^i,\sigma^j)\right)_{1\leq i<j\leq N} \in \mathcal{A}\right) \tag{1.10}$$

Note that these measures do of course give full measure to sets that respect the ultrametric triangle relations. In [BK2] we proved their convergence to a limiting distribution $\mathbb{Q}^{(n)}_\beta$. The Ghirlanda-Guerra identities (together with the fact that $1-d_N$ is an ultrametric distance) allow to compute $\mathbb{Q}^{(n+1)}_\beta$ in terms of $\mathbb{Q}^{(n)}_\beta$ recursively, while $\mathbb{Q}^{(2)}_\beta$ has distribution function $\mathbb{E}f_\beta(t)$. On the other hand, the full set of distributions $\mathbb{Q}^{(n)}_\beta$ determines the limiting random measures $\mathcal{K}_\beta$ through its moments.

The measures $\mathbb{Q}^{(n)}_\beta$ are the marginals of the probability distribution of an ultrametric ("genealogical") distance distribution on the positive integers. It is not difficult (and will be explained in Section 6) to show that the Ghirlanda-Guerra relations allow to relate these to the coalescent process introduced by Bolthausen and Sznitman [BS]. The work of Bertoin and Le Gall [BeLG] allows then to link this to Neveu's branching process. Implicitly, this also determines the limit of $\mathcal{K}_{\beta,N}$.

1.3. Aim of the paper. In this paper we want to close what we feel is a small final gap in our understanding. This is related to the question whether we can identify a limiting measure to which our Gibbs measures converge and that encodes the full geometric information contained in $\mathcal{K}_\beta$. As we will explain in some detail at the beginning of Section 2, this is not immediately possible. What will however be possible, is the following. We will introduce the notion of a flow of compatible probability measures on $[0,1]$ indexed by pairs of parameters $s\leq t\in I$ and with distribution functions satisfying the compatibility assumption (2.2). Next, we will associate to each of such flows a certain genealogical structure on $[0,1]$ described by a genealogical map $K_T\in M_1(M_1([0,1]))$ which is an empirical distribution of family sizes of all individuals as functions of degree of relatedness. Then we will provide a flow of compatible probability measures for each finite N with the genealogy describing efficiently the geometry of the Gibbs measure of the CREM: its genealogical map $K^{\beta,N}_T$ will equal the empirical distance distribution function $\mathcal{K}_{\beta,N}$. Finally, we will show that this flow of probability measures converges as $N \to \infty$ to the flow of compatible random probability measures with distribution functions that are normalized stable subordinators associated to Neveu's continuous state branching process via an appropriate deterministic time change. This convergence of flows is understood in the sense that their genealogical maps $\mathcal{K}_{\beta,N} = K_T^{\beta,N}$ converge. Thus the limiting geometry of the Gibbs measure of the CREM will be expressed in terms of the genealogy of Neveu's continuous state branching process modulo a time change determined only by $\mathbb{E}f_\beta(y)$ of (1.9).

1.4. Organization of the paper. The remainder of the paper is organized as follows. In Section 2 we define the notion of a flow of compatible probability measures in Definition 2.1 and associate to it a genealogical structure: a genealogical map $K_T \in M_1(M_1([0,1]))$ and a coalescent process on the integers. We show that K_T is determined by its moments that can be expressed as distances between integers of the corresponding coalescent.

In Section 3 we provide for all finite N a flow of compatible probability measures with the genealogical map $K_T^{\beta,N}$ that equals the empirical distance distribution function $\mathcal{K}_{\beta,N}$.

In section 4 we describe the flow of compatible probability measures associated to Neveu's continuous state branching process. Their probability distribution functions are normalized stable subordinators verifying the compatibility condition. The coalescent associated with this flow is the one of Bolthausen-Sznitman by [BeLG].

In Section 5 we formulate our main theorem. It states that the empirical distance distribution function $\mathcal{K}_{\beta,N}$ (which is the genealogical map of the flow of Section 3) converges to the genealogical map of the flow of measures associated with Neveu's branching process of Section 4 via an appropriate time change.

In Sections 6 and 7 we prove this theorem. As it was established in Section 2, it suffices to show the convergence of moments, that is that the n-replica distance distribution functions (1.10) of our spin glass model converge to the genealogical distance distribution functions of the Bolthausen-Sznitman coalescent under an appropriate time change. One way (short but indirect) to prove this is indicated in Section 6 and relies on the connection between Neveu's branching process and Ruelle's probability cascades established in [BeLG].

The second way (more direct) is given in Section 7: it consists in showing that the Bolthausen-Sznitman coalescent satisfies Ghirlanda-Guerra identities. For that purpose we use the Chinese restaurant process of J. Pitman [P].

We hope that the results presented in this class of models elucidate in a mathematically comprehensible context the fundamental and universal role played by Neveu's continuous state branching process as a universal random mechanism governing the extremal processes for a wide class of Gaussian processes. If one accepts the common belief of theoretical physicists, its role goes well beyond the class of models we discuss here. Even on a slightly less speculative level, Neveu's process will emerge in any model for which the Ghirlanda-Guerra relations hold in their strong form, which means in particular that this will be the case if not for the actual SK model, then at least for models where weak additional fields have been added to the Hamiltonians (see [GG,Le,T2]). Recent progress on the validity of Parisi's solution by Guerra [G02], Aizenman-Sims-Starr [ASS], and Talagrand [T5] on the level of the free energy makes it very credible that this will indeed be the case. Moreover, we also hope that these examples help to explain to a mathematical audience what physicist describe when they talk about "continuous replica symmetry breaking", and how such a phenomenon can actually arise.

Acknowledgements: We are grateful for numerous enlightening discussions with Jean Bertoin on continuous state branching, subordinators, and coalescents. A.B. also acknowledges inspiring discussions with Anton Wakolbinger. I.K. thanks Ph. Marchal for helpful discussions. Part of this work was done while A.B. was resident at the Erwin-Schroedinger Institute for Mathematical Physics (ESI) in Vienna in the programme "Mathematical population genetics and statistical physics". He thanks the organisers, Ellen Baake, Michael Baake, and Reinhard Bürger for the kind invitation and the ESI for financial support. I.K. thanks the Weierstrass Institute for Applied Analysis and Stochastics, Berlin, for its hospitality and financial support.

2. Genealogy of a flow of probability measures

In [B] one of us proposed to describe the infinite volume limit of the Gibbs measure in the Random Energy Model by considering its image on the unit interval through the map $r_N : \mathcal{S}_N \to (0,1]$ defined as

$$r_N(\sigma) \equiv 1 - \sum_{i=1}^{N} 2^{-i}(1+\sigma_i)/2. \tag{2.1}$$

It was shown that the phase transition in the REM manifests itself by the fact that the resulting image measure converges to Lebesgue measure in the high temperature phase ($\beta \leq \sqrt{2\ln 2}$) and towards a dense pure point measure in the low-temperature phase ($\beta > \sqrt{2\ln 2}$).

While this shows the existence of a phase transition, the limiting measure does not describe the geometry of the Gibbs measure. In the REM this is no problem, since the geometry is trivial. But in the GREMs a nontrivial geometry emerges, and our purpose would be to identify a measure on $[0,1]$ that represents the limiting Gibbs measure. It will be instructive to explain how a naive approach to do so fails. On the hypercube we are interested in the masses of sets $\{\sigma' : d_N(\sigma,\sigma') > t\}$ (1.5). If we map such sets on the unit interval via r_N, we obtain intervals $(r_{[Nt]} - 2^{-[tN]}, r_{[Nt]}]$ of length $2^{-[tN]}$. In fact there is no difficulty to express e.g. $\mathcal{K}_{\beta,N}$ for N fixed in terms of defined quantities with respect to the image measure on the hypercube. However, the construction involves masses of intervals of exponentially small size (in N). So what should one do in the limit when N is infinite? We cannot analyst the structure by looking at intervals of the size $2^{-t\infty}$.

What is needed is clearly a construction that does not refer explicitly to masses of intervals of exponentially small size while still revealing the fine structure of the measure at such a scale. In this section we show that a canonical construction exists when we consider a flow of probability measures on $[0,1]$.

2.1. Genealogical map of a flow of probability measures.

Definition 2.1: *A two-parameter family of measures with probability distribution functions $S^{(s,t)}$ on $[0,1]$, $s \leq t$, $s,t \in I \subset \mathbb{R}$, is called a flow of compatible probability measures on I, if and only if for any collection $t_1 \leq t_2 \leq \cdots \leq t_n \subset I$*

$$S^{(t_1,t_n)} = S^{(t_{n-1},t_n)} \circ S^{(t_{n-2},t_{n-1})} \circ \ldots S^{(t_2,t_3)} \circ S^{(t_1,t_2)} \tag{2.2}$$

holds.

Let us admit the following terminology. We say that each point $a \in [0,1]$ is an individual in generation s and its image $S^{(s,t)}(a) \in [0,1]$ is its offspring in generation t. Let us define for any distribution function $\Theta(x)$ its inverse function

$$\Theta^{-1}(x) = \inf\{a \mid \Theta(a) \geq x\}. \tag{2.3}$$

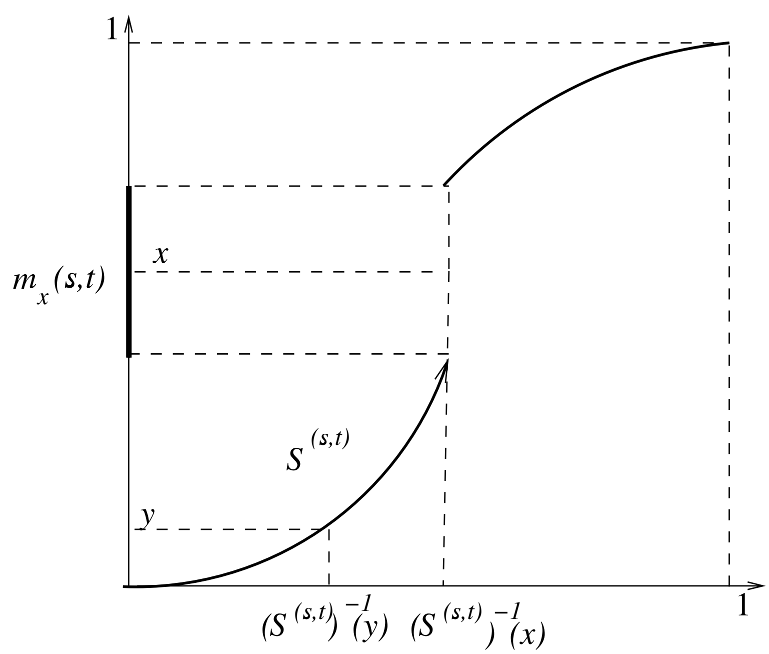
Then each individual $x \in [0,1]$ in generation t has an ancestor a in generation s which is $a = (S^{(s,t)})^{-1}(x)$.

Given an individual $x \in [0,1]$ in generation t, let us look for individuals x' having the same ancestor as x in generation s:

$$m_x(s,t) \equiv \{x' : (S^{(s,t)})^{-1}(x') = (S^{(s,t)})^{-1}(x)\}. \tag{2.4}$$

Two situations are possible. In the first case $S^{(s,t)}$ is continuous at $a = (S^{(s,t)})^{-1}(x)$. Then any individual $x' \neq x$ has a different ancestor from the one of x, i.e. $(S^{(s,t)})^{-1}(x') \neq (S^{(s,t)})^{-1}(x)$, $m_x(s,t) = \{x\}$. (In fact, by definition of $(S^{(s,t)})^{-1}$ it should be $S^{s,t}(a-\epsilon) < S^{s,t}(a)$ for any $\epsilon > 0$. Then for any $x' < x$ $(S^{(s,t)})^{-1}(x') < a$. If $S^{(s,t)}$ is strictly increasing at a, then clearly $(S^{(s,t)})^{-1}(x') > a$ for any $x' > x$. If $S^{s,t}$ is constant on $[a,b)$ and continuous at a, then $(S^{s,t})^{-1}(x') \geq b > a$ for any $x' > x$.) In the second case $S^{(s,t)}$ makes a jump at $a = (S^{(s,t)})^{-1}(x)$. Thus $S^{s,t}(a) > x$. In this case any individual $x' \in (\lim_{\epsilon\downarrow 0} S^{(s,t)}(a-\epsilon), S^{(s,t)}(a)]$ has the same ancestor as x in generation s i.e. $(S^{(s,t)})^{-1}(x') = a$. Hence, the family (2.4) of the individual x having the same ancestor as x in generation s is the following interval :

$$m_x(s,t) = \lim_{\epsilon\downarrow 0} \Big(S^{(s,t)}\big((S^{(s,t)})^{-1}(x) - \epsilon\big) \ , \ S^{(s,t)} \circ (S^{(s,t)})^{-1}(x) \Big] .$$



In Figure 2.1 the individual x in generation t has a family of "cousins" $m_x(s,t)$ having the same "grand-father" in generation s, while the individual y is the unique "grand-child" of his ancestor in generation s. We are mainly interested in a nontrivial case when functions $S^{(s,t)}$ make jumps.

The next lemma justifies our terminology. It says that any individual having an ancestor in common with x in generation s has necessarily an ancestor in common with x in any generation $s' < s$. In other words, if we partition the interval $[0,1]$ into families $m_x(s',t)$

having the same ancestor in generation s', then the partition into families $m_x(s,t)$ having the same ancestor in generation $s > s'$ is a refinement of the previous one.

Lemma 2.2: *Let $S^{(s,t)}$ be distribution functions of a flow of measures according to Definition 2.1. Then for all $x \in [0,1]$*

$$m_x(s,t) \subset m_x(s',t) \quad \forall s' < s \le t \in I. \tag{2.5}$$

Proof: From one hand, by definition (2.3) and compatibility (2.2) we have the inequality

$$(S^{(s,t)})^{-1}(x) \le S^{(s',s)} \circ (S^{(s',s)})^{-1} \circ (S^{(s,t)})^{-1}(x) = S^{(s',s)} \circ (S^{(s',t)})^{-1}(x) \tag{2.6}$$

leading to the upper bound

$$S^{(s,t)} \circ (S^{(s,t)})^{-1}(x) \le S^{(s,t)} \circ S^{(s',s)} \circ (S^{(s',t)})^{-1}(x) = S^{(s',t)} \circ (S^{(s',t)})^{-1}(x). \tag{2.7}$$

From the other hand, for any $\epsilon > 0$ by definition (2.3) $x > \Theta(\Theta^{-1}(x) - \epsilon)$. Then

$$(S^{(s,t)})^{-1}(x) > S^{(s',s)} \circ \big((S^{(s',s)})^{-1} \circ (S^{(s,t)})^{-1}(x) - \epsilon\big) = S^{(s',s)} \circ \big((S^{(s',t)})^{-1}(x) - \epsilon\big). \tag{2.8}$$

Then for any $\epsilon > 0$ one can find $\delta(\epsilon) > 0$ such that $\delta(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$ and

$$(S^{(s,t)})^{-1}(x) - \delta(\epsilon) > S^{(s',s)} \circ \big((S^{(s',t)})^{-1} - \epsilon\big). \tag{2.9}$$

Thus

$$S^{(s,t)}\big((S^{(s,t)})^{-1}(x) - \delta(\epsilon)\big) > S^{(s,t)} \circ S^{(s',s)}\big((S^{(s',t)})^{-1}(x) - \epsilon\big) = S^{(s',t)}\big((S^{(s',t)})^{-1}(x) - \epsilon\big). \tag{2.10}$$

Letting $\epsilon \downarrow 0$ in the inequality (2.10) yields the necessary lower bound, which together with (2.7) proves the lemma. $\diamondsuit$.

Whenever $t = T$ is fixed, the function $|m_x(\cdot, T)|$ is the family size of the individual x in generation T as a function of the degree of relatedness. By Lemma 2.2 it is a decreasing function on I. Finally, we define the associated empirical distribution of the functions $|m_x(\cdot, T)|$

$$\mathcal{K}_T = \int_0^1 dx \delta_{|m_x(\cdot, T)|}. \tag{2.11}$$

This construction (2.11) allows to associate to any flow of probability measures in the sense of Definition 2.1 an empirical distribution $\mathcal{K}_T$.

If we assume in addition that $[0,T] \subset I$ and $|m_x(\cdot,T)|$ are right-continuous, then $1-|m_x(\cdot,T)|$ are probability distribution functions. Then we will think of $\mathcal{K}_T$ as a map from flows of probability measures into $M_1(M_1([0,1]))$ which we call the genealogical map.

2.2 Coalescent associated with a flow of probability measures.

Now, let us define the exact degree of relatedness between two individuals $x, y \in [0,1]$ with respect to a flow of measures (2.2) as

$$\gamma_T(x,y) \equiv \sup\left(s \in I : y \in m_x(s,T)\right). \tag{2.12}$$

Lemma 2.3: *$1-\gamma_T$ defines an ultrametric distance on the unit interval.*

Proof: By Lemma 2.2, for all $x, y \in [0,1]$, if $s = \gamma_T(x,y)$, then

$$y \in m_x(s',T) \ \ \forall s' < s, \quad s' \in I. \tag{2.13}$$

It follows from (2.13) that for any $x,y,z \in [0,1]$ if $\gamma_T(x,y) \neq \gamma_T(x,z)$, then $\gamma_T(y,z) = \min\{\gamma_T(x,z), \gamma_T(x,y)\}$. In fact, let e.g. $\gamma_T(x,z) > \gamma_T(x,y)$. Then $z \in m_x(s,T)$ for all $s \in I$ such that $s \leq \gamma_T(x,y)$ and then $\gamma_T(y,z) \geq \gamma_T(x,y)$. From the other point of view, if $\gamma_T(y,z) > \gamma_T(x,y)$, then either $\gamma_T(x,z) \geq \gamma_T(y,z) > \gamma_T(x,y)$ or $\gamma_T(y,z) > \gamma_T(x,z) > \gamma(x,y)$. In the first case by (2.13) $x \in m_z(s,T)$ for all $s \in I$ such that $s < \gamma_T(y,z)$ and then $\gamma_T(x,y) \geq \gamma_T(y,z)$ which is impossible. In the second $y \in m_z(s,T)$ for all $s \in I$ such that $s \leq \gamma_T(x,z)$ from where $\gamma_T(x,y) \geq \gamma_T(x,z)$, which is again impossible. Thus $\gamma_T(y,z) = \gamma_T(x,y)$.

Note also that if $\gamma_T(x,y) = \gamma_T(x,z)$, then $\gamma_T(y,z) \geq \gamma_T(x,y) = \gamma_T(x,z)$. These observations imply that $1-\gamma_T$ is an ultrametric distance on $[0,1]$. $\diamondsuit$

The function γ_T is trivial if $S^{(s,t)}$ are all continuous, for then $\gamma_T(x,y) = T$ if $x = y$ and $\gamma_T(x,y) = -\infty$ if $x \neq y$, and in a strict sense nobody has any relatives. On the other hand, in the discontinuous case, rather large families exist, and the ultrametric structure of the interval can be very rich.

We will be interested in cases where the flow $S^{(s,t)}$ of Definition 2.1 is *random*. We will now describe a useful way of characterizing a random genealogical map K_T in this case.

Having defined a distance $1-\gamma_T$ on $[0,1]$, we can define in a very natural way the analogous distance on the integers. To do this, consider a family of i.i.d. random variables $\{U_i\}_{i\in\mathbb{N}}$ distributed according to the uniform law on $[0,1]$. Given such a family, we set

$$\rho_T(i,j)=\gamma_T(U_i,U_j). \tag{2.14}$$

Due to the ultrametric property of the γ_T and the independence of the U_i, for fixed T, the sets $B_i(s)\equiv\{j:\rho_T(i,j)\geq s\}$ form an exchangeable random partition of the integers. Moreover, the family of these partitions as a function of $T-s$ is a stochastic process on the space of integer partitions with the property that for any $s>s'$, the partition $B_i(s')$ is a coarsening of the partition $B_i(s)$. Such a process is called a *coalescent process* (see e.g. [Be1,Be2,BeLG,BePi,BeYo,BoS,P,Pi1,PPY]).

The key observation for our purposes is the possibility to express the moments of K_T in terms of this coalescent [Be1]. Namely, it is plain by the law of large numbers that

$$\lim_{n\uparrow\infty} n^{-1}\sum_{j=1}^{n} 1\!\!1_{j\in B_k(s)} = |m_{U_i}(s,T)| \quad \text{a.s.} \tag{2.15}$$

for any i such that $i\in B_k(s)$. This implies, for instance, as shown in [Be1] that

$$\mathbb{E}\int dx|m_x(s,T)|=\mathbb{P}[2\in B_1(s)] \tag{2.16}$$

and more generally that

$$\mathbb{E}\int dx|m_x(s,T)|^k=\mathbb{P}[2,3,\dots,k+1\in B_1(s)]. \tag{2.17}$$

Here the expectation $\mathbb{E}$ is with respect to the randomness of the family of measures $S^{(s,t)}$, and $\mathbb{P}$ is the law with respect to the random genealogy (depending both on the random measures and the i.i.d. r.v. U_i). We will need slightly more general expressions, namely a family of moments that determine the law of K_T. These can be written as follows. Let take any positive integer p, a collection of positive real numbers $0<t_1<\dots<t_p\leq T$, a positive integer ℓ, and non-negative integers $k_{11},\dots,k_{1p},k_{21},\dots,k_{2p},\dots,k_{\ell 1},\dots,k_{\ell p}$. Then we need

$$M(p,\underline{t},\underline{\underline{k}})\equiv\mathbb{E}\left(\int dx|m_x(t_1,T)|^{k_{11}}\dots|m_x(t_p,T)|^{k_{1p}}\right)\dots\left(\int dx|m_x(t_1,T)|^{k_{\ell 1}}\dots|m_x(t_p,T)|^{k_{\ell p}}\right). \tag{2.18}$$

By (2.15) we have that

$$\begin{aligned}&\int dx|m_x(t_1,T)|^{k_{11}}\dots|m_x(t_p,T)|^{k_{1p}}\\ &=\lim_{n\uparrow\infty} n^{-1-k_{11}-\dots-k_{1p}}\sum_{i=1}^{n}\sum_{j_1^1,\dots,j_{k_{11}}^1,\dots,j_1^p,\dots,j_{k_{1p}}^p}^{n} 1\!\!1_{j_1^1,\dots,j_{k_{11}}^1\in B_{k(i,t_1)}(t_1)}\cdots 1\!\!1_{j_1^p,\dots,j_{k_{1p}}^p\in B_{k(i,t_p)}(t_p)}\end{aligned} \tag{2.19}$$

where $k(i,t_p)$ is the smallest integer such that $k(i,t_p) \in B_i(t_p)$. Let us note first that in these expressions contributions from terms where two indices are equal can be neglected. Second, since

$$B_{k(i,t_p)}(t_p) \subset B_{k(i,t_{p-1})}(t_{p-1}) \subset \cdots \subset B_{k(i,t_1)}(t_1) \tag{2.20}$$

the summand in (2.19) is the same as

$$\mathbb{1}_{j_1^1,\dots,j_{k_{11}}^1,\dots,j_1^p,\dots,j_{k_{1p}}^p \in B_{k(i,t_1)}(t_1)} \cdots \mathbb{1}_{j_1^p,\dots,j_{k_{1p}}^p \in B_{k(i,t_p)}(t_p)} \tag{2.21}$$

Then

$$\begin{aligned} M(p,\underline{t},\underline{\underline{k}}) &= \lim_{n\uparrow\infty} n^{-\ell-k_{11}-\cdots-k_{\ell p}} \mathbb{E} \sum_{i_1,\dots,i_\ell} \sum_{j_1^{1,1},\dots,j_{k_{11}}^{1,1},\dots,j_1^{p,1},\dots,j_{k_{1p}}^{p,1}} \cdots \sum_{j_1^{1,\ell},\dots,j_{k_{\ell 1}}^{1,\ell},\dots,j_1^{p,\ell},\dots,j_{k_{\ell p}}^{p,\ell}} \\ &\mathbb{1}_{j_1^{1,1},\dots,j_{k_{11}}^{1,1},\dots,j_1^{p,1},\dots,j_{k_{1p}}^{p,1} \in B_{k(i_1,t_1)}(t_1)} \cdots\cdots \mathbb{1}_{j_1^{1,\ell},\dots,j_{k_{\ell 1}}^{1,\ell},\dots,j_1^{p,\ell},\dots,j_{k_{\ell p}}^{p,\ell} \in B_{k(i_\ell,t_1)}(t_1)} \\ &\dots \mathbb{1}_{j_1^{p,1},\dots,j_{k_{1p}}^{p,1} \in B_{k(i_1,t_p)}(t_p)} \cdots\cdots \mathbb{1}_{j_1^{p,\ell},\dots,j_{k_{\ell p}}^{p,\ell} \in B_{k(i_\ell,t_p)}(t_p)} \\ &= \mathbb{P}\left[J_{11} \in B_1(t_1), \dots, J_{1p} \in B_1(t_p), \dots J_{\ell 1} \in B_\ell(t_1), \dots, J_{\ell p} \in B_\ell(t_p)\right] \end{aligned} \tag{2.22}$$

where

(i) $J_{11},\dots,J_{\ell 1}$ is a disjoint partition of $\{\ell+1,\ell+2,\dots,k_{11}+\cdots+k_{\ell p}+\ell\}$,

(ii) For all $j=1,\dots,\ell$, $i=1,\dots,p$, $J_{ji} \supset J_{ji+1}$, and

(iii) $|J_{ji}| = k_{ji}+k_{ji+1}+\cdots+k_{jp}$.

By exchangeability, the choice of the partition and the subsets is irrelevant. The probabilities (2.22) can be expressed alternatively in the form

$$\mathbb{P}(\rho_T(i,j) \geq t_{m(i,j)}, \quad \forall i \in \{\ell+1,\dots,k_{11}+\cdots+k_{\ell p}+\ell\}, j \in \{1,\dots,\ell\}) \tag{2.23}$$

where $m(i,j) \in \{1,\dots,p\}$. Thus the genealogical map K_T is completely determined by the probabilities (2.22) or (2.23) of the corresponding coalescent through the family of its moments.

3. Finite N setting for the CREM.

We will now show that for finite N we can use the general construction from the preceding section to relate the geometric description of the Gibbs measure on $\mathcal{S}_N$ to the genealogical description of a family of embedded measures on $[0,1]$.

Recall that our basic objects are configurations $\sigma \in \mathcal{S}_N \equiv \{-1,1\}^N$ equipped with $d_N(\sigma,\tau) \equiv N^{-1}(\min(i : \sigma_i \neq \tau_i) - 1)$. For non-decreasing function $A : [0,1] \to [0,1]$, we have a Gaussian process X_σ with mean zero and covariance

$$\mathbb{E}X_\sigma X_\tau = A(d_N(\sigma,\tau)). \tag{3.1}$$

We can map the corresponding Gibbs measure $\mu_{\beta,N}(\sigma) = \frac{e^{\beta\sqrt{N}X_\sigma}}{\sum_\sigma e^{\beta\sqrt{N}X_\sigma}}$ on $\mathcal{S}_N$ to a measure $\tilde{\mu}_{\beta,N}$ on the unit interval through the map r_N defined in (2.1) via

$$\widetilde{\mu}_{\beta,N} \equiv \sum_{\sigma\in\mathcal{S}_N} \delta_{r_{[N]}(\sigma)}\mu_{\beta,N}(\sigma). \tag{3.2}$$

Let $\theta_{\beta,N}$ be the probability distribution function of $\widetilde{\mu}_{\beta,N}$:

$$\theta_{\beta,N}(x) = \mu_{\beta,N}(\sigma : r_{[N]}(\sigma) \leq x). \tag{3.3}$$

Let us take a parameter $s \in [0,1]$ and consider the map $r_{[sN]} : \mathcal{S}_N \to [0,1]$. Clearly, its image consists of $2^{[sN]}$ points, and for any σ,σ' with $d_N(\sigma,\sigma') > s$ we have $r_{[sN]}(\sigma) = r_{[sN]}(\sigma')$. Now we define a family of compatible distribution functions in the sense of Definition 2.1:

$$S^{(s,t)}_{\beta,N}(a) = \sum_\sigma \mu_{\beta,N}(\sigma)\mathbb{I}_{\{\theta(r_{[sN]}(\sigma))\leq a\}} \tag{3.4}$$

as states the following lemma.

Lemma 3.1: *The functions (3.4) verify the assumptions of Definition 2.1 with* $I = [0,1]$.

Proof: To understand better the construction of (3.4), let us take configurations $\sigma^1,\sigma^2,\dots,\sigma^{2^{[sN]}}$ having the first $[sN]$ spins different, i.e. with $d_N(\sigma^i,\sigma^j) \leq s$ and arrange them in order such that

$$0 < r_{[sN]}(\sigma^1) < r_{[sN]}(\sigma^2) < \cdots < r_{[sN]}(\sigma^{2^{[sN]}}) = 1.$$

Let

$$x^s_i = \mu_{\beta,N}(\sigma' : d_N(\sigma',\sigma^i) > s),\ i = 1,\dots,2^{[sN]}, \quad x_0 = 0. \tag{3.5}$$

Define

$$y^s_i \equiv x^s_0 + x^s_1 + \cdots + x^s_i = \theta(r_{[sN]}(\sigma^i)), \quad i = 0,1,\dots,2^{[sN]}. \tag{3.6}$$

Then we may write the representation

$$S^{(s,t)}_{\beta,N}(a) = \sum_{i=0}^{2^{[sN]}} y^s_i \mathbb{I}_{\{a\in[y^s_i,y^s_{i+1})\}}. \tag{3.7}$$

Thus $S^{(s,t)}_{\beta,N}$ is a right-continuous step function that makes jumps at points $y^s_1, y^s_2, \dots, 1$ for the values $x^s_1, x^s_2, \dots, x^s_{2^{[sN]}}$.

If we denote by $Y^s_{\beta,N} = \{y^s_0, y^s_1, \dots, 1\}$, then

$$S^{(s,t)}_{\beta,N}(y) = y \quad \forall y \in Y^s_{\beta,N}. \tag{3.8}$$

The function $S^{(s,t)}_{\beta,N}$ does not depend on the second parameter t. For any $s' > s$, $Y^{s'}_{\beta,N} \subset Y^s_{\beta,N}$, any segment $[y^{s'}_i, y^{s'}_{i+1})$ contains $2^{[sN]-[s'N]}$ points of $Y^s_{\beta,N}$ and $S^{(s',t)}_{\beta,N}$ is a coarse-grained version of $S^{(s,t)}_{\beta,N}$. Now it is straightforward to verify the assumption (2.2): for any $s' < s \le t$

$$S^{(s,t)}_{\beta,N} \circ S^{(s',s)}_{\beta,N} = S^{(s',t)}_{\beta,N}. \tag{3.9}$$

In fact, for any $a \in [0,1]$ $S^{(s',s)}_{\beta,N}(a) \in Y^{s'}_{\beta,N} \subset Y^s_{\beta,N}$, then by (3.8) $S^{(s,t)}_{\beta,N}(S^{(s',s)}_{\beta,N}(a)) = S^{(s',s)}_{\beta,N}(a) = S^{(s',t)}_{\beta,N}(a)$ since $S^{(s',t)}_{\beta,N}$ does not depend on t. $\diamondsuit$

Since the functions (3.4) satisfy Definition 2.1, we are entitled to apply to them the construction of the previous section. Their genealogy is

$$m_x(s,t) = (y^s_{i-1}, y^s_i] \text{ with } |m_x(s,t)| = |x^s_i|, \text{ if } x \in (y^s_{i-1}, y^s_i],\ i = 1, \dots, 2^{[sN]}. \tag{3.10}$$

We may associate to this genealogy the genealogical map (2.11) K_T and the coalescent process on the integers. The next lemma expresses the geometry of the Gibbs measure of the CREM contained in the empirical distance distribution function (1.4) $\mathcal{K}_{\beta,N}$ in terms of the genealogy induced by the functions (3.4).

Lemma 3.2: *We have*

$$\mathcal{K}_{\beta,N} = K^{\beta,N}_1,$$

where the empirical distance distribution function $\mathcal{K}_{\beta,N}$ is defined in (1.4) and $K^{\beta,N}_1$ is the genealogical map (2.11) with $T = 1$ of the flow of measures with probability distribution functions (3.4).

Proof: For any $\sigma \in \mathcal{S}_N$ one could find $i = 1, \dots, 2^{[sN]}$ such that $d_N(\sigma, \sigma^i) > s$, then

$$x^s_i = \mu_{\beta,N}(\sigma' : d_N(\sigma', \sigma) > s). \tag{3.11}$$

Then $r_N(\sigma) \in \left(r_{[sN]}(\sigma^i) - 2^{[sN]}, r_{[sN]}(\sigma^i)\right]$ and consequently

$$\theta(r_N(\sigma)) \in \left(\mu_{\beta,N}(\sigma' : r_N(\sigma') \le r_{[sN]}(\sigma^i) - 2^{[sN]}), \mu_{\beta,N}(\sigma' : r_N(\sigma') \le r_{[sN]}(\sigma^i))\right] = (y^s_{i-1}, y^s_i].$$

It follows from (3.10) that for any $\sigma \in \mathcal{S}_N$

$$m_{\theta(r_{[N]}(\sigma))}(s,t) = (y^s_{i-1}, y^s_i], \quad \forall s \le t. \tag{3.12}$$

Then by (3.11)

$$|m_{\theta(r_{[N]}(\sigma))}(s,t)| = x^s_i = \mu_{\beta,N}(\sigma' : d_N(\sigma', \sigma) > s).$$

Hence, in terms of the quantity $m_\sigma(s)$ defined in (1.5)

$$|m_{\theta(r_{[N]}(\sigma))}(s,t)| = m_\sigma(s). \tag{3.13}$$

This implies the statement of the lemma $\diamondsuit$.

Remark: If we would like to construct non-decreasing step functions compatible in the sense (2.2) with the genealogy (3.12), then for all $t \ge s$ it should be $S^{(s,t)}([0,1]) = Y^s_{\beta,N}$. By compatibility necessarily $S^{(s,t)}(S^{(s,s)}(x)) = S^{(s,t)}(x)$, from where $S^{(s,t)}(Y^s_{\beta,N}) = Y^s_{\beta,N}$. Since $S^{(s,t)}$ is non-decreasing, it must be $S^{(s,t)}(y) = y$ for all $y \in Y^s_{\beta,N}$. There are only two possibilities to construct a flow of non-decreasing functions such that $S^{(s,t)}(y) = y$ for all $y \in Y^s_{\beta,N}$ and $S^{(s,t)}([0,1]) = Y^s_{\beta,N}$. The one is (3.4) or equivalently (3.7). The other is

$$\hat{S}^{(s,t)}_{\beta,N}(a) = \sum_{i=0}^{2^{[sN]}-1} y^s_{i+1} \mathbb{1}_{\{a \in (y^s_i, y^s_{i+1}]\}}. \tag{3.14}$$

These non-decreasing functions are not probability distribution functions in a strict sense as they are left-continuous. They verify the compatibility (2.2). Since in the previous section we never used the right-continuity of the functions $S^{(s,t)}$ of Definition 2.1, we may apply to (3.14) the genealogical construction as well. Of course, the genealogical map associated with (3.14) coincides with the one of (3.7) and then with $\mathcal{K}_{\beta,N}$ by Lemma 3.2.

Remark: The functions (3.14) are in fact an example of a more general construction of a flow of functions always verifying (2.2). Let $\{\Theta_t\}_{t \in I \subset \mathbb{R}}$ be a one-parameter family of probability distribution functions on $(0,1]$. Let $\mathcal{I}_t$ denote the set of points of increase of the function Θ_t. We say that the family is refining, if for any $s < t \in I$, $\mathcal{I}_s \subset \mathcal{I}_t$. Let Θ_t be a refining family. Let us define the functions

$$S^{(s,t)} = \Theta_t((\Theta_s)^{-1}). \tag{3.15}$$

Then the functions $S^{(s,t)}$ verify the compatibility (2.2). In fact, by definition $S^{(s,t)} \circ S^{(s',s)}(x) = \Theta_t(\Theta_s^{-1}(\Theta_s(\Theta_{s'}^{-1}(x))))$. Since $y = \Theta_{s'}^{-1}(x)$ is the smallest value for which $\Theta_{s'}(y) \ge x$, for any $y' < y$ it must be true that $\Theta_{s'}(y') < \Theta_{s'}(y)$. Thus $\Theta_{s'}^{-1}(x) \in \mathcal{I}_{s'}$. But if y is a point of increase of $\Theta_{s'}$, by assumption, y is also a point of increase of Θ_s, hence $\Theta_s^{-1}(\Theta_s(y)) = y$ and therefore $S^{(s,t)} \circ S^{(s',t)}(x) = \Theta_t(y) = S^{(s',t)}(x)$.

Now define for the GREM the measures $\widetilde{\mu}^s_{\beta,N}$ indexed by the parameter $s \in [0,1]$ which are coarse-grained versions of $\tilde{\mu}_{\beta,N}$: $\tilde{\mu}^s_{\beta,N} = \sum_{\sigma \in \mathcal{S}_N} \mu_{\beta,N}(\sigma)\delta_{r_{[sN]}(\sigma)}$. Their probability distribution functions $\widetilde{\theta}_{\beta,N,s} = \mu_{\beta,N}(\sigma' : r_{[sN]}(\sigma') \le x)$ jump at points $r_{[sN]}(\sigma)$ and form a refining family depending on the parameter $s \in [0,1]$. Then we may define compatible left-continuous step functions (3.15) $\theta_{\beta,N,t}(\theta^{-1}_{\beta,N,s})$ which equal precisely the functions (3.14).

4. Genealogy of a continuous state branching process.

Another example of flows of probability measures satisfying Definition 2.1 arises in the context of continuous state branching process [BeLG]. The basic object here is a continuous state branching process $X(t)$ on $\mathbb{R}^+$ characterized by its Laplace exponent $u_t(\lambda)$. The process started in $a \ge 0$ will be denoted by $X(\cdot, a)$. This can be extended to a genuine two parameter process $(X(t,a), t, a \ge 0)$ using the fundamental branching property that states that if $X'(\cdot, b)$ and $X(\cdot, a)$ are independent copies, then $X(\cdot, a+b)$ has the same law as $X'(\cdot, b) + X(\cdot, a)$. The process $X(t,a)$ is characterized by the property that for any $a, b \ge 0$, $X(\cdot, a+b) - X(\cdot, a)$ is independent of the processes $X(\cdot, c)$, for all $c \le a$, and its law is the same as that of $X(\cdot, b)$. The right continuous version of $X(t, \cdot)$ is a subordinator. Bertoin and Le Gall [BeLG] prove the following proposition, based on the Markov property of this process.

Proposition 4.1: *On some probability space there exists a process $(\widetilde{S}^{(s,t)}(a), 0 \le s \le t, a \ge 0)$, such that*

(i) For any $0 \le s \le t$, $\widetilde{S}^{(s,t)}$ is a subordinator with Laplace exponent $u_{t-s}(\lambda)$.

(ii) For any integer $p \ge 3$ and $0 \le t_1 \le t_2 \le \dots \le t_p$, the subordinators $\widetilde{S}^{(t_1,t_2)}, \widetilde{S}^{(t_2,t_3)}, \dots, \widetilde{S}^{(t_{p-1},t_p)}$ are independent, and

$$\widetilde{S}^{(t_1,t_p)}(a) = \widetilde{S}^{(t_{p-1},t_p)} \circ \widetilde{S}^{(t_{p-1},t_p)} \circ \dots \circ \widetilde{S}^{(t_2,t_3)} \circ \widetilde{S}^{(t_1,t_2)}(a), \quad \forall a \ge 0, \quad a.s. \tag{4.1}$$

(iii) The processes $\widetilde{S}^{(0,t)}(a)$ and $X(t,a)$ have the same finite dimensional marginals.

The process $\widetilde{S}^{(s,t)}$ allows to construct a flow of probability distribution functions by setting

$$S^{(s,t)}(x) \equiv \frac{1}{X(t,1)} \widetilde{S}^{(s,t)}(X(s,1)x), \quad 0 \le s \le t \le 1. \tag{4.2}$$

Given I any countable subset of $\mathbb{R}^+$, they verify the assumptions of Definition 2.1 a.s.

We are interested in a particular case of Neveu's continuous state branching process X_t with

$$E(e^{-\lambda X_t} \mid X_0 = a) = e^{-u_t(\lambda)a}, \quad u_t(\lambda) = \lambda^{e^{-t}}. \tag{4.3}$$

In this case $\widetilde{S}^{(s,t)}$ are stable subordinators with index e^{s-t}. Then the normalized stable subordinators $S^{(s,t)}$ of (4.2) is a family of random probability distribution functions verifying Definition 2.1. Thus the genealogical construction of Section 2 applies to them.

Finally, note that if we take an increasing function $t(y) \geq 0$ for $y \in [0,1]$, then we may consider the time-changed flow $\bar{S}^{(y,z)} = S^{(t(y),t(z))}$, $0 \leq y \leq z$, satisfying again Definition 2.1 and therefore allowing the genealogical construction of Section 2.

Bertoin and Le Gall [BeLG] showed that the coalescent process on the integers induced by $S^{(s,t)}$ of (4.2) associated to Neveu's process (4.3) coincides with the coalescent process constructed by Bolthausen and Sznitman [BoS]. They also proved the following remarkable result connecting the collection of subordinators to Ruelle's Generalized Random Energy Model. Let us state this result for our convenience. Take the parameters $0 < x_1 < \cdots < x_p < 1$ and $0 < t_1 < \cdots < t_p$ linked by the identities

$$t_k = \ln x_{k+1} - \ln x_1 \tag{4.4}$$

for $k = 0, \ldots, p-1$, and $t_p = -\ln x_1$. Then the law of the family of jumps of the normalized subordinators $S^{(t_k,t_p)}$, for $k = 0, \ldots, p-1$, is the same as the law of Ruelle's probability cascades with parameters x_i, $i = 1, \ldots, p$.

Now consider a GREM with finitely many hierarchies and parameters such that the points $y_0 = 0$ and $0 < y_1 < \ldots < y_p \leq 1$ are the extremal points of the concave hull of A. Let us remind that $\lim_{N\to\infty} \mathbb{E} f_{\beta,N}(y) = \mathbb{E} f_\beta(y)$ can be computed by (1.8) for any $y \in [0,1]$. Now set

$$\mathbb{E} f_\beta(y_{i-1}) = x_i, \quad i = 1, \ldots, p, \tag{4.5}$$

where all of the $x_i < 1$. In Theorem 1.9 of [BK1] we proved that the point process

$$\sum_\sigma \delta_{\{\mu_{\beta,N}(\sigma' : d_N(\sigma,\sigma') > y_1), \ldots, \mu_{\beta,N}(\sigma' : d_N(\sigma,\sigma') > y_p)\}} \tag{4.6}$$

in $[0,1]^p$ converge to Ruelle's probability cascades with parameters x_i, $i = 1, \ldots, p$. (The convergence of the marginals of the process (4.6) for the GREM under the assumption that for any given hierarchy $i = 1, \ldots, p$ and $N > 0$ the number of configurations $\{\sigma' : d_N(\sigma,\sigma') > y_i\}$ is the same for all $\sigma \in \Sigma_N$, has been also established in Proposition 9.6 of [BoS2].)

Combining these two results yields

Proposition 4.2: *Let $\mu_{\beta,N}$ be the Gibbs measure associated to a GREM with finitely many hierarchies satisfying (4.5) at the extremal points y_i, $i = 1, \dots, p$ of the concave hull of the function A. Then the family of distribution functions $S_{\beta,N}^{(y_k,y_p)}$, $k = 1, 2, \dots, p$ defined according to (3.4) converges in law, and the limit has the same distribution as the family of normalized stable subordinators (4.2) $S^{(t_k,t_p)}$, $k = 0, 1, \dots, p-1$ in the sense that the joint distribution of their jumps has the same law, provided t_k are chosen according to (4.4), (4.5).*

5. Main result.

From the preceding proposition we expect that Neveu's process will provide the universal limit for all of our CREMs. The dependence on the particular model (i.e. the function A) and on the temperature must come from a rescaling of time. Set

$$x(y) \equiv \mathbb{E} f_\beta(y) = \begin{cases} \frac{\sqrt{2\ln 2}}{\beta\sqrt{\bar{a}(t)}}, & \text{if } t < t_\beta \\ 1, & \text{if } t \geq t_\beta \end{cases} \tag{5.1}$$

where $\bar{a}$ is the right-derivative of the convex hull of the function A, $t_\beta = \sup(t : \frac{\sqrt{2\ln 2}}{\beta\sqrt{\bar{a}(t)}} < 1)$ (here $\mathbb{E} f_\beta(y)$ is defined by the function A through (1.8)). Set also

$$T = -\ln x(0), \quad t(y) = T + \ln x(y). \tag{5.2}$$

Define the flow of probability distribution functions

$$\bar{S}^{(y,z)}(x) \equiv S^{(t(y),t(z))}(x) \tag{5.3}$$

where $S^{(s,t)}$ is the flow of functions (4.2) associated to Neveu's process (4.3). Let $\bar{\mathcal{K}}_T^{t(y)}$ be the genealogical map (2.11) associated to this flow.

Theorem 5.1: *Consider Continuous Random Energy Model with general function A such that A does not touch its convex hull $\bar{A}$ in the interior of any interval where $\bar{A}$ is linear. Then*

$$\mathcal{K}_{\beta,N} = K_1^{\beta,N} \stackrel{\mathcal{D}}{\to} \bar{\mathcal{K}}_1^{t(y)}. \tag{5.4}$$

Here $\mathcal{K}_{\beta,N}$ is the empirical distance distribution function (1.4), $K_1^{\beta,N}$ is the genealogical map (2.11) of the flow of probability distribution functions (3.4) and the equality $\mathcal{K}_{\beta,N} = K_1^{\beta,N}$ holds by Lemma 3.2. Theorem 5.1 is the main result of this paper. It expresses the geometry of the limiting Gibbs measure contained in $\mathcal{K}_{\beta,N}$ in terms of the genealogy of Neveu's branching process via the deterministic time change (5.2).

6. Coalescence and Ghirlanda-Guerra identities.

In this section we prove Theorem 5.1. As it was remarked in Section 2, K_T associated with a flow of measures is completely determined by its moments (2.18) which can be expressed via genealogical distance distributions of the corresponding coalescent (2.23). So, we will prove that the moments of $\mathcal{K}_{\beta,N}$, which are the n-replica distance distributions in our spin glass model (1.10), converge to the genealogical distance distributions on the integers (2.23) constructed from the flow of compatible measures with distribution functions $\bar{S}^{(y,z)}$ (5.3). But the flow $\bar{S}^{(y,z)}$ is the time changed flow (4.2) of Neveu's branching process (4.3) that by [BeLG] corresponds to the coalescent of Bolthausen-Sznitman. Therefore, its genealogical distance distributions on the integers are those of Bolthausen-Sznitman coalescent under this time change (5.2). Then the proof of Theorem 5.1 is reduced to the following Theorem 6.1 that gives in addition the connection between the n-replica distance distribution function of the CREM with the genealogical distance distribution function of the Bolthausen-Sznitman coalescent.

Theorem 6.1: *Under the same assumptions as in Theorem 5.1, for any $n \in \mathbb{N}$,*

$$\begin{aligned} &\lim_{N\uparrow\infty} \mathbb{E}\mu_{\beta,N}^{\otimes n}\left(d_N(\sigma^1,\sigma^2) \le y_1,, \dots, d_N(\sigma^{n-1},\sigma^n) \le y_{n(n-1)/2}\right) \\ &= \mathbb{P}\left(\rho_T(1,2) \le t(y_1), \dots, \rho_T(n-1,n) \le t(y_{n(n-1)/2})\right) \end{aligned} \tag{6.1}$$

where $t(y)$ is defined in (5.2) via (5.1). The distance ρ_T is the distance on integers for the Bolthausen-Sznitman coalescent, induced through (2.14) by the genealogical distance γ_T of the flow of measures $S^{(s,t)}$ (4.2) of Neveu's branching process (4.3).

Proof: The fact that in Bolthausen-Sznitman coalescent $\mathbb{P}(\rho_T(1,2) \le t) = e^{t-T}$ and the convergence (1.8) imply the statement of the theorem for $n = 2$:

$$\mathbb{E}\mu_{\beta,N}^{\otimes 2}(d_N(\sigma,\sigma') \le y) \to x(y) = e^{t(y)-T} = \mathbb{P}(\rho_T(1,2) \le t(y)).$$

The proof of the theorem for $n > 2$, and in fact the entire identification of the limiting processes with objects constructed from Neveu's branching process, relies on the Ghirlanda-Guerra identities [GG] that were derived for the models considered here in [BK2]. We restate this result in a slightly modified form. Let us remind that the family of measures (1.10) $\mathbb{Q}_{\beta,N}^{(n)}$ is determined on the space $[0,1]^{n(n-1)/2}$ as $\mathbb{E}\mu_{N,\beta}^{\otimes n}(\underline{d_N} \in \cdot)$ where $\underline{d_N}$ denotes the vector of replica distances $d_N(\sigma^k,\sigma^l)$, $1 \le k < l \le n$. Denote by $\mathcal{B}_k$ the vector of the first $k(k-1)/2$ coordinates.

Theorem 6.2: [BK3] *The family of measures $\mathbb{Q}^{(n)}_{\beta,N}$ converge to limiting measures $\mathbb{Q}^{(n)}_{\beta}$ for all finite n, as $N \uparrow \infty$. Moreover, these measures are uniquely determined by the distance distribution functions $\mathbb{E} f_\beta(y) = x(y)$ (1.8). They satisfy, for any $y \in [0,1]$, $n \in \mathbb{N}$ and $k \leq n$,*

$$\mathbb{Q}^{(n+1)}_{\beta}\left(d(k,n+1) \leq y | \mathcal{B}_n\right) = \frac{x(y)}{n} + \frac{1}{n} \sum_{\substack{l \neq k}}^{n} \mathbb{Q}^{(n)}_{\beta}\left(d(k,l) \leq y | \mathcal{B}_n\right). \tag{6.2}$$

Let us recall that due to the ultrametric property of d_N, these identities determine the measures $\mathbb{Q}^{(n)}_{\beta}$ uniquely. Thus, we must show that the right-hand side of (6.1) satisfies, for $t < T$,

$$\mathbb{P}\left(\rho_T(k,n+1) \leq t \mid \mathcal{B}_n\right) = \frac{1}{n} e^{t-T} + \frac{1}{n} \sum_{l \leq n, l \neq k} \mathbb{P}\left(\rho_T(k,l) \leq t \mid \mathcal{B}_n\right) \tag{6.3}$$

that can be equivalently written as

$$\mathbb{P}\left(\rho_T(k,n+1) > t \mid \mathcal{B}_n\right) = \frac{|l \in \{1,\ldots,n\} : \rho_T(k,l) > t| - e^{t-T}}{n} \qquad (6.4).$$

There are *two* ways to verify that (6.3) holds for the Bolthausen-Sznitman coalescent.

The first one is to observe that relation (6.3) involves only the marginals of the coalescent at a finite set of times. By Theorem 5 of Bertoin-Le Gall [BeLG], these can be expressed in terms of Ruelle's probability cascades modulo the appropriate time change. Thus, by Theorem 1.9 of [BK1] these probabilities can be expressed as limits of a suitably constructed GREM (with finitely many hierarchies) for which the Ghirlanda-Guerra relations do hold by Proposition 1.12 of [BK1]. Thus (6.3) is satisfied. $\diamondsuit$

The second way is to verify directly that Ghirlanda-Guerra relations (6.4) hold for the Bolthausen-Sznitman coalescent. This is the subject of the next Section 7.

7. Ghirlanda-Guerra identities and Chinese restaurant processes

Let us first give the following definition. Given the sequence of normalized jumps of the stable subordinator (Δ_i/T) with index x and given $U_1, U_2, \ldots$ independent uniform random variables on [0, 1], the partition of positive integers Π distributed as a partition of blocks of indices of U_i belonging to the same intervals $\Delta_i/T \in [0,1]$ is called $(x,0)$-*partition*, see [P].

Let us introduce an operation of coagulation on partitions, see [Pi1]: for a partition $\pi = (A_1, A_2, \ldots,)$ and $\Pi = (B_1, B_2, \ldots)$, the Π-coagulation of π consists of blocks of the form $\bigcup_{j \in B_i} A_j$.

By [BS] *the Markov kernels* $(e^{-t},0)$*-coagulation,* $t\geq 0$*, on partitions of* $\mathbb{N}$ *form a semi-group. The Markov process*

$$\mathbb{P}^{\pi}(\Pi(t+)\in\cdot)=(e^{t-T},0)-\textit{coagulation of}\ \pi \tag{7.1}$$

is distributed as the Bolthausen-Sznitman coalescent. It starts from a partition of singletons at time T and finishes by a partition of one block $\mathbb{N}$ at time $-\infty$. (The semi-group property can be also seen from the fact that the limiting frequencies of $(e^{-t},0)$-partitions are distributed as normalized jumps of stable subordinators and from their matching condition (4.1).)

Next, consider exchangeable random partitions Π on $\mathbb{N}$, introduced by J. Pitman under the name of Chinese restaurant processes. For each parameter $0<x<1$ this partition can be constructed as follows. Let Π_n denote the restriction of Π to the first n positive integers. Then, conditionally given $\Pi_n=\{A_1,\dots,A_k\}$ for any particular partition of $\{1,2,\dots,n\}$ into k subsets (tables) A_i of sizes n_i, $i=1,\dots,k$, the partition Π_{n+1} is an extension of Π_n such that the number $n+1$ (new customer) is attached to the class (table) A_i with probability $(n_i-x)/n$, and forms a new class (sits at a new table) with probability kx/n. Let us denote by $p(n_1,\dots,n_k)$ the probability of partitions Π with Π_n a particular partition of k classes of sizes $n_1,\dots,n_k$ respectively. Then

$$p(n_1+1,n_2,\dots,n_k)=\frac{n_1-x}{n}p(n_1,\dots,n_k) \tag{7.2}.$$

The crucial fact it that *the partition* Π *of the Chinese restaurant process with parameter* x *is* $(x,0)-$ *partition.* This fact noticed in [P] follows from the combination of the results of [Pi1] and [PPY]. To see this, it should be said first that Π is a partially exchangeable random partition in the sense of [Pi1]. Then, given the sequence of its a.s. limiting relative frequencies of classes P_i *in order of appearance*, the conditional distribution of Π given the whole sequence (P_i) is as follows: for each n conditionally given P_i and $\Pi_n=\{A_1,\dots,A_k\}$ where A_i are in order of appearance, Π_{n+1} is an extension of Π_n in which $n+1$ attaches to class A_i with probability P_i, $1\leq i\leq k$ and forms a new class with probability $1-\sum_{i=1}^k P_i$. In other words

$$p(n_1,...,n_k)=\mathbb{E}\Big[\prod_{i=1}^k P_i^{n_i-1}\prod_{i=1}^{k-1}\Big(1-\sum_{j=1}^i P_j\Big)\Big] \tag{7.3}$$

In the case of the Chinese restaurant process

$$p(n_1,...,n_k)=\frac{x\times 2x\times\cdots\times kx}{n!}\prod_{i=1}^k(1-x)(2-x)\cdots(n_i-x) \tag{7.4}$$

The function $p(n_1, ..., n_k)$ being symmetric, Π is an exchangeable random partition according to [Pi1]. Furthermore, again due to [Pi1], computing the moments from (7.3) and (7.4) one checks that the limiting relative frequencies in order of appearance in the Chinese restaurant process are $P_i = (1-W_1)(1-W_2)\cdots(1-W_{i-1})W_i$ with W_i independent beta $(1-x, ix)$. From the other point of view it has been shown in [PPY] that if $T = \sum_i \Delta_i$ has a stable distribution with index x, with $\Delta_1 > \Delta_2 > \dots$ being points of the Poisson point process on $(0, \infty)$, then the sequence $\Delta_{(i)}/T$ in size-biased order (this means, that given the whole sequence Δ_i/T, and U_i independent random variables uniform on $[0,1]$, then $U_1 \in \Delta_{(1)}/T$, $U_{\min\{j: U_j \notin \Delta_{(1)}/T\}} \in \Delta_{(2)}/T$ etc) has the same distribution of products of independent beta random variables. It follows, that the limiting frequencies of the Chinese restaurant process Π ranked by size are distributed as Δ_i/T, i.e. they have Poisson-Dirichlet distribution with parameter x. Hence, the Chinese restaurant process is $(x,0)-$ partition.

Thus, by (7.1) the marginals of Bolthausen-Sznitman coalescent $\Pi(t)$ at times $0 = t_0 < t_1 < \cdots < t_{p-1} < t_p = T$ can be constructed as the following sequence of Chinese restaurant processes. Let $x_i = e^{t_{i-1}-t_p}$, $0 < x_1 < x_2 < \cdots < x_p < 1$. Then $\Pi(t_{p-1}+)$ is distributed as $(x_p, 0)-$ partition, i.e. as the Chinese restaurant process with parameter x_p. Next, we define the partition $\Pi(t_{p-2}+)$ as the Chinese restaurant process on the classes of partition $\Pi(t_{p-1}+)$ with parameter $x_{p-1}/x_p = e^{t_{p-2}-t_{p-1}}$: this means that given already the classes $A_1^{p-1}, \dots, A_k^{p-1}$ obtained from $A_1^p, \dots, A_l^p$, where A_i^{p-1} consists of l_i blocks of Π^p, $i = 1, \dots, k$, $l_1 + \cdots + l_k = l$, the block A_{l+1}^p joins A_i^{p-1} with probability $(l_i^{p-1} - x_{p-1}/x_p)/l$ and forms a new class with probability $kx_{p-1}/(x_p l)$. One iterates this procedure with parameters $x_{p-2}/x_{p-1}, \dots, x_1/x_2$ to construct the partitions $\Pi(t_{p-3}+), \dots, \Pi(t_0+)$. By the semi-group property of $(e^{-t}, 0)-$ coagulations, $\Pi(t_i+)$ is distributed as a Chinese restaurant process with parameter $x_{i+1} = e^{t_i - t_p}$ for all $i = 0, 1, \dots, p-1$ verifying (7.2). Now (6.4) is immediate from the Chinese restaurant property (7.2).

Recently Ph. Marchal found another beautiful way to identify the Chinese restaurant process with $(x,0)$-partitions and also the iterated Chinese restaurant process with the Bolthausen-Sznitman coalescent, see [M].

References

[AC] M. Aizenman and P. Contucci, On the stability of the quenched state in mean field spin-glass models, J. Statist. Phys. **92**, 765-783 (1998).

[ASS] M. Aizenman M, R. Sims, S.L. Starr, (2003) An Extended Variational Principle for the SK Spin-Glass Model. preprint, `cond-mat/0306386` (2003).

[Be1] J. Bertoin, Homogeneous fragmentation processes, Probab. Theory Related Fields **121**, 301–318 (2001).

[Be2] J. Bertoin, Self-similar fragmentations, Ann. Inst. H. Poincaré (Probab. Statist.) **38**, 319–340 (2002).

[BeLG] J. Bertoin and J.F. Le Gall, The Bolthausen-Sznitman coalescent and the genealogy of continuous state branching processes, Probab. Theor. Tel. Fields **117**, 249-266 (2000).

[BePi] J. Bertoin and J. Pitman, Two coalescents derived from the ranges of stable subordinators, Electron. J. Probab. **5**, no. 7 (2000).

[BeYo] J. Bertoin and M. Yor, On subordinators, self-similar Markov processes and some factorizations of the exponential variable, Electron. Comm. Probab. **6**, 95–106 (2001)

[BoS] E. Bolthausen and A.-S. Sznitman, On Ruelle's probability cascades and an abstract cavity method, Commun. Math. Phys. **107**, 247-276 (1998).

[BoS2] E. Bolthausen and A.-S. Sznitman, *Ten Lectures on Random Media* Birkhäuser Verlag, Basel-Boston-Berlin (2002).

[BK1] A. Bovier and I. Kurkova, Derrida's Generalized random Energy models 1. Models with finitely many hierarchies. *Annals de l'Institut Henri Poincaré* **40** (4), 439–480 (2004).

[BK2] A. Bovier and I. Kurkova, Derrida's Generalized Random Energy models 2. Models with continuous hierarchies. *Annals de l'Institut Henri Poincaré* **40** (4), 481 –495 (2004).

[BKL] A. Bovier, I. Kurkova, and M. Löwe, Fluctuations of the free energy in the REM and the p-spin SK-models, Ann. Probab. **30**, 605-651 (2002).

[B] A. Bovier, Statistical mechanics of disordered systems, MaPhySto Lecture Notes 10, Aarhus, 2001.

[CCP] D. Capocaccia, M. Cassandro, and P. Picco, On the existence of thermodynamics for the Generalized Random Energy Model. J. Stat. Physics **46**, 493–505 (1987).

[D] B. Derrida, A generalization of the random energy model that includes correlations between the energies, J. Physique. Lett. **46**, 401–407 (1985).

[DG1] B. Derrida and E. Gardner, Solution of the generalized random energy model, J. Phys. C **19**, 2253–2274 (1986).

[DG2] B. Derrida and E. Gardner, Magnetic properties and function $q(x)$ of the generalized random energy model, J. Phys. C **19**, 5783-5798 (1986)

[GG] S. Ghirlanda, F. Guerra. General properties of the overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity, J. Phys. A **31**, 9144-9155 (1998).

[G03] F. Guerra, Broken replica symmetry bounds in the mean field spin glass model, Comm. Math. Phys. **233**, 1-12 (2003).

[GT] F. Guerra and F. Toninelli, The thermodynamics limit in mean field spin glass models, Commun. Math. Phys. **230**, 71-79 (2002).

[GMP] A. Galvez, S. Martinez, and P. Picco, Fluctuations in Derrida's random energy and generalized random energy models, J. Stat. Phys. **54**, 515-529 (1989).

[Le] M. Ledoux, On the distribution of overlaps in the Sherrington-Kirkpatrick spin glass model, J. Statist. Phys. **100** (2000), 871-892.

[M] Ph. Marchal Regenerative sets, random partitions and the Bolthausen-Sznitman coalescent, Preprint 2003.

[MPV] M. Mézard, G. Parisi, M.A. Virasoro, Spin-glass theory and beyond. World Scientific, Singapore (1988).

[Ne] J. Neveu, A continuous state branching process in relation with the GREM model of spin glass theory, Rapport interne no. 267, École Polytechnique

[Ru] D. Ruelle, A mathematical reformulation of Derrida's REM and GREM. Commun. Math. Phys. **108**, 225-239 (1987).

[P] J. Pitman, Combinatorial stochastic processes, Lecture Notes of the 2002 École des Probabilités de St. Flour; to appear as Springer Lecture Notes in Mathematics.

[Pi1] J. Pitman, Exchangeable and partially exchangeable random partitions, Probab. Theor. Relat. Fields **102**, 145–158 (1995).

[PPY] M. Perman, J. Pitman, M. Yor, Size-biased sampling of Poisson point processes and excursions, Probab. Theor. Relat. Fields **92**, 21–39 (1992).

[SK] D. Sherrington and S. Kirkpatrick, Solvable model of a spin glass, Phys. Rev. Lett. **35**, 1792-1796 (1972).

[T1] M. Talagrand, Rigorous low temperature results for mean field p-spin interaction models, Probab. Theor. Rel. Fields. **117**, 303-360 (2000).

[T2] M. Talagrand, Self organization in the low-temperature region of a spin glass model, Rev. Math. Phys. **15**, 1-78 (2003).

[T3] M. Talagrand, Mean field models for spin glasses: a first course, Lecture Notes of the 2000 École des Probabilités de St. Flour; to appear as Springer Lecture Notes in Mathematics.

[T4] M. Talagrand, Spin Glasses: a Challenge to Mathematicians, Ergebnisse der Mathematik und ihrer Grenzgebiete Vol. 46, Springer, Berlin-Heidelberg-New York, 2003.

[T5] M. Talagrand, The generalized Parisi formula. C. R. Math. Acad. Sci. Paris **337**, 111-114 (2003).