Statistics 291: Lecture 20 (April 04, 2024)

Towards the Spherical Parisi Formula: Ruelle Probability Cascades

Instructor: Mark Sellke

Scribe: Kevin Luo

Ruelle probability cascades were instrumental in the proof of the Parisi formula, as they give a description of the Gibbs measure, though this fact we will not fully prove. They will furthermore be needed for our interpolation.

1 Preliminaries

We are interested in understanding a heuristic for the fine-grained structure on 1-RSB Gibbs measures. Here, where the temperature is low, we expect to have, as in our discussion of shattering, a collection of clusters $\mathscr{C}_1, \mathscr{C}_2, \ldots$; at the center of each is an extremal critical point, as illustrated in Figure 1. We will

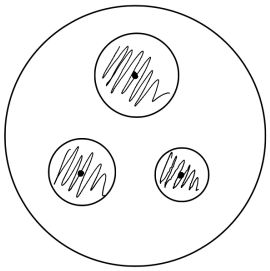

Figure 1: Structure on a 1-RSB Gibbs measure.

take these to be ordered, so that $\mu_\beta(\mathscr{C}_1) \geq \mu_\beta(\mathscr{C}_2) \geq \cdots$. The key distinction between the 1-RSB phase and the shattered phase is that, since the temperature is even lower, rather than each cluster having an exponentially small fraction of the mass, now, heuristically, the top clusters occupy a constant fraction of the mass. One can think that this means

$$\sum_{i=1}^{100} \mu_\beta(\mathscr{C}_i) \geq \frac{1}{2} \tag{1}$$

with probability 99%; more formally,

$$\lim_{N\to\infty} \sum_{i=1}^{k} \mu_\beta(\mathscr{C}_i) \geq 1-\epsilon \tag{2}$$

with probability $1-\epsilon$ for $k \geq k_0(\epsilon)$.

Our goal today will be to understand the joint distribution of the weights of the top clusters, which is a much more detailed question than just understanding the total free energy on the model. We will do this by identifying each cluster with the energy of the critical point at its center. Our working models for today will be spherical pure p-spin models (which indeed turn out to be 1-RSB).

2 The Top Critical Points and Clusters

2.1 Poisson Point Process Approximation

Let the top critical points be $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2, \ldots$, so that $H_N(\boldsymbol{\sigma}^1) \geq H_N(\boldsymbol{\sigma}^2) \geq \cdots$. We know through Kac-Rice that the "intensity" of these points at various levels of the energy behaves as the curve in Figure 2.

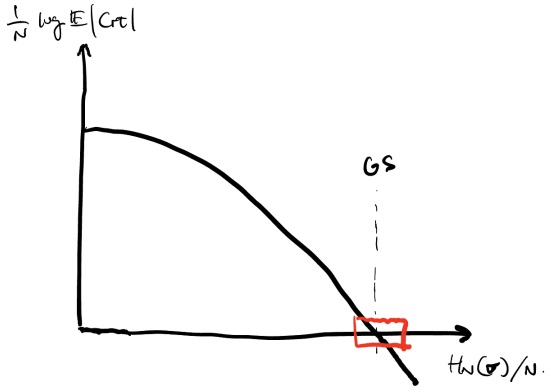

Figure 2: Curve of annealed counts given by Kac-Rice

There are a lot of critical points at low energy, and the number diminishes until we reach the cross-over point, which in this case further turns out to be the ground state (recall that before we had only argued it was an upper bound).

If we try to zoom in very closely near the cross-over point, because the average number of critical points diminishes exponentially fast according to this curve, in the limit, we expect the energy of the realized critical points in this region to have intensity measure something like $e^{-x}\,\mathrm{d}x$, for $x = H_N(\boldsymbol{\sigma}) - E_0$, where E_0 is some scaling shift (and we are setting an arbitrary constant factor in the exponent to 1). Moreover it is natural to hope for a Poisson point process in the limit. One can imagine setting E_0 so that, with probability at least $1/2$, the maximum critical point is below this level. Note that we are not considering the normalized energy $H_N(\cdot)/N$, but actually the energies themselves – this corresponds to looking in an $O(1)$ size window on the graph above. These are shown in Figure 3.

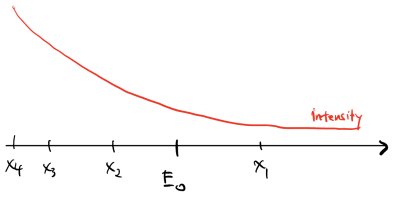

Figure 3: Heuristic distribution of maximum points

Definition 2.1 (Poisson Point Process)**.** A Poisson point process with intensity measure $\Lambda\,\mathrm{d}x$ is the unique point process on $\mathbb{R}$ such that, the number of points in $[a,b]$ is distributed as $\mathrm{Pois}(\int_a^b e^{-x}\,\mathrm{d}x)$ and disjoint

intervals are independent.

For the rest of this lecture, we will assume that the approximation above holds. Furthermore, we will guess that

$$\mu_\beta(\mathscr{C}_i) \propto e^{\beta H_N(\boldsymbol{\sigma}_i)} = e^{\beta x_i}. \tag{3}$$

That is, the mass on each cluster is proportional to the exponential of the energy at its central critical point. Let $u_i = e^{\beta H_N(\boldsymbol{\sigma}_i)} = e^{\beta x_i} = e^{x_i/m}$; from here on, we will write things in terms of the temperature m rather than β. Then

$$\mu_\beta(\mathscr{C}_i) = v_i := \frac{u_i}{\sum u_i}. \tag{4}$$

Under what conditions will the normalizing sum in the denominator be finite? The correct condition is that $m < 1$. The reasoning for this is that for any fixed energy level x, there should be something like $O(e^{-x})$ points of that energy level (using the PPP approximation above), so there are a lot of such points as the energy becomes small. In order for this sum to converge, we need to the contribution to the mass of each of these terms to be small, and this is satisfied when $e^{x/m} \cdot e^{-x} < 1$ for large, negative x, which exactly occurs when $m < 1$.

Now, if one pushes through this analysis and performs the chain rule, then the u_i can be found to also follow a Poisson process, where $(u_1, u_2, \dots) \sim \mathrm{PPP}(C \cdot u^{-1-m}\, \mathrm{d}u)$. Note that this Poisson process is singular around 0, so there will be an infinite number of points. Furthermore, although their expected sum $\int_0^\infty u^{-m}\, \mathrm{d}u$ is infinite due to the tail for large u, there will always be a finite number of points past any nonzero threshold, so we are fine, and the bulk of points near 0 have finite contribution to the sum.

2.2 Useful Identities

Consider now $x_i + \ell_i - \log\mathbb{E}[e^\ell]$, for i.i.d. distributed ℓ_i with finite exponential moment. Consider sorting these quantities into $\tilde{x}_1 \geq \tilde{x}_2 \geq \cdots$, so that $\tilde{x}_{\pi(i)} = x_i + \ell_i - \log\mathbb{E}[e^\ell]$, where π denotes the sorting permutation.

Proposition 2.2. *With $(x_1, x_2, \dots)$ and $(\tilde{x}_1, \tilde{x}_2, \dots)$ defined above, one has the distributional equality*

$$(x_1, x_2, \dots) \overset{d}{=} (\tilde{x}_1, \tilde{x}_2, \dots). \tag{5}$$

Before beginning the proof, we give a little intuition for this procedure. One should think of the ℓ_i as being a Gaussian perturbation, where this process consists of partially renoising the Hamiltonian, i.e.

$$H_N \to \sqrt{1 - t/N} H_N + \sqrt{t/N} \tilde{H}_N \tag{6}$$

for $t = O(1)$.

By applying this small amount of noise, we expect the clusters themselves to stay intact, but for their energies to change a little bit. One can view this noising procedure as a Markov process on the x_i's, which we view as corresponding to the cluster masses $\mu_\beta(\mathscr{C}_i)$. Since the marginal law of the entire spin glass remains the same, we should expect that the behavior of the clusters, once reordered, remains the same.

Proof of Proposition 2.2. The intensity of $x_i + \ell_i - \log\mathbb{E}[e^\ell]$ at $y \in \mathbb{R}$ is

$$\int_{\mathbb{R}} e^{-y+\ell-\log\mathbb{E}[e^\ell]}\, \mathrm{d}p(\ell) = e^{-y}, \tag{7}$$

which is indeed the same intensity measure as for the x_i, where $p(\ell)$ corresponds to the law of ℓ. In essence, above, we integrate over the intensity of a corresponding x, coupled with the density of ℓ such that the final sum is y. □

Since we defined v_i the normalized cluster weights, and each cluster weight is proportional to the energy at its critical point, this further implies

$$(v_1, v_2, \ldots) \stackrel{d}{=} (\tilde{v}_1, \tilde{v}_2, \ldots) \tag{8}$$

where

$$\tilde{v}_i = \frac{\tilde{u}_i}{\sum \tilde{u}_j}, \text{ where } \tilde{u}_j = e^{\tilde{x}_j/m}. \tag{9}$$

Corollary 2.3. *Let $L_i = e^{\ell_i/m}$. Then*

$$(u_1, u_2, \ldots) \stackrel{d}{=} (\tilde{u}_1, \tilde{u}_2, \ldots) \tag{10}$$

where

$$\widetilde{u}_{\pi(i)} = \frac{u_i L_i}{\mathbb{E}[L^m]^{1/m}}, \tag{11}$$

where again $\pi(i)$ is just the sorting permutation.

The above follows from exponentiating x_i/m, which is exactly the definition of u_i.

Lemma 2.4. *For i.i.d. L_i,*

$$\mathbb{E}\log\sum v_i L_i = \log(\mathbb{E}[L^m]^{1/m}) = \frac{1}{m}\log\mathbb{E}[L^m]. \tag{12}$$

Proof.

$$\mathbb{E}\log\frac{\sum u_i L_i}{\sum u_i} \stackrel{(a)}{=} \mathbb{E}\log\frac{\sum u_i L_i}{\sum \tilde{u}_i} \stackrel{(b)}{=} \log(\mathbb{E}[L^m]^{1/m}). \tag{13}$$

(a) corresponds to applying the distributional equality shown in Corollary 2.3, while (b) follows from the definition of $\tilde{u}_i$. □

To gain some intuition for this, it helps to think about what happens when $m = 1$, even though it is not truly allowed. When one sets this, the above expression reduces to $\log\mathbb{E}[L]$, which corresponds to an annealed approximation of $\mathbb{E}\log\sum v_i L_i$. This is what one would expect if all the weights v_i are extremely small, in which case the law of large numbers would give $\sum_{v_i} L_i \approx \mathbb{E}[L_i]$ with high probability. Indeed for $m = 1$, the Poisson point process for the u_i has intensity proportional to u^{-2}, which corresponds to there being even more points near zero, and less large clusters. For large m, indeed all of the clusters should be very small – there are no dominating weights, and hence one can essentially apply the law of large numbers as above.

2.3 Tangentially Related Facts

Given two i.i.d. samples σ^1, σ^2, drawn such that

$$\mathbb{P}[\sigma^1 = i] = \mathbb{P}[\sigma^2 = i] = v_i, \tag{14}$$

one can ask what $\mathbb{P}[\sigma^1 = \sigma^2]$ would be. The answer turns out to be $1 - m$. In essence, as the temperature goes down, the overlap goes to 1. Roughly speaking, if we expect all overlaps to be orthogonal, with centers at $\sqrt{qN}$, then the overlap distribution has a piecewise constant CDF, as shown in Figure 4. This is what is called a 1-RSB overlap distribution.

This is related to the Chinese Restaurant Process.

Definition 2.5 (Chinese Restaurant Process)**.** One has an infinite collection of infinitely large tables. People arrive at the restaurant 1 by 1. The first person goes to a random table. The second person now joins the first person with probability 1/2, or starts a new one with probability 1/2. The third person joins person 1 with probability 1/3, joins person 2 with probability 1/3, and starts their own table with probability 1/3, and so on.

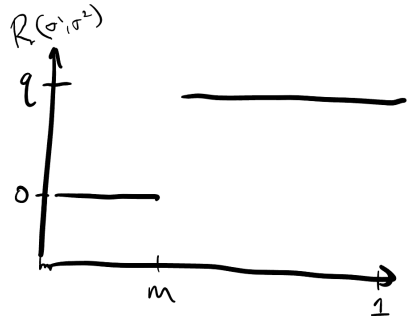

Figure 4: 1-RSB overlap distribution

This is a type of Pólya urn model. If one looks at the number of people at Table 1 versus the number not at table 1, then this is a Pólya urn model where the fraction at table 1 is a $v_1 = y_1 \sim \text{Unif}([0,1])$. Likewise, if one looks at the proportion of people at Table 2 versus the number at table numbers three or higher, this is an independent Pólya urn, where the proportion follows $v_2 = (1-y_1)y_2$, where $y_2 \sim \text{Unif}([0,1])$. This procedure continues, where $v_3 = (1-y_1)(1-y_2)y_3$, and so on. In these models, these y_i are drawn from $\text{Beta}(1,1)$ distributions. For our clusters, instead we have $y_1 \sim \text{Beta}(1-m,m)$, $y_2 \sim \text{Beta}(1-m,2m)$, and so on.

How can this be represented in terms of the restaurant story? If the current tables have sizes $s_1, s_2, \ldots, s_k$, and n people total, then the $n+1$st person joins table s_i with probability $\frac{s_i - m}{n}$; otherwise, they start a new table with probability km/n.

3 Ruelle Probability Cascades

We now want to perform a multi-level analog of our analysis in the prior section. This corresponds to analyzing a tree of finite depth r, but where each node has infinite nodes: Along each edge from the root,

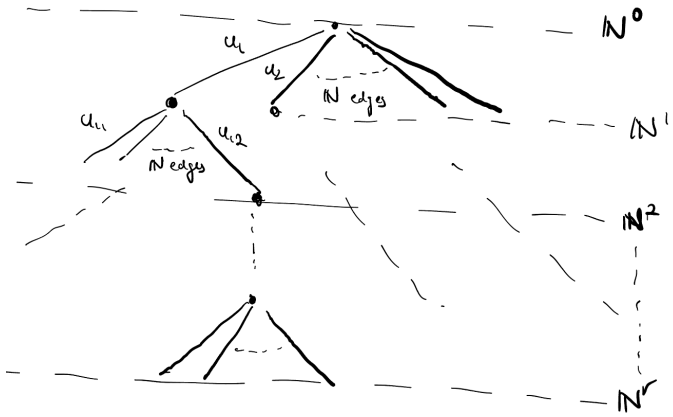

Figure 5: The tree corresponding to a Ruelle Probability Cascade.

we assign weights $u_1, u_2, \ldots$. Similarly, we assign weights $u_{11}, u_{12}, \ldots$ for the edges from the first node on

the 2nd level, and continue down the tree, similarly to how we labeled the edges last lecture. Now, fixing $\alpha \in \mathbb{N}^d$, where $\alpha = \alpha_1\alpha_2\cdots\alpha_d$, we define the weight $w_\alpha = u_{\alpha_1}u_{\alpha_2}u_{\alpha_1\alpha_2\alpha_3}\cdots u_{\alpha_1\cdots\alpha_r}$.

At the bottom level, for $\gamma \in \mathbb{N}^r$, we further define the weights

$$v_\gamma = \frac{w_\gamma}{\sum_{\alpha\in\mathbb{N}^r} w_\alpha}. \tag{15}$$

We generate $u_1, u_2, \cdots \sim \text{PPP}(u^{-1-m_0})$, and more generally, for $\alpha \in \mathbb{N}^d$, we generate $u_{\alpha 1}, u_{\alpha 2}, \cdots \sim \text{PPP}(u^{-1-m_d})$, where $m_0 < m_1 < \cdots < m_{r-1} < m_r = 1$.

Note that the above Poisson point processes are again not integrable around 0, so we once more have infinitely many points, but in that region, the points contribute finite weight, by the same analysis as before.

The idea here is again that v_γ should represent the weights of the clusters in an r-step RSB model.

As in the previous lecture, the tree structure naturally yields

$$\mathbb{P}[\alpha \wedge \gamma = d] = m_d - m_{d-1}. \tag{16}$$

We want to be able to do the renoising analysis we did in the prior section, but now we need to conduct it in such a way that the hierarchy of the tree is respected. Before, this was not so much an issue, since the tree only had one level.

Our approach will be to generate independent y_γ for each $\gamma \in \mathbb{N}^{\le r}$. For $\alpha = \alpha_1\alpha_2\cdots\alpha_r$ (so α is a leaf), now define

$$F_\alpha = F_r(y_{\alpha_1}, y_{\alpha_1\alpha_2}, y_{\alpha_1\alpha_2\alpha_3}, \cdots, y_{\alpha_1\cdots\alpha_r}). \tag{17}$$

We then have the following:

Theorem 3.1.

$$\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha F_\alpha = \log F_0, \tag{18}$$

where F_0 is defined by the following backwards recursion:

$$F_d(y^1, y^2, \cdots, y^d) = \mathbb{E}^{y^{d+1}}\left[F_{d+1}(y^1, \cdots, y^{d+1})^{m_{d-1}} \mid (y^1, \cdots, y^d)\right]^{1/m_{d-1}}. \tag{19}$$

As an example, when $r = 2$, this results in

$$F_0 = \mathbb{E}^{y^1}\left[\mathbb{E}^{y^2}\left[F_2(y^1, y^2)^{m_1} \mid y^1\right]^{m_0/m_1}\right]^{1/m_0}, \tag{20}$$

and by construction, one can see that the correlations essentially respect the hierarchy.

Proof. We show the proof in the case of $r = 2$. The rest are similar. First, we rewrite

$$\mathbb{E}\log\sum_{i,j} u_i u_{ij} F_2(y_i, y_{ij}) = \mathbb{E}\log\sum_i u_i\left(\sum_j u_{ij}F_2(y_i, y_{ij})\right) \tag{21}$$

Note that for each fixed i, the collection $\{F_2(y_i, y_{ij})\}_{j\ge1}$ above, conditioned on y_i, are in fact i.i.d. Hence we can apply Corollary 2.3, which provides that, for each i, conditional on y_i,

$$\sum_{ij} u_{ij}F_2(y_i, y_{ij}) \overset{d}{=} \sum_{ij}\tilde{u}_{ij}\mathbb{E}[F_2(y_i, y_{ij})^{m_1} \mid y_i]^{1/m_1}. \tag{22}$$

Hence we can rewrite the above as

$$\mathbb{E}\log\sum_i u_i\left(\sum_j u_{ij}F_2(y_i, y_{ij})\right) = \mathbb{E}\log\sum_i u_i\left[\left(\sum_j \tilde{u}_{ij}\right)\mathbb{E}\left[F_2(y_i, y_{ij})^{m_1} \mid y_i\right]^{1/m_1}\right]. \tag{23}$$

We are now in a position to apply Corollary 2.3 again, where the i.i.d. variable is now the bracketed term above. The corollary provides

$$\sum_i u_i \left[\left(\sum_j \tilde{u}_{ij} \right) \mathbb{E}\left[F_2(y_i, y_{ij})^{m_1} \mid y_i \right]^{1/m_1} \right] \stackrel{d}{=} \sum_i \tilde{u}_i \left(\sum_j \tilde{u}_{ij} \right) \mathbb{E}\left[\mathbb{E}\left[F_2(y_i, y_{ij})^{m_1} \mid y_i \right]^{m_0/m_1} \right]^{1/m_0}, \tag{24}$$

thus obtaining

$$\mathbb{E}\log \sum_i u_i \left[\left(\sum_j \tilde{u}_{ij} \right) \mathbb{E}\left[F_2(y_i, y_{ij})^{m_1} \mid y_i \right]^{1/m_1} \right] = \mathbb{E}\log \sum_{i,j} \tilde{u}_i \tilde{u}_{ij} \mathbb{E}\left[\mathbb{E}\left[F_2(y_i, y_{ij})^{m_1} \mid y_i \right]^{m_0/m_1} \right]^{1/m_0} \tag{25}$$

$$= \mathbb{E}\log \sum_{i,j} \tilde{u}_i \tilde{u}_{ij} \mathbb{E}\left[\mathbb{E}\left[F_2(y_1, y_2)^{m_1} \mid y_1 \right]^{m_0/m_1} \right]^{1/m_0}. \tag{26}$$

To conclude, note that

$$\mathbb{E}\log \sum_{i,j} v_{ij} F_2(y_i, y_{ij}) = \mathbb{E}\log \left(\frac{\sum_{ij} u_i u_{ij} F_2(y_i, y_{ij})}{\sum u_i u_{ij}} \right) \tag{27}$$

Proposition 2.2 once again provides $\sum_{i,j} u_i u_{ij} \stackrel{d}{=} \sum_{i,j} \tilde{u}_i \tilde{u}_{ij}$

$$= \mathbb{E}\log \left(\frac{\sum_{i,j} u_i u_{ij} F_2(y_i, y_{ij})}{\sum_{i,j} \tilde{u}_i \tilde{u}_{ij}} \right) \tag{28}$$

$$= \log F_0 \tag{29}$$

where this follows from (26) and the definition of F_0. □

4 Preview of the remainder

The interpolation scheme we will use is, for $\boldsymbol{\sigma} \in \mathscr{S}_N$ and $\alpha \in \mathbb{N}^r$,

$$H_{N,t}(\boldsymbol{\sigma}, \alpha) = \log v_\alpha + \left(\sin(t) \left(H_N(\boldsymbol{\sigma}) + g_\theta(\alpha) \right) + \cos(t) \left\langle G_{\xi'}(\alpha), \boldsymbol{\sigma} \right\rangle \right), \tag{30}$$

where

- $\theta(y) = y\xi''(y) - \xi(y)$,
- g_θ is a centered Gaussian process with covariance function $\mathbb{E} g_\theta(\alpha) g_\theta(\gamma) = \theta(q_{\alpha \wedge \gamma})$, for some increasing sequence $q_0 < q_1 < q_2 < \cdots < q_d$,
- $G_{\xi'}$ is similar but a spherical Gaussian process instead, i.e. $G_{\xi'}$ is a random vector in $\mathbb{R}^N$ with IID coordinates having the law of $g_{\xi'}$.

The $\log v_\alpha$ at the start just amounts to using the weights v_α as the "base" measure on $\mathbb{N}^r$, just like the uniform measure on $\mathscr{S}_N$ has been our base measure throughout the class.