Exchangeable random arrays

Tim Austin

Abstract

These notes were written to accompany a sequence of workshop lectures at the Indian Institute of Science, Bangalore in January 2013. My lectures had two purposes. The first was to introduce some of the main ideas of exchangeability theory at a level suitable for graduate students (assuming familiarity with measure-theoretic probability). The second was to show the key rôle exchangeability has played in recent work on mean-field spin glass models, particularly Panchenko's proof of a version of the Ultrametricity Conjecture. To my taste this may be the single most exciting reason to learn about exchangeability.

Of course, these dual purposes constrained the choice of material: I had to focus on those varieties of exchangeability that arise in spin glass theory. Also, the spin glass results were far too complicated to cover in depth, so in those lecture I omitted almost all detail, and the notes are the same. A much more complete treatment of similar material will appear in Panchenko's monograph [Panar]. These notes are certainly not intended to compete with that, but offer a more basic overview that may benefit the newcomer to the field.

Contents

Part I

Exchangeability Theory

1 Exchangeable sequences and arrays

Some notation

- $\mathbb{N} := \{1, 2, \ldots\}$;
- for $n \in \mathbb{N}$, $[n] := \{1, 2, \ldots, n\}$;
- $\mathrm{Sym}(n)$ is the group of all permutations of $[n]$ and $\mathrm{Sym}(\mathbb{N})$ is the group of all permutations of $\mathbb{N}$;
- for any set A and $k \geq 1$, $A^{(k)}$ is the set of subsets of A of size k, and $A^{(\leq k)} := \bigcup_{j=0}^{k} A^{(j)}$.
- if E is a standard Borel space[1], then $\Pr E$ is the space of all Borel probability measures on E;
- if E and E' are measurable spaces, μ is a probability measure on E, and $f : E \longrightarrow E'$ is measurable, then $f_*\mu$ denotes the pushforward measure:

$$f_*\mu(A) := \mu(f^{-1}(A)) \quad \forall A \subseteq_{\text{measurable}} E;$$

- a background probability space will be denoted by $(\Omega, \mathcal{F}, \mathsf{P})$, with expectation denoted E (although in keeping with convention among probabilists, Ω will usually be kept hidden in the background);
- if X and Y are r.v.s valued in the same space, then $X \stackrel{\mathrm{d}}{=} Y$ if they have the same distribution.

Objects of the theory

Exchangeability theory is concerned with families of random variables whose joint distribution is unchanged when they are permuted by some group of permutations.

[1]Meaning it is isomorphic, as a measure space, to a Borel subset of a complete, separable metric space with the Borel σ-algebra. This assumption is needed when working with conditional probabilities. It covers all 'nice' spaces that one meets in practice.

For particular groups of permutations, it seeks to describe all possible joint distributions that have this property. Of course, if the index set is a discrete group Γ, the random variables are indexed by Γ, and they are permuted by the regular representation, then this task encompasses the whole of ergodic theory. In that case a complete description is generally impossible. But in some other cases, in which the relevant group of permutations is 'very large' relative to the indexing set, one can deduce rather more complete results on the distributions of such families than in general ergodic theory.

Basic objects of the theory:

- **Exchangeable sequences:** Let E be $\{0,1\}$, $\mathbb{R}$, or any other standard Borel space. A sequence of E-valued random variables $(X_n)_{n\in\mathbb{N}}$ is **exchangeable** if
$$(X_n)_n \stackrel{\mathrm{d}}{=} (X_{\sigma(n)})_n \quad \forall\sigma \in \mathrm{Sym}(\mathbb{N}).$$
Note that this is really an assertion about the measure μ on $E^{\mathbb{N}}$ which is the joint law of the r.v.s (X_n): it is invariant under the action of $\mathrm{Sym}(\mathbb{N})$ on $E^{\mathbb{N}}$ by the permutation of coordinates. When $E = \{0,1\}$ these were studied by de Finetti in the 1930's; for more general E by Hewitt and Savage in the 1950's.

- **Exchangeable arrays:** More generally, for any $k \geq 1$ we can consider an array of E-valued r.v.s $(X_e)_{e\in\mathbb{N}^{(k)}}$ indexed by size-k subsets of $\mathbb{N}$, and say it is **exchangeable** if $(X_e)_e \stackrel{\mathrm{d}}{=} (X_{\sigma(e)})_e$ for any $\sigma \in \mathrm{Sym}(\mathbb{N})$, where if $e = \{n_1,\dots,n_k\}$ then $\sigma(e) := \{\sigma(n_1),\dots,\sigma(n_k)\}$. So now exchangeability is an assertion about the law μ on $E^{\mathbb{N}^{(k)}}$. Exchangeable sequences are the case $k = 1$. General arrays were studied by Hoover [Hoo79, Hoo82], Aldous [Ald81, Ald82, Ald85], Fremlin and Talagrand [FT85] and Kallenberg [Kal89, Kal92]. Important ideas were also suggested (though not published) by Kingman, who had previously studied random partitions with a similar symmetry and proved his related 'paintbox' theorem: see [Kin78b, Kin78a].

Recommended reading on basic exchangeability theory: the really essential reference is still [Ald85]. The recent textbook [Kal05] offers a more modern and definitive account (see Chapter 7 in particular). Three surveys with different emphases are [Aus08, DJ07, Ald]: the first two give special attention to the connection with limit objects for finite graphs.

Why are these important?

Exchangeable random structures are important because they are the natural output of sampling at random from discrete structures.

Example 1 Most simply, for any E and any probability measure ν on E, the product measure $\nu^{\otimes \mathbb{N}^{(k)}}$ is always the law of an E-valued exchangeable array.

Example 2 Suppose now that E is arbitrary and that $\{\nu_1, \dots, \nu_m\}$ is a finite set of probability measures on E. Choose an E-valued array $(X_e)_{e \in \mathbb{N}^{(k)}}$ as follows: first pick $\ell \in \{1, 2, \dots, m\}$ uniformly at random, and then conditionally on this choose $(X_e)_{e \in \mathbb{N}^{(k)}}$ i.i.d. from ν_ℓ. (This clearly agrees with the previous example when $m = 1$.)

Example 3 Let $E = \{0, 1\}$ and let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a finite graph, possibly with loops. Let $(V_n)_{n \in \mathbb{N}}$ be a random sequence of vertices sampled independently from the uniform distribution on $\mathcal{V}$, and now define $(X_e)_{e \in \mathbb{N}^{(2)}}$ by letting

$$X_{ij} := \begin{cases} 1 & \text{if } V_i V_j \in E \\ 0 & \text{else} \end{cases}$$

(so there is no additional randomness once the V_n have been chosen). Note that in general the notation ij stands for the unordered pair $\{i, j\}$.

One can easily find ways to generalize these examples (for instance, how could Example 3 be made to give a random array with a different space E?). We will soon introduce a broad framework for discussing these. Remarkably, it turns out that once a suitably general notion of 'sampling' has been defined, it is the *only* way one can produce an exchangeable sequence or array.

2 Statement of the Representation Theorem

In order to generalize Examples 1–3 above, one must first observe that to construct a k-set exchangeable random array, randomness can be introduced at any 'level' between 0 and k. In order to make this formal, consider the **uniform random array**: this is a family of random variables $(U_a)_{a \subseteq \mathbb{N},\, |a| \leq k}$ which are i.i.d. $\mathrm{U}[0, 1]$. Its law is simply the product of copies of Lebesgue measure on the space

$$[0, 1] \times [0, 1]^{\mathbb{N}} \times [0, 1]^{\mathbb{N}^{(2)}} \times \cdots \times [0, 1]^{\mathbb{N}^{(k)}}.$$

Now consider also any measurable function

$$f : [0,1] \times [0,1]^k \times [0,1]^{\binom{k}{2}} \times \cdots \times [0,1]^{\binom{k}{k-1}} \times [0,1] \longrightarrow E$$

which is symmetric under the action of the permutation group $\mathrm{Sym}(k)$:

$$f(x, (x_i)_i, (x_{ij})_{ij}, \ldots, x_{[k]}) = f(x, (x_{\sigma(i)})_i, (x_{\sigma(i)\sigma(j)})_{ij}, \ldots, x_{[k]}) \quad \forall \sigma \in \mathrm{Sym}(k).$$

Such a function will be referred to as **middle-symmetric**.

Then we may obtain an exchangeable random array as follows: let the U_a for $a \subset \mathbb{N}$, $|a| \leq k$ be uniform i.i.d. as above, and set

$$X_e := f\big(U_\emptyset, (U_i)_{i\in e}, (U_a)_{a\in e^{(2)}}, \ldots, (U_a)_{a\in e^{(k-1)}}, U_e\big) \quad \text{for } e \in \mathbb{N}^{(k)}.$$

This is well-defined owing to the middle-symmetry of f.

Definition 2.1 *This is the exchangeable random array* ***directed by*** f. *We denote its law by* $\mathrm{Samp}(f)$.

Let's revisit the previous examples:

Example 1 By abstract measure theory, any Borel probability measure ν on E is the pushforward of $\mathrm{U}[0,1]$ under some measurable $f_0 : [0,1] \longrightarrow E$: that is, $\mathrm{law}(f_0) = \nu$. So now let

$$f(x, (x_i)_i, (x_{ij})_{ij}, \ldots, x_{[k]}) = f_0(x_{[k]}).$$

This example is using only 'level-k randomness'.

Example 2 Building on the above, let $f_\ell : [0,1] \longrightarrow E$ be such that $\mathrm{law}(f_\ell) = \nu_\ell$ for $1 \leq \ell \leq m$, and also let $\mathcal{P} = (I_1, \ldots, I_m)$ be a partition of $[0,1]$ into m equal subintervals. Now let

$$f(x, (x_i)_i, (x_{ij})_{ij}, \ldots, x_{[k]}) = f_\ell(x_{[k]}) \quad \text{whenever } x \in I_\ell.$$

This example is using the randomness from levels 0 and k.

Example 3 Finally, let $\mathcal{P} = (I_v)_{v\in\mathcal{V}}$ be a partition of $[0,1]$ into $|\mathcal{V}|$-many equal subintervals, and now define

$$f(x, x_1, x_2, x_{12}) := \begin{cases} 1 & \text{if } (x_1, x_2) \in I_u \times I_v \text{ for some } uv \in E \\ 0 & \text{else.} \end{cases}$$

So this example (with $k = 2$) is using the randomness from level 1.

Theorem 2.2 (Representation Theorem for exchangeable arrays) *Any exchangeable array $(X_e)_{e\in\mathbb{N}^{(k)}}$ has law equal to* $\mathrm{Samp}(f)$ *for some f as above.*

This is due to de Finetti ($k = 1$, $E = \{0,1\}$), Hewitt and Savage ($k = 1$), Hoover and separately Aldous ($k = 2$, different proofs), and Kallenberg (all k). (Aldous partly attributes his proof to Kingman.)

We will prove the cases $k = 1$ and $k = 2$. The latter already contains the main difficulties, except that the general case requires an induction on k which needs some careful management. That will be left as an exercise, or see [Aus08].

3 A tool: the Noise-Outsourcing Lemma

The following soft fact from measure theory provides a valuable tool for simplifying and clarifying the structures we will examine. It is treated in many standard probability texts; for instance, a more general version is given as Theorem 6.10 in Kallenberg [Kal02].

Lemma 3.1 (Noise-Outsourcing lemma) *If X, Y are r.v.s taking values in standard Borel spaces S and T, then (possibly after enlarging the background probability space) there are a r.v. $U \sim \mathrm{U}[0,1]$ and a Borel function $f : S \times [0,1] \longrightarrow T$ such that U is independent from X and*

$$(X, Y) = (X, f(X, U)) \quad a.s..$$

$\square$

In case S is a one-point space, say $S = \{*\}$, and X is deterministic, this is just the assertion that any standard-Borel-valued r.v. has law equal to an image of $\mathrm{U}[0,1]$

4 Proof of de Finetti's Theorem

We will prove de Finetti's Theorem in this section, and the Aldous-Hoover Theorem in the next.

Thus, suppose that E is a standard Borel space and that $(X_i)_i$ is an exchangeable sequence of E-valued r.v.s. We must find a Borel function $g : [0,1]\times[0,1] \longrightarrow E$ such that

$$(X_i)_i \stackrel{\mathrm{d}}{=} (g(U,U_i))_i,$$

where U and U_i, $i \in \mathbb{N}$, are i.i.d. $\mathrm{U}[0,1]$.

Obtaining conditional independence The key to the proof is finding a coupling of $(X_i)_i$ to a new r.v. Z (possibly after enlarging $(\Omega, \mathcal{F}, \mathsf{P})$) such that:

(i) one still has exchangeability, now in the enhanced form

$$(Z, X_1, X_2, \dots) \stackrel{\mathrm{d}}{=} (Z, X_{\sigma(1)}, X_{\sigma(2)}, \dots) \quad \forall \sigma \in \mathrm{Sym}(\mathbb{N})$$

(so Z is not moved by the permutation action);

(ii) the r.v.s X_i are conditionally independent over Z: this means that

$$\mathsf{P}(X_1 \in \mathrm{d}x_1,\, X_2 \in \mathrm{d}x_2,\, \dots,\, X_k \in \mathrm{d}x_k \mid Z) = \prod_{j=1}^{k} \mathsf{P}(X_i \in \mathrm{d}x_i \mid Z)$$

for any finite k and any $x_1, \dots, x_k \in E$. In terms of conditional expectations, this asserts that

$$\mathsf{E}(f_1(X_1)\cdots f_k(X_k) \,|\, Z) = \prod_{i=1}^{k} \mathsf{E}(f_i(X_i) \,|\, Z)$$

for any bounded measurable functions $f_1, \dots, f_k : E \longrightarrow \mathbb{R}$.

This coupling is obtained from a simple 'duplication' trick. Observe that exchangeability of $(X_i)_i$ implies also

$$(X_i)_i \stackrel{\mathrm{d}}{=} (X_{\gamma(i)})_{\gamma(i)}$$

whenever $\gamma : \mathbb{N} \longrightarrow \mathbb{N}$ is an *injection* (not necessarily a permutation). This is because the distributions of these two arrays are determined by their finite-dimensional marginals, and for any finite collection $i_1, i_2, \dots, i_m \in \mathbb{N}$ we can find a genuine permutation $\sigma : \mathbb{N} \longrightarrow \mathbb{N}$ such that $\sigma(i_r) = \gamma(i_r)$ for all r (but σ and γ differ elsewhere if necessary).

Now, of course, our previous choice of $\mathbb{N}$ as index set was rather arbitrary; we could have used, say, $\mathbb{Z}$ instead. But if we switch to indexing by $\mathbb{Z}$, we now rediscover the original family $(X_i)_{i\in\mathbb{N}}$ inside the new family $(X_i)_{i\in\mathbb{Z}}$, in the sense that this sub-family has the same distribution as the $\mathbb{N}$-indexed family that we started with. This is simply because we can let $\gamma : \mathbb{Z} \longrightarrow \mathbb{Z}$ be an injection with image equal to $\mathbb{N}$ and apply the reasoning above.

This completely trivial observation is important, because it provides a large collection of extra r.v.s $(X_i)_{i\in\mathbb{Z}\setminus\mathbb{N}}$ from which to synthesize the new r.v. Z. Let $\mathbb{N}^{\mathrm{c}} := \mathbb{Z}\setminus\mathbb{N}$. Letting $Z = (X_i)_{i\in\mathbb{N}^{\mathrm{c}}}$, a r.v. valued in $E^{\mathbb{N}^{\mathrm{c}}}$, we will show that this has the desired properties.

To see property (i), observe that if $\sigma : \mathbb{N} \longrightarrow \mathbb{N}$ is any permutation, then we may define a permutation $\tilde{\sigma} : \mathbb{Z} \longrightarrow \mathbb{Z}$ to agree with σ on $\mathbb{N}$ and to be the identity on $\mathbb{N}^{\mathrm{c}}$, and now one has

$$\begin{aligned}
(Z, X_{\sigma(1)}, X_{\sigma(2)}, \dots) &= ((X_i)_{i\le 0}, X_{\sigma(1)}, X_{\sigma(2)}, \dots)\\
&= ((X_{\tilde{\sigma}(i)})_{i\le 0}, X_{\tilde{\sigma}(1)}, X_{\tilde{\sigma}(2)}, \dots)\\
&\stackrel{\mathrm{d}}{=} ((X_i)_{i\le 0}, X_1, X_2, \dots)\\
&= (Z, X_1, X_2, \dots),
\end{aligned}$$

where the equality of distributions in the middle follows from the exchangeability of the original sequence.

Property (ii) needs a deeper idea. By induction on k it suffices to prove that

$$\mathsf{E}(f_1(X_1)\cdots f_k(X_k)\,|\,Z) = \mathsf{E}(f_1(X_1)\cdots f_{k-1}(X_{k-1})\,|\,Z)\mathsf{E}(f_k(X_k)\,|\,Z)$$

for any bounded measurable functions $f_1, \dots, f_k : E \longrightarrow \mathbb{R}$, and this, in turn, is really asserting that

$$\mathsf{E}(f_k(X_k)\,|\,Z, X_1, \dots, X_{k-1}) = \mathsf{E}(f_k(X_k)\,|\,Z)$$

(i.e., if we condition $f_k(X_k)$ on Z, X_1,..., X_{k-1}, then no more information is retained about it than if we condition on Z alone).

To prove this, let $\mathcal{F}_1$ be the σ-algebra generated by Z and $\mathcal{F}_2$ the σ-algebra generated by $(Z, X_1, \dots, X_{k-1})$. Then $\mathcal{F}_1 \subseteq \mathcal{F}_2$, and so the law of iterated conditional expectation gives

$$\mathsf{E}(f_k(X_k)\,|\,\mathcal{F}_1) = \mathsf{E}\big(\mathsf{E}(f_k(X_k)\,|\,\mathcal{F}_2)\,\big|\,\mathcal{F}_1\big).$$

In particular, this implies that

$$\|\mathsf{E}(f_k(X_k)\,|\,\mathcal{F}_1)\|_2 \le \|\mathsf{E}(f_k(X_k)\,|\,\mathcal{F}_2)\|_2$$

with equality if and only if the functions themselves are equal. This is because, in Hilbert-space terms, conditional expectations are orthogonal projections.

Therefore we need only prove this equality of norms. However, recalling our definition of Z, we have

$$\mathcal{F}_1 = \sigma\text{-alg}\big((X_j)_{j\leq 0}\big) \quad \text{and} \quad \mathcal{F}_2 = \sigma\text{-alg}\big((X_j)_{j\leq k-1}\big).$$

Now let $\tau : \mathbb{N} \longrightarrow \mathbb{N}$ be the injection

$$\tau(\ell) = \begin{cases} \ell & \text{if } \ell \geq k \\ \ell - (k-1) & \text{if } \ell \leq k-1, \end{cases}$$

and observe that

$$\mathcal{F}_1 = \sigma\text{-alg}\big((X_{\tau(j)})_{j\leq k-1}\big).$$

Exchangeability applied to this τ implies that

$$((X_{\tau(i)})_{i\leq k-1}, X_{\tau(k)} = X_k) \stackrel{\mathrm{d}}{=} ((X_i)_{i\leq k-1}, X_k),$$

and hence

$$\|\mathsf{E}(f_k(X_k)\,|\,\mathcal{F}_1)\|_2 = \|\mathsf{E}(f_k(X_k)\,|\,\mathcal{F}_2)\|_2,$$

completing the proof of (ii).

Finishing the proof Now we need only the Noise-Outsourcing Lemma and some routine bookkeeping. That lemma gives a Borel function $g' : [0,1] \times [0,1] \longrightarrow E$ such that

$$(Z, X_i) \stackrel{\mathrm{d}}{=} (Z, g'(Z, U_i))$$

for each i, where the U_i are i.i.d. $\mathrm{U}[0,1]$, independent from Z. The same function g' works for each i, since exchangeability implies that all pairs (Z, X_i) have the same distribution. Given this, the conditional independence over Z proved in (ii) implies for the whole sequence that

$$(Z, (X_i)_{i\in\mathbb{N}}) \stackrel{\mathrm{d}}{=} (Z, (g'(Z, U_i))_{i\in\mathbb{N}}).$$

Finally, another appeal to Lemma 3.1, this time in the simple case $|S| = 1$, gives a Borel function $h : [0,1] \longrightarrow E^{\mathbb{N}^{\mathrm{c}}}$ such that $Z \stackrel{\mathrm{d}}{=} h(U)$ when $U \sim \mathrm{U}[0,1]$. Substituting this into the above gives

$$(Z, (X_i)_{i\in\mathbb{N}}) \stackrel{\mathrm{d}}{=} (h(U), (g'(h(U), U_i))_{i\in\mathbb{N}}),$$

where U is independent from $(U_i)_i$, and hence completes the proof with $g(x,y) := g'(h(x),y)$. $\square$

Remark Instead of constructing the new r.v. Z as above, many proofs of de Finetti's Theorem prove that the r.v.s X_i's are conditionally independent over their own tail σ-alegbra. Although possibly cleaner, this argument does not generalize so directly to the case of higher-dimensional arrays, so I have avoided it. $\triangleleft$

Finite sequences

The analog of de Finetti's Theorem does not hold for finite sequences. One can see that the proof given above makes important use of the infinitude of $\mathbb{N}$, most obviously through the existence of injections $\gamma : \mathbb{N} \longrightarrow \mathbb{N}$ for which $\mathbb{N} \setminus \gamma(\mathbb{N})$ is infinite.

A rather weaker characterization is possible for finite exchangeable sequences, however. Given a finite sequence $\mathbf{x} = (x_i)_{i=1}^n \in E^n$, let

$$\mathcal{E}(\mathbf{x}) := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$$

be its empirical distribution. Clearly $\mathcal{E}$ is a $\mathrm{Sym}(n)$-invariant function $E^n \longrightarrow \Pr E$. On the other hand, if $\nu \in \Pr E$ lies in the image of $\mathcal{E}$ (meaning that it is a sum of n equal-weight atoms, not necessarily distinct), then let $\nu^{(n)}$ be the uniform distribution on the finite set of sequences $\mathcal{E}^{-1}\{\nu\}$ (that is, all sequences whose frequencies are given by ν).

Proposition 4.1 *If μ is a $\mathrm{Sym}(n)$-invariant probability on E^n, then*

$$\mu(\mathrm{d}\mathbf{x} \,|\, \mathcal{E}(\mathbf{x}) = \nu) = \nu^{(n)}(\mathrm{d}\mathbf{x}),$$

so μ is a mixture of the measures $\nu^{(n)}$:

$$\mu = \int_{\Pr E} \nu^{(n)}\, \mathcal{E}_*\mu(\mathrm{d}\nu).$$

$\square$

This is an elementary calculation, and we omit the proof. However, it is worth knowing that this gives another approach to de Finetti's Theorem. Given the finite exchangeable law

$$\mu = \int_{\Pr E} \nu^{(n)}\, \mathcal{E}_*\mu(\mathrm{d}\nu),$$

a fairly easy estimate can be used to compare it with the mixture of product measures

$$\int_{\Pr E} \nu^{\otimes n} \, \mathcal{E}_*\mu(\mathrm{d}\nu).$$

The difference between $\nu^{\otimes n}$ and $\nu^{(n)}$ is essentially that between sampling from a set of n samples with and without replacement. (In the latter case, $\nu^{(n)}$ is the law of the classical **urn sequence** obtained by sampling without replacement from the set of atoms of ν, possibly with some multiplicities.) For the first k outcomes of this sample, this difference is small when the sample size n is $\gg k$. A simple quantitative estimate gives the following.

Corollary 4.2 *If μ is a $\mathrm{Sym}(n)$-invariant probability on E^n and μ_k is the marginal of μ on any k coordinates, then*

$$\Big\|\mu_k - \int_{\Pr E} \nu^{\otimes k} \, \mathcal{E}_*\mu(\mathrm{d}\nu)\Big\|_{\mathrm{TV}} \leq \frac{k(k-1)}{n}.$$

□

Letting $n \longrightarrow \infty$ and then $k \longrightarrow \infty$ yields another proof of de Finetti's Theorem. See Section 1.2 of Kallenberg [Kal05] for more on these ideas, or Diaconis and Freedman [DF80] for better estimates that can be obtained given some extra restrictions on E.

5 Proof of the Aldous-Hoover Theorem

Now suppose $(X_{ij})_{ij\in\mathbb{N}^{(2)}}$ is an exchangeable random E-valued array. Recall that for us $ij \in \mathbb{N}^{(2)}$ implies $i \neq j$. We wish to show that there is a function f : $[0,1] \times [0,1]^2 \times [0,1] \longrightarrow E$ which is middle-symmetric and such that

$$(X_{ij})_{ij} \stackrel{\mathrm{d}}{=} (f(U, U_i, U_j, U_{ij}))_{ij\in\mathbb{N}^{(2)}}$$

where U, U_i, U_j and U_{ij} for $i, j \in \mathbb{N}$ are all i.i.d. $\mathrm{U}[0,1]$.

We will give the 'classical' proof, essentially following Aldous [Ald82, Ald85], which builds on de Finetti's Theorem.

Obtaining conditional independence Similarly to the proof of de Finetti's Theorem, the first step is to construct some new r.v.s coupled to $(X_{ij})_{ij}$ which give some conditional independence. This time we will find a whole sequence of r.v.s $(Y_i)_i$, coupled to $(X_{ij})_{ij}$ and valued in some auxiliary standard Borel space, such that the following hold:

(i) The whole enlarged array $(Y_i, X_{ij})_{i,j}$ is still exchangeable, i.e.

$$(Y_i, X_{ij})_{i\neq j\in\mathbb{N}\times\mathbb{N}} \stackrel{\mathrm{d}}{=} (Y_{\pi(i)}, X_{\pi(i)\pi(j)})_{i\neq j\in\mathbb{N}\times\mathbb{N}}$$

for any permutation $\pi : \mathbb{N} \longrightarrow \mathbb{N}$. Note that this array, unlike $(X_{ij})_{ij}$, is now indexed by *directed* edges (i, j).

(ii) The r.v.s X_{ij} are conditionally independent over the r.v.s Y_i, in the following specific sense:

$$\mathsf{E}\big(f_1(X_{i_1j_1})f_2(X_{i_2j_2})\cdots f_m(X_{i_mj_m})\,\big|\,(Y_i)_{i\in\mathbb{N}}\big) = \prod_{r=1}^{m}\mathsf{E}\big(f_r(X_{i_rj_r})\,\big|\,Y_{i_r},\,Y_{j_r}\big) \quad (1)$$

for any family of distinct pairs $i_rj_r \in \mathbb{N}^{(2)}$, $1 \leq r \leq m$ and any bounded measurable functions $f_1, \ldots, f_m : E \longrightarrow \mathbb{R}$. To be precise, this amounts to conditional independence over $(Y_i)_i$, and also the assertion that when $X_{i_1j_1}$ is conditioned on the whole sequence $(Y_i)_i$, it actually depends only on the two values Y_{i_1} and Y_{j_1}.

Now, just as in the case of de Finetti's Theorem, exchangeability of the family $(X_{ij})_{ij\in\mathbb{N}^{(2)}}$ implies that also

$$(X_{ij})_{ij\in\mathbb{N}^{(2)}} \stackrel{\mathrm{d}}{=} (X_{\gamma(i)\gamma(j)})_{ij\in\mathbb{N}^{(2)}}$$

for any injection $\gamma : \mathbb{N} \longrightarrow \mathbb{N}$. Just as before, it follows that we may assume $(X_{ij})_{ij\in\mathbb{N}^{(2)}}$ is actually a sub-array of a larger exchangeable array indexed by $\mathbb{Z}^{(2)}$.

The extra random variables X_{ij}, for which at least one of i, j lies in $\mathbb{N}^{\mathrm{c}}$, will be used to construct the Y_i. This time, we define the random variables Y_i for $i \in \mathbb{N}$ by

$$Y_i = \big((X_{i'j'})_{i'j'\in(\mathbb{N}^{\mathrm{c}})^{(2)}}, (X_{i'i})_{i'\in\mathbb{N}^{\mathrm{c}}}\big), \quad (2)$$

so these take values in the product space

$$\bar{E} := E^{(\mathbb{N}^{\mathrm{c}})^{(2)}} \times E^{\mathbb{N}^{\mathrm{c}}}.$$

Thus, for each $i \in \mathbb{N}$, Y_i simply records *all* of the values $X_{i'j'}$ where $i'j'$ is either an edge in $\mathbb{N}^{\mathrm{c}}$, or an edge that joins i to $\mathbb{N}^{\mathrm{c}}$.

The invariance property (i) of the family $\big(Y_i, X_{ij}\big)_{i\neq j\in\mathbb{N}\times\mathbb{N}}$ is now an immediate consequence of the exchangeability of the whole $\mathbb{Z}^{(2)}$-indexed array, just as in the case of de Finetti's Theorem.

It remains to prove (1). By induction on m and the law of iterated conditional expectation, it suffices to show that

$$\mathsf{E}\big(f(X_{i_mj_m})\,\big|\,X_{i_1j_1},\ldots,X_{i_{m-1}j_{m-1}},(Y_i)_{i\in\mathbb{N}}\big) = \mathsf{E}\big(f(X_{i_mj_m})\,\big|\,Y_{i_m},\,Y_{j_m}\big)$$

for any bounded measurable function $f: E \longrightarrow \mathbb{R}$.

Let $\mathcal{F}_2$ be the σ-algebra generated by all the random variables $X_{i_1j_1}$, $\ldots$, $X_{i_{m-1}j_{m-1}}$ and $(Y_i)_{i\in\mathbb{N}}$, and $\mathcal{F}_1$ the σ-algebra generated by just Y_{i_m} and Y_{j_m}. Hence $\mathcal{F}_1 \subseteq \mathcal{F}_2$, and by another appeal to iterated conditional expectation we know that

$$\mathsf{E}(f(X_{i_mj_m})\,|\,\mathcal{F}_1) = \mathsf{E}(\mathsf{E}(f(X_{i_mj_m})\,|\,\mathcal{F}_2)\,|\,\mathcal{F}_1).$$

We wish to show that $\mathsf{E}(f(X_{i_mj_m})\,|\,\mathcal{F}_1) = \mathsf{E}(f(X_{i_mj_m})\,|\,\mathcal{F}_2)$, and once again the norm-contracting property of conditional expectation shows that

$$\|\mathsf{E}(f(X_{i_mj_m})\,|\,\mathcal{F}_1)\|_2 \leq \|\mathsf{E}(f(X_{i_mj_m})\,|\,\mathcal{F}_2)\|_2 \tag{3}$$

with equality of the functions if and only if their norms are equal.

We now perform the analog of the re-arrangement trick that proved de Finetti's Theorem, but in this case it will require slightly more care. Let $T \subset \mathbb{N}^{\mathrm{c}}$ be a further subset such that T and $\mathbb{N}^{\mathrm{c}} \setminus T$ are both infinite. Given this infinitude, we may now choose an injection $\gamma : \mathbb{Z} \longrightarrow \mathbb{Z}$ with the following properties:

- $\gamma(i_m) = i_m$ and $\gamma(j_m) = j_m$,
- $\gamma(\mathbb{N}^{\mathrm{c}}) = T$, and
- $\gamma(\mathbb{N} \setminus \{i_m, j_m\}) = \mathbb{N}^{\mathrm{c}} \setminus T$

(so all indices except i_m and j_m end up somewhere in $\mathbb{N}^{\mathrm{c}}$).

After applying this map to the indices, the r.v.s $X_{i_rj_r}$ are sent to $X_{\gamma(i_r)\gamma(j_r)}$, and the r.v.s Y_i are replaced by

$$Y'_{\gamma(i)} := \big((X_{i'j'})_{i'j'\in T^{(2)}}, (X_{i'\gamma(i)})_{i'\in T}\big)$$

(recall (2)), and the joint exchangeability of all our r.v.s promises that

$$\begin{aligned}\mathsf{E}\big(f(X_{i_mj_m})\,\big|\,X_{i_1j_1},\ldots,X_{i_{m-1}j_{m-1}},(Y_i)_{i\in\mathbb{N}}\big)&\\ \stackrel{\mathrm{d}}{=} \mathsf{E}\big(f(X_{i_mj_m})\,\big|\,X_{\gamma(i_1)\gamma(j_1)},\ldots,X_{\gamma(i_{m-1})\gamma(j_{m-1})},(Y'_{\gamma(i)})_{i\in\mathbb{N}}\big).&\end{aligned}$$

Taking L^2 norms, this implies

$$\|\mathsf{E}(f(X_{i_m j_m}) \,|\, \mathcal{F}_2)\|_2 = \|\mathsf{E}(f(X_{i_m j_m}) \,|\, \mathcal{F}_3)\|_2,$$

where $\mathcal{F}_3$ is the σ-algebra generated by $X_{\gamma(i_1)\gamma(j_1)}, \dots, X_{\gamma(i_{m-1})\gamma(j_{m-1})}$ and $(Y'_{\gamma(i)})_{i\in\mathbb{N}}$. (Note that in the case of de Finetti's Theorem this third σ-algebra was not needed; the difference is that in the present case we may be unable to find an injection γ that converts $\mathcal{F}_2$ exactly into $\mathcal{F}_1$.)

Upon unraveling the definition of γ, one sees that $\mathcal{F}_3$ is generated by some r.v.s of the form $X_{i'j'}$ where $i'j'$ is either an edge in $\mathbb{N}^{\mathrm{c}}$ or is of the form $i'i_m$ or $i'j_m$ for some $i' \in \mathbb{N}^{\mathrm{c}}$. This is because $\gamma(\mathbb{Z}) = \mathbb{N}^{\mathrm{c}} \cup \{i_m, j_m\}$, and the edge $i_m j_m$ is distinct from $i_r j_r$ for $r \leq m-1$. This particular subcollection of the random variables $X_{i'j'}$ is determined by the various coordinates appearing in the definition (2) of Y_{i_m} and Y_{j_m}. Therefore one has the inclusion $\mathcal{F}_3 \subseteq \mathcal{F}_1$, and hence

$$\|\mathsf{E}(f(X_{i_m j_m}) \,|\, \mathcal{F}_3)\|_2 \leq \|\mathsf{E}(f(X_{i_m j_m}) \,|\, \mathcal{F}_1)\|_2,$$

by using again the norm-contracting property of conditional expectation. Since this left-hand side is equal to $\|\mathsf{E}(f(X_{i_m j_m}) \,|\, \mathcal{F}_2)\|_2$, we deduce the desired equality of norms by sandwiching with the previous inequality.

Completion of the proof: using de Finetti By the Noise-Outsourcing Lemma, considering any given pair $ij \in \mathbb{N}^{(2)}$ we can choose a Borel function $g : [0,1]^2 \times [0,1] \longrightarrow [0,1]$ such that

$$(Y_i, Y_j, X_{i,j}) \stackrel{\mathrm{d}}{=} (Y_i, Y_j, g(Y_i, Y_j, U)),$$

where $U \sim \mathrm{U}[0,1]$ is independent from everything else, and since the left-hand distribution is symmetric in i and j we may choose g to be symmetric in its first two arguments. Now by exchangeability, this same g must work for every ij.

However, in view of (1) this now implies that for the joint distribution of the whole process $(Y_i, X_{ij})_{i\neq j}$ we have

$$(Y_i, X_{ij})_{i\neq j} \stackrel{\mathrm{d}}{=} (Y_i, g(Y_i, Y_j, U_{ij}))_{i\neq j},$$

where $U_{ij} \sim \mathrm{U}[0,1]$ are i.i.d. and independent from everything else.

Finally, de Finetti's Theorem applied to $(Y_i)_i$ gives $h : [0,1] \times [0,1] \longrightarrow \bar{E}$ such that $(Y_i)_i \stackrel{\mathrm{d}}{=} (h(U, U_i))_i$, and so combining with the above we have

$$(Y_i, X_{i,j})_{i\neq j} \stackrel{\mathrm{d}}{=} (h(U, U_i), g(h(U, U_i), h(U, U_j), U_{ij}))_{i\neq j},$$

where U, U_i, $U_{i,j}$ are from the full array of i.i.d. $\mathrm{U}[0,1]$. This implies the desired conclusion with the middle-symmetric function

$$f(x, x_1, x_2, x_{12}) := g(h(x, x_1), h(x, x_2), x_{12}).$$

$\square$

Remarks 1. This proof looks rather like magic, because it's hard to locate where we did anything nontrivial. Perhaps the first place one should point to is equality (3). The essence of this theorem is that conditioning $X_{i_m j_m}$ on all the other random variables that gave rise to $\mathcal{F}_2$ is the same as conditioning on only Y_{i_m} and Y_{j_m}. For the proof, the key realization is that this assertion can be made 'quantitative', in that it requires only the equality of the L^2-norms appearing in (3).

2. Just as for de Finetti's Theorem, the analog of the Aldous-Hoover Theorem fails for finite arrays. Once again there is an approximate version of the story instead, but here it is substantially more delicate than for sequences, requiring the study of general structural results for large dense graphs (in particular, a version of the famous Szemerédi Regularity Lemma from graph theory). This is a part of the theory of 'limit objects' for sequences of dense graphs, which has recently been the subject of considerable study by combinatorists (see, for instance, Lovász and Szegedy [LS06]). We set that aside here; its connection to exchangeability is surveyed in [Aus08] and [DJ07]. $\triangleleft$

6 Random partitions and mass partitions

The paintbox theorem

As suggested previously, the importance of exchangeable random structures is their appearance as a result of infinite sampling. The general philosophy is nicely expressed by Aldous in Section 3 of [Ald]:

> 'One way of examining a complex mathematical structure is to sample i.i.d. random points and look at some form of induced substructure relating the random points.'

Our first and most classical example is Kingman's study of exchangeable random partitions. More recently this has become a central ingredient in the formulation and study of certain coagulation and fragmentation processes, but we will not approach that subject here; see, for instance, Bertoin [Ber06], and Schweinsberg's course at this workshop.

First, a **mass partition** is a measure on $\mathbb{N}$ of mass at most 1 and with atoms of non-increasing size; equivalently, it is a non-negative, non-increasing sequence $(s_k)_{k\in\mathbb{N}}$ such that $\sum_k s_k \leq 1$. Let $\mathcal{P}_{\mathrm{m}}$ denote the space of mass partitions. It is easily shown to be compact for the topology inherited from the product topology on $[0,1]^{\mathbb{N}}$.

Next, let Ptn denote the space of all partitions of $\mathbb{N}$. Given $\Pi \in \mathrm{Ptn}$ and $A \subseteq \mathbb{N}$, we write $\Pi|_A$ for the restriction of Π to A. The space Ptn is also easily seen to be a compact metrizable space by letting two partitions Π, Π' be close if $\Pi|_{[n]} = \Pi'|_{[n]}$ for some large n. Partitions Π are in bijective correspondence with equivalence relations on Π, where the associated relation $\sim_\Pi$ is defined by

$$m \sim_\Pi n \quad \Longleftrightarrow \quad m, n \text{ are in the same cell of } \Pi.$$

There is a natural action of $\mathrm{Sym}(\mathbb{N})$ on Ptn coming from the action on $\mathbb{N}$ itself: $\sigma(\Pi) := \{\sigma^{-1}(A) : A \in \Pi\}$. A probability measure on Ptn is **exchangeable** if it is invariant for this action.

Example Random mass partitions give a simple construction of exchangeable random partitions. Suppose that $\nu \in \mathrm{Pr}\,\mathcal{P}_{\mathrm{m}}$, and now draw a random element of Ptn as follows. First, pick $(s_k)_k \sim \nu$ at random. Having done so, pick a sequence $(m_n)_n$ at random so that the m_n are i.i.d. elements of $\mathbb{N} \cup \{\infty\}$ with law

$$\mathsf{P}(m_n = k) = s_k, \quad \mathsf{P}(m_n = \infty) = 1 - \sum_k s_k.$$

Finally, let Π consist of the classes

$$\{n \in \mathbb{N} : m_n = k\}$$

for all $k \in \mathbb{N}$, together with all the singletons $\{n\}$ for which $m_n = \infty$.

It is a simple calculation to check that this process is an exchangeable random partition, i.e. its law is a $\mathrm{Sym}(\mathbb{N})$-invariant element of $\mathrm{Pr}(\mathrm{Ptn})$. We refer to this process as $\mathrm{Samp}(\nu)$; it is also called the **paintbox** partition obtained from ν. This name derives from the following intuitive picture. We think of $\mathbb{N}$ as a paintbox from which we choose a colour for each $n \in \mathbb{N}$ independently at random according to the probabilities $(s_1, s_2, \ldots)$, or choose not to colour n with probability $1 - \sum_k s_k$. Then $m, n \in \mathbb{N}$ lie in the same cell of $\boldsymbol{\Pi}$ if and only if they have both been painted and are the same colour. $\triangleleft$

Just as for sequences and arrays, the point here is that the 'natural' examples turn out to be the only ones.

Theorem 6.1 (Kingman's Paintbox Theorem) *Every exchangeable random partition* $\mathbf{\Pi}$ *has the same distribution as* $\mathrm{Samp}(\nu)$ *for some* $\nu \in \mathrm{Pr}\,\mathcal{P}_{\mathrm{m}}$.

Proof This is a consequence of de Finetti's Theorem. To make contact with that theorem, we construct a $[0,1]$-valued process $(V_n)_{n\in\mathbb{N}}$ as follows:

- first, choose a sample of the random partition $\mathbf{\Pi}$;
- then, for each cell $C \in \mathbf{\Pi}$ choose an independent $\mathrm{U}[0,1]$ r.v. V_C;
- finally, let $V_n := V_C$ where C is the cell containing n.

Observe that:

(i) The r.v.s $(V_n)_n$ a.s. determine the partition $\mathbf{\Pi}$, because a.s. we have that all V_C for different cells C are distinct (since there are only countably many cells), and hence

$$V_n = V_m \quad \text{iff} \quad n,m \text{ lie in the same cell of } \mathbf{\Pi}. \tag{4}$$

(ii) The sequence $(V_n)_n$ is exchangeable. Indeed, after fixing the sample $\mathbf{\Pi}$, the process $(V_n)_n$ arises simply from independent choices of constant values within each cell of $\mathbf{\Pi}$, and hence

$$\mathrm{law}((V_{\sigma(n)})_{n\in\mathbb{N}}\,|\,\sigma(\mathbf{\Pi})) = \mathrm{law}((V_n)_{n\in\mathbb{N}}\,|\,\mathbf{\Pi}).$$

Since $\mathbf{\Pi}$ is exchangeable, i.e. $\sigma(\mathbf{\Pi}) \stackrel{\mathrm{d}}{=} \mathbf{\Pi}$, averaging over the distribution of $\mathbf{\Pi}$ now gives that $(V_n)_n$ is exchangeable.

By de Finetti's Theorem, there is some Borel $f : [0,1]\times[0,1] \longrightarrow [0,1]$ such that

$$(V_n)_n \stackrel{\mathrm{d}}{=} (f(U,U_n))_n.$$

Finally, let $\theta(x) \in \mathrm{Pr}[0,1]$ be the law of $f(x,U)$ when $U \sim \mathrm{U}[0,1]$, and let $s(x) = (s_n(x))_n$ be the sequence of masses of the atoms of $\theta(x)$ arranged in non-increasing order. This defines a measurable function $s : [0,1] \longrightarrow \mathcal{P}_{\mathrm{m}}$. Conditionally on $x \in [0,1]$, the rule (4) gives that if $m,n \in \mathbb{N}$ are distinct then they lie in the same cell of our random partition if and only if $f(x,U_m)$ and $f(x,U_n)$ land on the same atom of $\theta(x)$. This is clearly equivalent to the description in the example, so $\mathbf{\Pi}$ has law $\mathrm{Samp}(\nu)$ with ν the law of $s(U)$. $\square$

The Chinese Restaurant and the Poisson-Dirichlet distributions

Having introduced exchangeable random partitions, we take the chance to introduce also an important family of examples.

Fix a parameter $0 < \alpha < 1$, and consider the measure $m_\alpha(\mathrm{d}x) = x^{-\alpha-1}\mathrm{d}x$ on $(0,\infty)$. Simple calculus gives

$$m_\alpha([\varepsilon,\infty)) < \infty \ \ \forall \varepsilon > 0 \quad \text{and} \quad \int_0^1 x\, m_\alpha(\mathrm{d}x) < \infty,$$

using $\alpha > 0$ and $\alpha < 1$ respectively.

Now let Λ be a Poisson point process on $(0,\infty)$ with intensity measure m_α. This is a random countable subset of $(0,\infty)$, and the first inequality above translates into the fact that $|\Lambda \cap [\varepsilon,\infty)| < \infty$ a.s. for all $\varepsilon > 0$. This means that we may enumerate the points of Λ in non-increasing order, say as $(u_k)_k$. In addition,

$$\mathsf{E}\Big(\sum_k u_k 1_{\{u_k \le 1\}}\Big) = \int_0^1 x\, m_\alpha(\mathrm{d}x) < \infty,$$

so these two facts together imply that $\sum_k u_k$ is finite a.s. We may therefore consider the sequence

$$s_k := \frac{u_k}{\sum_k u_k}.$$

This is now a random element of $\mathcal{P}_{\mathrm{m}}$.

Thus we have constructed a family of probability measures on $\mathcal{P}_{\mathrm{m}}$ indexed by $\alpha \in (0,1)$. They are called the **Poisson-Dirichlet** distributions and are denoted by $\mathrm{PD}(\alpha, 0)$. (They are part of a larger two-parameter family $\mathrm{PD}(\alpha,\theta)$, whose others members will not concern us.) They were introduced by Pitman and Yor in [PY97], and have since shown up in a remarkable range of applications. We will meet them again later; at this point we simply record some basic facts about the associated paintbox random partitions.

To do so, first consider any random partition $\mathbf{\Pi}$ (not necessarily exchangeable). The law of $\mathbf{\Pi}$ is determined by the laws of all its finite restrictions $\mathbf{\Pi}|_{[n]}$, and hence by the function

$$p(B_1, B_2, \dots, B_k) := \mathsf{P}(\mathbf{\Pi}|_{[n]} \text{ has cells } B_1, B_2, \dots, B_k), \tag{5}$$

defined for any partition $\{B_1, \dots, B_k\}$ of a finite set $[n]$.

If this quantity depends only on $|B_1|$, ..., $|B_k|$ then $\mathbf{\Pi}$ is said to be **weakly exchangeable**; clearly this is implied by exchangeability. In this case p is written as a function of these cardinalities and is called the **exchangeable partition probability function** ('EPPF').

In the case of $\mathrm{Samp}(\mathrm{PD}(\alpha, 0))$, some clever calculus (omitted here) now yields the **Pitman sampling formula** for the EPPF:

$$p(n_1, n_2, \ldots, n_k) = \frac{\prod_{i=1}^k \big(\alpha(1-\alpha)(2-\alpha)\cdots(n_i - \alpha)\big)}{(n-1)!/(k-1)!}. \tag{6}$$

Using this, one can prove that the following remarkable construction also gives rise to the law $\mathrm{Samp}(\mathrm{PD}(\alpha, 0))$. Consider a restaurant with an infinite number of tables in a row, all of them infinitely large. Initially all tables are empty. At subsequent times, customers arrive and pick tables according to the following random process. The first customer simply sits at table 1. Now suppose that at time $n \geq 1$, the first k tables already have at least one customer. Then the $(n+1)^{\mathrm{th}}$ customer chooses to sit at the first unoccupied table with probability $k\alpha/n$, or chooses to sit at the i^{th} occupied table with probability

$$\frac{(\text{number of people already at table } i) - \alpha}{n}.$$

For future reference, let us note an alternative way of writing this rule: it asserts that the probability of customer $n+1$ choosing to sit at the same table as any given customer $m \in \{1, 2, \ldots, n\}$ is given by

$$\frac{(\text{number of other people already sitting with } m) + 1 - \alpha}{n}. \tag{7}$$

Over time, this process reveals a partition of $\mathbb{N}$ whose classes are the sets of customers sitting at each table. This is often referred to as the **Chinese Restaurant Process**, and was introduced in work of Dubins and Pitman; see [Pit95].

Proposition 6.2 *The random partition* $\mathbf{\Pi}$ *resulting from the Chinese Restaurant process is exchangeable with law* $\mathrm{Samp}(\mathrm{PD}(\alpha, 0))$. $\square$

This can be proved directly by computing the probabilities (5) for this partition by induction on $|B_1| + \cdots + |B_k| = n$, and verifying that they agree with (6). Without this calculation, it is not even obvious that the Chinese Restaurant Process gives a weakly exchangeable partition.

The Poisson-Dirichlet processes also arise naturally from a remarkable number of other discrete random structures: a nice overview is given in Section 11 of Aldous [Ald85].

7 Gram-de Finetti matrices and probabilities on Hilbert space

Our next class of exchangeable structures is the following.

Definition 7.1 (Gram-de Finetti matrices) *A* ***Gram-de Finetti matrix*** *is a symmetric exchangeable array* $(R_{i,j})_{(i,j)\in\mathbb{N}^2}$ *of* $\mathbb{R}$*-valued random variables such that the matrix* $(R_{i,j})_{i,j}$ *is almost surely non-negative definite.*

This time, the natural 'sampling' examples are the following.

Examples Suppose that $\boldsymbol{\mu}$ is a random probability measure on a Hilbert space $\mathfrak{H}$, which will always be assumed real and separable, and construct a random array from it as follows. First, sample $\boldsymbol{\mu}$ at random. Having chosen $\boldsymbol{\mu}$, now draw from it an i.i.d. sequence of vectors $(\xi_i)_{i\in\mathbb{N}}$, and finally set

$$R_{i,j} := \xi_i \cdot \xi_j,$$

the matrix of pairwise inner products in $\mathfrak{H}$. This defines a Gram-de Finetti matrix, whose law we denote by $\mathrm{Samp}(\boldsymbol{\mu})$.

To make this a little more general, suppose instead that $\boldsymbol{\mu}$ is a random probability measure on $\mathfrak{H} \times [0,\infty)$, and modify the above construction as follows: after choosing $\boldsymbol{\mu}$ at random, let (ξ_i, a_i) be an i.i.d. $\sim \boldsymbol{\mu}$ sequence and set

$$R_{i,j} := \xi_i \cdot \xi_j + a_i\delta_{ij},$$

where δ_{ij} is the Kronecker delta. This law will still be referred to as $\mathrm{Samp}(\boldsymbol{\mu})$. $\lhd$

If we fix $M > 0$, then the space of non-negative definite arrays with all entries bounded by M is compact for the product topology, so we may naturally talk of vague (= weak*) convergence for probability measures on this space, and moreover a limit of exchangeable measures is easily seen to be still exchangeable. This will be important for applications later.

The Dovbysh-Sudakov representation

The crucial fact which makes Gram-de Finetti matrices useful is that, just as for other exchangeable arrays, they *all* arise from sampling. This is the main result of this section.

Theorem 7.2 (Dovbysh-Sudakov representation; [DS82, Hes86, Pan10]) *For any Gram-de Finetti matrix R there is a random probability measure $\boldsymbol{\mu}$ on $\ell_2 \times [0, \infty)$ such that* $\mathrm{law}(R) = \mathrm{Samp}(\boldsymbol{\mu})$.

A suitable choice of $\boldsymbol{\mu}$ is called a **directing random measure** for R; in the next section we will address the issue of its uniqueness.

We will prove Theorem 7.2 only in the special case that R takes values in $[-1, 1]$ and $R_{i,i} \equiv 1$. Both of these assumptions can be removed with just a little more work, but we leave that to the references for the sake of brevity; the special case is enough for our later applications to spin glasses. In this special case we will show that there is a random probability measure $\boldsymbol{\mu}$ on the unit ball $B \subset \ell_2$ (rather than on $\ell_2 \times [0, \infty)$) such that

$$(R_{i,j})_{i \neq j \in \mathbb{N} \times \mathbb{N}} \stackrel{\mathrm{d}}{=} (\xi_i \cdot \xi_j)_{i \neq j \in \mathbb{N} \times \mathbb{N}},$$

where $(\xi_i)_i$ is a conditionally i.i.d. sequence drawn from $\boldsymbol{\mu}$. The diagonal terms are then taken care of simply by setting

$$a_i := 1 - \|\xi_i\|^2.$$

Proof of special case, following [Pan10][2] Since $(R_{i,j})_{i \neq j}$ is a symmetric exchangeable array, by the Aldous-Hoover Theorem we have

$$\mathrm{law}((R_{i,j})_{i \neq j}) = \mathrm{Samp}(f)$$

for some middle-symmetric $f : [0, 1] \times [0, 1]^2 \times [0, 1] \longrightarrow [-1, 1]$.

Instead of the arbitrary measurable function f, we want the richer geometric structure of sampling points from a random probability measure on ℓ_2. The rest of the proof goes into synthesizing the latter from the former.

Step 1: Letting $f_u := f(u, \cdot, \cdot, \cdot)$, it is easy to see that $\mathrm{Samp}(f)$ is a.s. non-negative definite if and only if $\mathrm{Samp}(f_u)$ is a.s. non-negative definite for a.e. u.

[2] I think this is similar to the proof of [Hes86], but I haven't been able to access that.

It therefore suffices to show that $\mathrm{Samp}(f_u)$ must arise from sampling from some measure μ_u on $B \subset \ell_2$ for a.e. u, where the measure μ_u is now non-random. From this, general measure theory gives a measurable selection $u \mapsto \mu_u$, which defines the desired random measure.

So suppose henceforth that f is a function of only (u_1, u_2, u_{12}), and that $\mathrm{Samp}(f)$ is a.s. non-negative definite. We may also suppose that

$$(R_{i,j})_{i\neq j} = (f(U_i, U_j, U_{ij}))_{i\neq j}$$

(not just in law), simply by taking this as our new definition of $(R_{i,j})_{i\neq j}$. We still have $R_{i,i} \equiv 1$.

Step 2: Next we show that $f(u_1, u_2, u_{12})$ cannot depend non-trivially on u_{12} without violating the a.s. non-negative definiteness of R. To make use of the non-negative definiteness, observe that for any $n \geq 1$ and any bounded measurable functions $h_1, \ldots, h_n : [0,1] \longrightarrow \mathbb{R}$ one has

$$\frac{1}{n}\sum_{i,j=1}^{n} R_{i,j} h_i(u_i) h_j(u_j) \geq 0 \quad \text{a.s..} \tag{8}$$

We will apply this with the following careful choice of functions. Let $n = 4m$, let $A_1, A_2 \subseteq [0,1]$ be any measurable subsets, and let

$$h_i(x) := \begin{cases} 1_{A_1}(x) & 1 \leq i \leq m, \\ -1_{A_1}(x) & m+1 \leq i \leq 2m, \\ 1_{A_2}(x) & 2m+1 \leq i \leq 3m, \\ -1_{A_2}(x) & 3m+1 \leq i \leq 4m. \end{cases}$$

Now consider

$$\int_{[0,1]^n} \frac{1}{n}\sum_{i,j=1}^{n} f(u_i, u_j, u_{ij}) h_i(u_i) h_j(u_j) \prod_{i=1}^{n} \mathrm{d}u_i,$$

which is an average over the u_i (but not the u_{ij}) of an expression like (8). Taking the sum outside the integral, we may write this as

$$D + I_{11} + I_{12} + \ldots + I_{34} + I_{44},$$

where D contains the diagonal terms ($i = j$) and each $I_{k\ell}$ consists of those terms with $i \neq j$, $(k-1)m + 1 \leq i \leq km$ and $(\ell - 1)m + 1 \leq j \leq \ell m$.

Since $|f| \leq 1$, D consists of an average of n terms that are uniformly bounded by 1. On the other hand, each $I_{k\ell}$ is $(1/n)$ times a sum of terms of the form

$$\pm \int_A \int_{A'} f(u, v, u_{ij}) \, \mathrm{d}u\mathrm{d}v$$

with each of A, A' equal to either A_1 or A_2. If $k \neq \ell$ there are m^2 of these terms in $I_{k\ell}$, and if $k = \ell$ then there are $m^2 - m$ (because the diagonal terms are in D instead). Also, as we vary k and ℓ the $\pm$-signs almost exactly cancel, in that each integral $\int_A \int_{A'}$ appears in all of the $I_{k,\ell}$ together the same number of times as $-\int_A \int_{A'}$, apart from a small correction owing to the diagonal terms.

However, each of these individual integrals appearing in one of the sums $I_{k\ell}$ depends on a different variable u_{ij}. If their dependence on this variable is not trivial up to a negligible set, then a simple estimate using the Central Limit Theorem shows that these off-diagonal sums must have approximately a centred Gaussian distribution as functions of the uniform r.v.s $(U_{ij})_{i,j \leq n}$, with variance of order 1. In particular, there is some small positive probability in these uniform r.v.s that the sum $I_{11} + I_{12} + \ldots + I_{44}$ will be negative and have absolute value much larger than $|D| \leq 1$. This would make the whole sum $D + I_{11} + I_{12} + \ldots + I_{44}$ negative, and this would contradict non-negative definiteness.

So instead one must have that for every $A_1, A_2 \subseteq [0, 1]$ the quantity

$$\int_{A_1} \int_{A_2} f(u, v, w) \, \mathrm{d}u\mathrm{d}v$$

is independent of w outside a Lebesgue-negligible set of w. Since f itself may be approximated by a linear combination of indicator functions of the form $1_{A_1 \times A_2}$, this implies that $f(u, v, w)$ does not depend on w outside of some negligible set, as required.

Henceforth we write f as a function of only (u_1, u_2).

Step 3: Now consider the linear operator A on $L^2[0, 1]$ defined by

$$Ag(x) := \int_0^1 g(y) f(x, y) \, \mathrm{d}y.$$

Since f is uniformly bounded by 1, this is a bounded operator, and moreover an easy exercise shows that it is compact. It is self-adjoint owing to the symmetry of f.

Therefore the Spectral Theorem for compact self-adjoint operators provides a sequence of real eigenvalues λ_i with $|\lambda_i| \longrightarrow 0$ and eigenfunctions $\varphi_i \in L^2[0, 1]$

such that

$$f(x,y) = \sum_{i=1}^{\infty} \lambda_i \overline{\varphi_i(x)} \varphi_i(y),$$

where the series converges in $L^2([0,1]^2)$. (If you're not familiar with this spectral theorem, then think of it as putting the 'symmetric $[0,1] \times [0,1]$ matrix $f(\cdot,\cdot)$' into 'diagonal form'; and see any standard book covering Hilbert space operators, such as Conway [Con90].)

Step 4: Another property of the operator A is that it is non-negative definite for a.e. u; once again, this is necessary in order that R be a.s. non-negative definite. To see this, simply observe that for any measurable function $h : [0,1] \longrightarrow \mathbb{R}$, non-negative definiteness and the Law of Large Numbers applied to the independent r.v.s $U_i \sim \mathrm{U}[0,1]$ give

$$0 \leq \frac{1}{n^2} \sum_{i,j=1}^{n} f(U_i, U_j) h(U_i) h(U_j) \xrightarrow{\text{a.s.}} \int_{[0,1]^2} f(u,v) h(u) h(v) \, \mathrm{d}u \mathrm{d}v = \langle h, Ah \rangle$$

as $n \longrightarrow \infty$, where this denotes the inner product in $L^2[0,1]$. Therefore we also have $\lambda_i \geq 0$ for all i a.s.

Step 5: Now define a measurable function $F : [0,1] \longrightarrow \mathbb{C}^{\mathbb{N}}$ by

$$F(x) := (\sqrt{\lambda_1}\varphi_1(x), \sqrt{\lambda_2}\varphi_2(x), \ldots).$$

We will next argue that it takes values in $B \subset \ell_2$ a.s. To be specific, let $x \in [0,1]$ be a Lebesgue density point for every φ_i simultaneously: that is,

$$\frac{1}{2\delta} \int_{(x-\delta,x+\delta)} \varphi_i(y) \, \mathrm{d}y \longrightarrow \varphi_i(x) \quad \forall i \text{ as } \delta \longrightarrow 0.$$

These points are co-negligible in $[0,1]$ by the Lebesgue Differentiation Theorem, which applies since $\varphi_i \in L^2[0,1] \subset L^1[0,1]$.

Now we can compute that

$$\begin{aligned} \sum_{i\geq 1} \lambda_i \Big| \frac{1}{2\delta} \int_{(x-\delta,x+\delta)} \varphi_i(y) \, \mathrm{d}y \Big|^2 &= \sum_{i\geq 1} \lambda_i \frac{1}{(2\delta)^2} \int_{(x-\delta,x+\delta)^2} \varphi_i(u) \overline{\varphi_i(v)} \, \mathrm{d}u \mathrm{d}v \\ &= \frac{1}{(2\delta)^2} \int_{(x-\delta,x+\delta)^2} f(u,v) \, \mathrm{d}u \mathrm{d}v, \end{aligned}$$

and this is ≤ 1 because f is pointwise bounded by 1. Letting $\delta \longrightarrow 0$ shows that

$$\sum_{i\geq 1} \lambda_i |\varphi_i(x)|^2 \leq 1 \quad \text{for a.e. } x,$$

as required.

In terms of F we now have the relation

$$f(x,y) = F(x) \cdot F(y),$$

where the right-hand side is the inner product in ℓ_2. Let μ be the distribution of $F(X)$ on B when $X \sim \mathrm{U}[0,1]$ (that is, the push-forward of Lebesgue measure under F).

Step 6: Lastly, recall that $\mathrm{Samp}(f)$ is the law of R, and write this relation explicitly as

$$\begin{aligned} \mathsf{P}(R_{i,j} \in \mathrm{d}r_{i,j}\ \forall i,j \leq N) &= \int_{[0,1]^N} 1_{\{f(u_i,u_j) \in \mathrm{d}r_{i,j}\ \forall i,j \leq N\}} \prod_{i \leq N} \mathrm{d}u_i \\ &= \int_{[0,1]^N} 1_{\{F(u_i)\cdot F(u_j) \in \mathrm{d}r_{i,j}\ \forall i,j \leq N\}} \prod_{i \leq N} \mathrm{d}u_i \end{aligned}$$

for any $N \geq 1$ and any non-negative definite matrix $(r_{ij})_{i,j \leq N}$. Finally, we recognize that

$$\int_{[0,1]^N} 1_{\{F(u_i)\cdot F(u_j) \in \mathrm{d}r_{i,j}\ \forall i,j \leq N\}} \prod_{i \leq N} \mathrm{d}u_i = \mu^{\otimes N}\{(\xi_i)_{i=1}^N : \xi_i \cdot \xi_j \in \mathrm{d}r_{ij}\ \forall i,j \leq N\}.$$

This is precisely the assertion that $\mathrm{law}(R) = \mathrm{Samp}(\mu)$, so the proof is complete. □

An important consequence of Theorem 7.2 is a sensible notion of convergence for random Hilbert space probability measures. If $(\boldsymbol{\mu}_n)_n$ is a sequence of random probability measures on (possibly different) Hilbert spaces, then they **sampling-converge** if the resulting laws $\mathrm{Samp}(\boldsymbol{\mu}_n)$ converge vaguely as probability measures on the space of non-negative definite matrices, and in this case their **limit object** is any choice of directing random measure for the limiting random matrix. If all these random probability measures have uniformly bounded support then the resulting Gram-de Finetti matrices will be uniformly bounded, and so there will always at least be subsequential limits.

This is worth comparing with the theory of limit objects for dense finite graphs [LS06, DJ07, Aus08], for which it is the direct analog of left-convergence of homomorphism densities. It is also very much in the spirit of the more general discussion of probability distributions on distance matrices that characterize a general metric probability space – see Sections $3\frac{1}{2}.4$ through $3\frac{1}{2}.7$ in Gromov [Gro99], and the references given there to works of Vershik.

8 Some uniqueness results

In the setting of the Structure Theorem 2.2, it is natural to ask which pairs of middle-symmetric functions f, f' give $\mathrm{Samp}(f) = \mathrm{Samp}(f')$. The necessary and sufficient condition is a little subtle, and is explained in detail in Chapter 7 of Kallenberg [Kal05]. However, in the settings of the Kingman and Dovbysh-Sudakov Theorems it is easier to be precise, so here we will focus on these.

In the case of random partitions, the representation is unique.

Proposition 8.1 *If* $\nu, \nu' \in \mathcal{P}_{\mathrm{m}}$ *and* $\mathrm{Samp}(\nu) = \mathrm{Samp}(\nu')$*, then* $\nu = \nu'$.

Proof Recall the construction of $\mathbf{\Pi} \sim \mathrm{Samp}(\nu)$: first one chooses $(s_k)_k \sim \nu$; then one chooses $m_n \in \mathbb{N} \cup \{\infty\}$ i.i.d. from the distribution $(s_1, s_2, \dots, 1 - \sum_k s_k)$; and finally one defines $\ell \sim_{\mathbf{\Pi}} n$ if and only if $m_\ell = m_n \in \mathbb{N}$.

In this construction, after fixing the mass partition $(s_k)_k$, it follows from the Law of Large Numbers that for each $k \in \mathbb{N}$ one has

$$\frac{|\{n \leq N : m_n = k\}|}{N} = \frac{|C_k \cap [N]|}{N} \longrightarrow s_k$$

a.s. in the choice of the sequence $(m_n)_n$, where $C_k = \{n : m_n = k\}$ is the corresponding cell of $\mathbf{\Pi}$. Therefore, it holds a.s. that the asymptotic frequency

$$f(C) := \lim_{N \longrightarrow \infty} \frac{|C \cap [N]|}{N}$$

exists for every cell $C \in \mathbf{\Pi}$, and the set of positive asymptotic frequencies is equal to the set of values s_k which are positive, counted with multiplicities. Hence $(s_k)_k$ is a.s. equal to $(t_k(\mathbf{\Pi}))_k$, defined to be the set of values

$$f(C), \quad C \in \mathbf{\Pi},$$

arranged in non-increasing order. It follows that ν is equal to the law of the sequence $(t_k(\mathbf{\Pi}))_k$ as a function of the random partition $\mathbf{\Pi}$, and so the law of $\mathbf{\Pi}$ determines the law of $(s_k)_k$. $\square$

The situation is not quite so simple for Gram-de Finetti matrices. For example, if μ, μ' are probability measures on Hilbert spaces $\mathfrak{H}$, $\mathfrak{H}'$ such that there is a linear isometry $\Phi : \mathfrak{H} \longrightarrow \mathfrak{H}'$ with $\mu' = \Phi_* \mu$, then one easily calculates that $\mathrm{Samp}(\mu) = \mathrm{Samp}(\mu')$. However, it turns out that this kind of degeneracy, suitably generalized, is the only possibility.

To formulate this, we will use the following notation: if μ is a probability measure on $\mathfrak{H}\times[0,\infty)$, we will write $\mathrm{spt}_1\mu$ for the projection of $\mathrm{spt}\,\mu \subseteq \mathfrak{H}\times[0,\infty)$ onto $\mathfrak{H}$ (or, equivalently, the support of the projection of μ onto $\mathfrak{H}$), and will write

$$\overline{\mathrm{span}}(\mathrm{spt}_1\mu)$$

for the closed subspace of $\mathfrak{H}$ generated by $\mathrm{spt}_1\mu$.

Before giving the main proposition, it is worth proving the following lemma separately.

Lemma 8.2 *There are measurable functions*

$$f_n : [-1,1]^{\binom{n}{2}} \longrightarrow [0,1], \quad n \geq 2,$$

with the following property. Suppose that $\mathfrak{H}$ is a real Hilbert space and ξ_1, ξ_2, ... is a sequence of vectors in $\mathfrak{H}$ such that $\|\xi_i\| \leq 1$ for all i and

$$\xi_1 \in \overline{\mathrm{span}}(\xi_2, \xi_3, \dots).$$

Then the values

$$f_n\big((\xi_i \cdot \xi_j)_{1\leq i<j\leq n}\big)$$

converge to $\|\xi_1\|$.

Proof By enlarging $\mathfrak{H}$ if necessary, we may assume that e_1, e_2, ... is an orthonormal sequence which is also orthogonal to every ξ_i. Now let

$$\zeta_i := \xi_i + (\sqrt{1-\|\xi_i\|^2})e_i \quad \text{for each } i \geq 1,$$

so that $\|\zeta_i\| = 1$ and

$$\zeta_i \cdot \zeta_j = \xi_i \cdot \xi_j \quad \text{if } i \neq j.$$

Since ξ_1 lies in $\overline{\mathrm{span}}(\xi_2, \xi_3, \dots)$, it is equal to its projection onto that subspace. That, in turn, is equal to the projection of ζ_1 onto $\overline{\mathrm{span}}(\zeta_2, \zeta_3, \dots)$, since the vectors e_i are orthogonal to each other and to everything else.

However, for the ζ_is (unlike for the ξ_is) we know that all their lengths are equal to 1. Therefore, by implementing the Gram-Schmidt procedure and computing the resulting change of basis, for any $n \geq 2$ the length of the projection of ζ_1 onto $\mathrm{span}(\zeta_2, \dots, \zeta_n)$ is given by a measurable function f_n of the inner products

$$\zeta_i \cdot \zeta_j = \xi_i \cdot \xi_j, \quad 1 \leq i < j \leq n.$$

Letting $n \longrightarrow \infty$ gives the result. $\square$

Proposition 8.3 *Suppose that $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ are random probability measures on the respective spaces $\mathfrak{H} \times [0,\infty)$ and $\mathfrak{H}' \times [0,\infty)$ such that $\mathrm{Samp}(\boldsymbol{\mu}) = \mathrm{Samp}(\boldsymbol{\mu}')$. Then there is a coupling of random variables $(\boldsymbol{\mu}, \boldsymbol{\mu'}, \boldsymbol{\Phi})$ in which*

- *$\boldsymbol{\Phi}$ is almost surely a linear isometry*

$$\overline{\mathrm{span}}(\mathrm{spt}_1\boldsymbol{\mu}) \longrightarrow \overline{\mathrm{span}}(\mathrm{spt}_1\boldsymbol{\mu}');$$

- *one has*

$$(\boldsymbol{\Phi} \times \mathrm{id}_{[0,\infty)})_*\boldsymbol{\mu} = \boldsymbol{\mu}' \quad a.s..$$

As for the Dovbysh-Sudakov Theorem, for the proof we restrict to the special case $|R_{i,j}| \leq 1$, $R_{i,i} \equiv 1$. Note that this does *not* imply $a_i = 0$. The general case is treated in [Panar], but for the last step in the proof below I have taken a different route from Panchenko, using an idea of Vershik from the more general setting of exchangeable random metrics on $\mathbb{N}$ (see Section $3\frac{1}{2}.7$ of Gromov [Gro99]).

Proof in special case Recall the definition of the process with law $\mathrm{Samp}(\boldsymbol{\mu})$: first one samples $\boldsymbol{\mu}$ at random, and then one samples $(\xi_i, a_i)_{i\in\mathbb{N}}$ i.i.d. $\sim \boldsymbol{\mu}$ and forms the matrix

$$R_{i,j} := \xi_i \cdot \xi_j + \delta_{i,j}a_i.$$

If we retain all of the random choices made in this procedure, it actually defines a coupled collection of random variables

$$(\boldsymbol{\mu}, (\xi_i, a_i)_{i\in\mathbb{N}}, (R_{i,j})_{(i,j)\in\mathbb{N}^2}),$$

in which the $R_{i,j}$s are determined by the (ξ_i, a_i)s.

Step 1: First we show that under the joint distribution of these random data, the norms $\|\xi_i\|$, $i \in \mathbb{N}$, are a.s. determined by the matrix $(R_{i,j})_{i,j}$. In case $a_i \equiv 0$ this is obvious, since then $\|\xi_i\|^2 = R_{i,i}$, but in general we must instead make use of the off-diagonal terms of the matrix. The key observation is that almost surely one has

$$\xi_i \in \mathrm{spt}_1\boldsymbol{\mu} \quad \forall i,$$

and given this it also holds almost surely that:

> for every i and every $\varepsilon > 0$ there are infinitely many j such that
>
> $$\|\xi_i - \xi_j\| < \varepsilon.$$

Therefore

$$\xi_i \in \overline{\mathrm{span}}(\xi_j \,:\, j \neq i) \quad \forall i \quad \text{a.s..}$$

Now Lemma 8.2 shows that on this probability-1 event, each of the norms $\|\xi_i\|$ is equal to a limit of measurable functions of the off-diagonal inner products $\xi_i \cdot \xi_j = R_{i,j}$ for $j \neq i$, and hence the norm itself is a measurable function of these off-diagonal R-entries.

Having shown this, it follows that $a_i = R_{i,i} - \|\xi_i\|^2$ is also a measurable function of the R-entries.

Step 2: If one knows $R_{i,j}$ for all $i \neq j$ and also $\|\xi_i\|$ for all i, then these quantities determine the distances

$$\|\xi_i - \xi_j\| = \sqrt{\|\xi_i\|^2 + \|\xi_j\|^2 - 2R_{i,j}}.$$

Therefore, in the sextuple of random data

$$\big(\boldsymbol{\mu}, (\xi_i, a_i)_i, (R_{i,j})_{i,j}, (\|\xi_i\|)_i, (\|\xi_i - \xi_j\|)_{i<j}, (a_i)_i\big),$$

the matrix R a.s. determines all the distances in the fourth and fifth entries and all the values in the sixth entry.

Step 3: Now suppose that $\boldsymbol{\mu}'$ is another random measure giving $\mathrm{Samp}(\boldsymbol{\mu}') = \mathrm{Samp}(\boldsymbol{\mu})$, and form also its collection of random data

$$\big(\boldsymbol{\mu}', (\xi_i', a_i')_i, (R_{i,j}')_{i,j}, (\|\xi_i'\|)_i, (\|\xi_i' - \xi_j'\|)_{i<j}, (a_i')_i\big)$$

in the same way. Our assumption is that $(R_{i,j})_{i,j} \stackrel{\mathrm{d}}{=} (R_{i,j}')_{i,j}$, and so (using that we work on standard Borel spaces) there is a coupling of these random collections of data under which

$$(R_{i,j})_{i,j} = (R_{i,j}')_{i,j} \quad \text{a.s..}$$

Having formed this coupling, Step 2 implies that also

$$(\|\xi_i\|, a_i) = (\|\xi_i'\|, a_i') \quad \forall i \quad \text{and} \quad \|\xi_i - \xi_j\| = \|\xi_i' - \xi_j'\| \quad \forall i,j \quad \text{a.s..}$$

Therefore the random map

$$\boldsymbol{\Phi}^0 : \{0\} \cup \{\xi_i :\ i \geq 1\} \longrightarrow \{0\} \cup \{\xi_i' :\ i \geq 1\}$$

defined by $\boldsymbol{\Phi}^0(0) = 0$ and $\boldsymbol{\Phi}^0(\xi_i) = \xi_i'$ is almost surely an isometry.

On the other hand, by the Law of Large Numbers, another almost sure event is that the empirical distributions of the sequence $(\xi_i, a_i)_i$ are tight and satisfy

$$\overline{\{\xi_i : i \geq 1\}} = \mathrm{spt}_1\boldsymbol{\mu} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} \delta_{(\xi_i, a_i)} \longrightarrow \boldsymbol{\mu}$$

in the vague topology, and similarly for the pairs (ξ_i', a_i') and the random measure $\boldsymbol{\mu}'$.

Finally, when these a.s. events both hold and the map $\mathbf{\Phi}^0$ is an isometry, that map may be extended uniquely to a linear isometry

$$\mathbf{\Phi} : \overline{\mathrm{span}}(\mathrm{spt}_1\boldsymbol{\mu}) \longrightarrow \overline{\mathrm{span}}(\mathrm{spt}_1\boldsymbol{\mu}')$$

(since an origin-preserving isometry between subsets of Hilbert spaces uniquely extends to a linear isometry of the subspaces they generate). Now applying $\mathbf{\Phi}$ to the convergence of the empirical distributions gives

$$(\mathbf{\Phi} \times \mathrm{id}_{[0,\infty)})_*\boldsymbol{\mu} = \lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{i=1}^{n} \delta_{(\mathbf{\Phi}(\xi_i), a_i)} = \lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{i=1}^{n} \delta_{(\xi_i', a_i)} = \boldsymbol{\mu}'.$$

This completes the proof. □

9 Comparison of random partitions and Gram-de Finetti matrices

Although Kingman's Paintbox Theorem is much simpler than Dovbysh-Sudakov, it is worth noting that the former is a special case of the latter. This is simply because if $\mathbf{\Pi}$ is an exchangeable random partition, then it defines a $\{0,1\}$-valued Gram-de Finetti matrix by setting

$$R_{i,j} := \begin{cases} 1 & \text{if } i \sim_{\mathbf{\Pi}} j \\ 0 & \text{else} \end{cases}$$

(an easy exercise shows that this is non-negative definite).

Applying the Dovbysh-Sudakov Theorem to this R gives a random measure $\boldsymbol{\mu}$ on $\mathfrak{H} \times [0,\infty)$ such that

$$(R_{i,j})_{i,j} \stackrel{\mathrm{d}}{=} (\xi_i \cdot \xi_j + \delta_{ij} a_i)_{i,j}$$

for an i.i.d.($\boldsymbol{\mu}$) sample $(\xi_i, a_i)_i$. Since R is $\{0,1\}$-valued, this implies that $\xi \cdot \xi'$ lies in $\{0,1\}$ almost surely when ξ, ξ' are drawn independently from $\boldsymbol{\mu}$. By some simple analysis, this is possible only if the random set $\mathrm{spt}_1\boldsymbol{\mu}$ almost surely consists of either a finite or infinite orthonormal sequence, or an orthonormal sequence together with the origin, or just the origin.

In the third case one has $R_{i,j} = 0$ whenever $i \neq j$, so this corresponds to the trivial partition of $\mathbb{N}$ into singletons. In either of the first two cases, let $(e_k)_k$ be the orthonormal sequence ordered so that the weights $s_k := \boldsymbol{\mu}\{e_k\}$ are non-increasing, and made infinite by including extra vectors if necessary. The support $\mathrm{spt}_1\boldsymbol{\mu}$ also contains 0 precisely when $\sum_k s_k < 1$, in which case $\boldsymbol{\mu}\{0\} = 1 - \sum_k s_k$. Rewriting the representation of R as $\mathrm{Samp}(\boldsymbol{\mu})$ in terms of $\boldsymbol{\Pi}$, we find that the random sequence of weights $(s_k)_k$ is precisely the random mass partition that directs $\boldsymbol{\Pi}$ according to the paintbox construction.

Concerning the uniqueness results of the preceding section, one sees a closer parallel by choosing a slightly different formulation of paintbox processes. If one does not insist that mass partitions be non-increasing, then two mass partitions $(s_k)_k$, $(s'_k)_k$ give the same paintbox process if and only if one is a re-ordering of the other. This is the analog of the redundancy that we found for Gram-de Finetti matrices driven by Hilbert space measures, except that the relevant symmetry group is $\mathrm{Sym}(\mathbb{N})$ rather than the orthogonal group of $\mathfrak{H}$. With this less restrictive notion of random mass partitions, two measures $\nu, \nu' \in \Pr \mathcal{P}_{\mathrm{m}}$ give $\mathrm{Samp}(\nu) = \mathrm{Samp}(\nu')$ if and only if there is a coupling $\lambda \in \Pr(\mathcal{P}_{\mathrm{m}} \times \mathcal{P}_{\mathrm{m}})$ of ν and ν' such that λ is supported on the pairs $((s_k)_k, (s'_k)_k)$ for which $(s'_k)_k$ is a re-ordering of the sequence $(s_k)_k$.

10 Other symmetries for stochastic processes

Several other symmetry principles for the laws of stochastic processes have been studied by methods more-or-less similar to those above, often using the basic Structure Theorem 2.2 to do the heavy lifting and then adding some refinements, as we did for partitions and Garm-de Finetti matrices.

Some well-known examples, with references, include:

- contractible sequences and arrays ([Ald85, Section 6] and [Kal05, Chapters 1 and 7]);
- separately (or 'row-column') exchangeable arrays ([Ald85, Section 14] and

[Kal05, Chapters 1 and 7]);

- notions of exchangeability for continuous-time processes ([Ald85, Section 10] and [Kal05, Section 1.3]);
- tree-indexed processes which have the symmetries of the tree [Ald85, Section 13];
- rotatable arrays ([Ald85, Subsection 15.7] and [Kal05, Chapter 8]);
- exchangeable random sets [Ald85, Section 17] and [Kal05, Chapter 6]);
- invariant point processes [Ald85, Subsection 21.2], and more generally symmetric random measures on rectangles in Euclidean spaces [Kal05, Chapter 9].

Part II

Spin glasses

11 Introduction to spin glasses

Some terminology from physics:

'**Glass**': a material which is hard and inflexible (like a solid) but is not completely ordered or lattice-like in its microscropic structure (thus, unlike a crystal or simple metals).

'**Spin**': pertaining to the magnetic spins of ions in a material.

'**Spin glass**': a material, effectively solid, which contains some irregular distribution of magnetizable ions.

Basic laboratory examples: Cadmium telluride (a non-magnetizable crystalline compound of cadmium and tellurium) doped with some easily-magnetized atoms, e.g. of iron or nickel.

Spin glasses are complicated because the magnetic interactions between ions depend heavily on the exact distance between them, the irregular distances in a spin glass give rise to irregular interactions and hence a locally highly-complicated response to an external magnetic field.

The basic model

In order to model a spin glass, consider the state space $\{-1,1\}^N$, where an element $\sigma = (\sigma_n)_{n\leq N}$ is interpreted as an assignment of spins to each of N magnetic ions. In this model each spin can only be either 'up' or 'down'; more sophisticated models might use $(\mathrm{S}^2)^N$ or suchlike.

Basic procedure of modelling in statistical physics: determine (from physical considerations) a 'Hamiltonian'

$$H : \{-1,1\}^N \longrightarrow \mathbb{R},$$

with the interpretation that $H(\sigma)$ is the internal magnetic energy of the system when it is in state σ. Now suppose the material interacts with its environment at some temperature T, so that its state keeps changing due to microscopic interactions. Then the basic prescription from thermodynamics is that the *proportion of time spent in different states* is given by the **Gibbs measure at temperature** T:

$$\gamma_\beta\{\sigma\} := \frac{1}{Z(\beta)}\exp(-\beta H(\sigma)),$$

where $\beta := 1/T$ and

$$Z(\beta) := \sum_\sigma \exp(-\beta H(\sigma))$$

is the normalizing constant, which is called the **partition function**.

Following standard practice in physics, we will sometimes use $\langle - \rangle_\beta$ to denote an average over any number of independent states drawn from γ_β. For instance, if $f : (\{-1,1\}^N)^2 \longrightarrow \mathbb{R}$ then

$$\langle f\rangle_\beta := \int_{(\{-1,1\}^N)^2} f(\sigma^1,\sigma^2)\,\gamma_\beta^{\otimes 2}(\mathrm{d}\sigma^1,\mathrm{d}\sigma^2). \tag{9}$$

Subscripts such as 'β' may also be dropped when this can cause no confusion. When several independently-chosen states are invoked, as here, they are sometimes referred to as 'replicas'.

In most sensible models, $Z(\beta)$ is exponentially large as a function of N, and the quantity of greatest interest is the first-order behaviour of the exponent. For this reason one now defines the **free energy** to be the quantity

$$\frac{1}{N}\log Z(\beta).$$

In order to understand the material in thermal equilibrium, one wishes to describe the measure γ_β or the quantity $F(\beta)$ as well as possible.

The key feature of a spin glass is that different pairs of spins interact in very different ways. This is reflected by a Hamiltonian of the form

$$H(\sigma) = \sum_{i,j} g_{ij}\sigma_i\sigma_j$$

in which the interaction constants g_{ij} vary irregularly with the pair $\{i, j\}$ and can take either sign, so some pairs prefer to be aligned and some anti-aligned. Notice that among three indices i, j and k, it can happen that two pairs prefer alignment but the third prefers anti-alignment, or that all three prefer anti-alignment, in which case the assignment of spins that minimizes H may be far from obvious. This phenomenon (and extension to larger numbers of indices) is called **frustration**.

In realistic models, the indices i, j are locations in space, and one assumes that $g_{i,j} = 0$ if $|i - j|$ is large. The irregularity appears for nearby i, j. Among nonzero interactions, a simple way to produce a spin glass model is to *choose the interaction constants themselves at random* (so now there are two levels of randomness involved). The basic model here is the Edwards-Anderson model [MPV87]. Almost nothing is known rigorously about this model.

To simplify, one can consider a mean field model, in which one ignores the spatial locations of the spins. Most classical is the **Sherrington-Kirkpatrick ('SK') model** (see the papers of Sherrington and Kirkpatrick in [MPV87]): let g_{ij} be independent standard Gaussians on some background probability space $(\Omega, \mathcal{F}, \mathsf{P})$, and let H be the random function

$$H(\sigma) := \frac{1}{\sqrt{N}} \sum_{ij} g_{ij}\sigma_i\sigma_j. \tag{10}$$

The normalization is chosen so that each of the random variables $H(\sigma)$ has variance N, which turns out to be the regime of greatest interest. This model is still very complicated, but recent work has thrown considerable light onto its structure.

An even simpler toy model, which nevertheless begins to show some interesting behaviour, is the **Random Energy Model ('REM') model**, introduced by Derrida [Der81]. In this case one lets the values $H(\sigma)$ be centred Gaussians of variance N and all simply independent for different σ.

Having chosen one of these models, let $\gamma_{\beta,N}$ be the random Gibbs measure on $\{-1, 1\}^N$ resulting from this Hamiltonian, and let $F_N(\beta)$ be the **expected free**

energy:

$$F_N(\beta) := \mathsf{E}\frac{1}{N}\log\sum_\sigma \exp(-\beta H(\sigma)). \tag{11}$$

Basic (vague) question: What are the values of $F_N(\beta)$ or the typical structure of $\gamma_{\beta,N}$ as functions of these random interactions?

Recommended reading: [ASS07, Tal03, Pan10, Pan12, Panar]. A new version of Talagrand's comprehensive book [Tal03] is in preparation, with Volume 1 already available [Tal11]. The classic physicists' text on spin glasses is [MPV87].

Connection to random optimization

When physicists choose to study a mean-field model of a situation which in the real world involves some spatial variation, they do so simply because the mean-field model is simpler. Their hope is that it will still reveal some non-trivial structure which can then suggest fruitful questions to ask about a more realistic model. For instance, the Curie-Weiss model exhibits a phase transition much like the Ising model in two or three dimensions.

In fact, it remains contentious whether the SK model is worthwhile as a toy version of spatially-extended models. However, it was quickly realized that mean-field spin glass models have a much more solid connection with random optimization.

Consider again the random function $H : \{-1,1\}^N \longrightarrow \mathbb{R}$ defined in (10). When β is large, the Gibbs measure $\gamma_{\beta,N}$, for which $\gamma_{\beta,N}\{\sigma\}$ is proportional to $\exp(-\beta H(\sigma))$, should be concentrated on those configurations $\sigma \in \{-1,1\}^N$ where $(-H(\sigma))$ is large [3]. Also, trivial estimates give

$$\exp(\beta \max_\sigma(-H(\sigma))) \leq \sum_\sigma \exp(-\beta H(\sigma)) \leq 2^N \exp(\beta\max_\sigma(-H(\sigma))),$$

and hence

$$\mathsf{E}\frac{\beta}{N}\max_\sigma(-H(\sigma)) \leq F_N(\beta) \leq \log 2 + \mathsf{E}\frac{\beta}{N}\max_\sigma(-H(\sigma)).$$

Dividing by β and letting $\beta \longrightarrow \infty$, we conclude that

$$\mathsf{E}\frac{1}{N}\max_\sigma(-H(\sigma)) = \lim_{\beta\longrightarrow\infty}\frac{F_N(\beta)}{\beta},$$

[3] Unfortunately the sign conventions of statistical physics mean that we will be working with $(-H)$, rather that H, throughout this discussion of optimization.

with a rate of convergence that does not depend on N. Therefore we may take expectations, let $N \longrightarrow \infty$ and change the order of the limits, and hence obtain

$$\lim_{N\longrightarrow\infty} \mathsf{E}\frac{1}{N}\max_{\sigma}(-H(\sigma)) = \lim_{\beta\longrightarrow\infty}\frac{F(\beta)}{\beta} \quad \text{with high probability,}$$

where $F(\beta)$ is the limiting free energy introduced above. Moreover, in many cases the random quantity $\frac{1}{N}\max_{\sigma}(-H(\sigma))$ is known to concentrate around this limiting value as $N \longrightarrow \infty$ as a result of Gaussian concentration phenomena.

Thus, if we have a formula for $F(\beta)$, then this gives a formula for the leading-order behaviour of the random optimization problem $\max_{\sigma}(-H(\sigma))$. This amounts to determining the maximum over σ of a typical instance of the random function

$$\frac{1}{\sqrt{N}}\sum_{i,j=1}^{N} g_{ij}\sigma_i\sigma_j$$

(dropping a minus-sign now, since $-g_{ij} \overset{\mathrm{d}}{=} g_{ij}$). This is an instance of the classical *Dean's problem*: given a population of individuals $\{1, 2, \dots, N\}$ in which the like or dislike between individuals i and j is given by the (positive or negative) value g_{ij}, the Dean would like to separate them into two classes (which will correspond to $\{i : \ \sigma_i = 1\}$ and $\{i : \ \sigma_i = -1\}$) so that the overall level of comfort is maximized: that is, s/he needs the optimum way to put pairs that like each other into the same class, and pairs that dislike each other into different classes.

Amazingly, a good enough understanding of the SK model allows one to give an expression for $F(\beta)$ (albeit very complicated), so that in principle one could actually estimate $\lim_{\beta\longrightarrow\infty} F(\beta)/\beta$ this way and so obtain the correct constant c in the estimate

$$\max_{\sigma} H(\sigma) = cN + \mathrm{o}(N).$$

The expression for $F(\beta)$ is called the Parisi formula, and is given later in these notes. Although not simple, it can be estimated numerically, and using this the physicists made the prediction that c (exists and) is approximately $0.7633\dots$. The Parisi formula is now known rigorously [Tal06, Panar], so this is now a theorem up to the quality of those numerical methods.

What's more, the exact distribution of the random coefficients g_{ij} is not essential for this calculation. Provided these r.v.s have mean zero, have variance one and have bounded third moments, the resulting behaviour of the free energy as $N \longrightarrow \infty$ will be the same: see Carmona and Hu [CH06].

This means that other random variants of the Dean's problem can be solved this way. Perhaps the most classical is an instance of MAX-CUT (whose exact solution for deterministic graphs is NP-complete in general). In this case the pairwise interactions g_{ij} are not Gaussian, but take two values according to the adjacency matrix of an Erdős-Rényi random graph $\mathcal{G}(N,p)$, with those two values chosen to give mean zero and variance one. A concentration argument shows that the expected best cut must cut $N^2p/4 + \mathrm{o}(N^2)$ of the edges, but analysis of this spin-glass model gives a value for the next term. The prediction of the physicists is that for this random MAX-CUT problem one has

$$\mathrm{OPT} = N^2p/4 + \frac{\sqrt{p(1-p)}}{2}cN^{3/2} + \mathrm{o}(N^{3/2}) \quad \text{w.h.p.}$$

where again $c = 0.7633\dots$. With the rigorous proof fo the Parisi formula, this, too, is now a theorem, up to the quality of some numerical methods. See [MPV87, Chapter IX] for a more complete account of applications to random optimization.

12 Describing asymptotic structure

We return to our basic questions about the SK model. As shown in the previous section, the application to random optimization will be served by having a good enough asymptotic formula for $F_N(\beta)$. However, we will see that this problem is intimately related to understanding the geometry of the Gibbs measures $\gamma_{\beta,N}$.

First we must decide in what terms to try to describe $\gamma_{\beta,N}$. The basic idea here is that two configurations $\sigma, \sigma' \in \{-1,1\}^N$ are similar if they agree in most coordinates. More formally, this means that they are close in Hamming metric, or, equivalently, in any of the metrics inherited by regarding $\{-1,1\}^N$ as a subset of ℓ_p^N (i.e., $\mathbb{R}^N$ with the ℓ_p-norm) for any fixed choice of $p \in [1,\infty)$.

In fact, in this setting it is most natural to think of $\{-1,1\}^N$ as a subset of ℓ_2^N, because that space captures the structure of the random variables $H(\sigma)$, and specifically their covariances. To see this, compute

$$\mathrm{Cov}(H(\sigma), H(\sigma')) = \mathsf{E}H(\sigma)H(\sigma') = \frac{1}{N}\sum_{ij,i'j'} \mathsf{E}(g_{ij}g_{i'j'})\sigma_i\sigma_j\sigma'_{i'}\sigma'_{j'}.$$

Since $\mathsf{E}(g_{ij}g_{i'j'})$ is zero unless $ij = i'j'$, because the interactions are independent, this simplifies to

$$\mathrm{Cov}(H(\sigma), H(\sigma')) = \frac{1}{N}\sum_{ij}\sigma_i\sigma'_i\sigma_j\sigma'_j = \frac{1}{N}\Big(\sum_i \sigma_i\sigma'_i\Big)^2 = N(\sigma\cdot\sigma')^2,$$

where $\sigma \cdot \sigma' := \frac{1}{N}\sum_i \sigma_i \sigma'_i$.

So the covariances of the random function H are given by the structure of $\{-1,1\}^N$ as a subset of Hilbert space; in particular, the problems of estimating the structure of $\gamma_{\beta,N}$ and the value of $F_N(\beta)$ are unchanged if we change $\{-1,1\}^N$ by a rigid rotation in ℓ_2^N. Motivated by this, we will think of $\gamma_{\beta,N}$ as a random probability measure on a Hilbert space, and try to give a 'coarse' description of it as such.

By a 'coarse' description of $\gamma_{\beta,N}$, we really want an idea of the limiting behaviour of $\gamma_{\beta,N}$ as $N \longrightarrow \infty$ in terms of some meaningful notion of convergence for random probability measures on Hilbert spaces. Since we really care only about the structure of $\gamma_{\beta,N}$ up to orthogonal rotations of ℓ_2^N, convergence of the associated Gram-de Finetti matrices obtained by sampling offers an ideal such notion:

> Do the random measures $\gamma_{\beta,N}$ sampling-converge, and if so what is their limit?

13 More tools: facts about Gaussians

Before proceeding with spin glasses, we need to describe two basic tools from the study of Guassian processes. The first is a concentration inequality; see, for instance, Ledoux [Led01].

Proposition 13.1 *Suppose that $F : \mathbb{R}^M \longrightarrow \mathbb{R}$ is a function such that*

$$|F(x) - F(y)| \leq K\|x - y\| \quad \forall x, y \in \mathbb{R}^M,$$

and let $g = (g_1, \dots, g_M)$ be a sequence of independent standard Gaussian r.v.s. Then for each $t > 0$ we have

$$\mathsf{P}(|F(g) - \mathsf{E}F(g)| \geq t) \leq 2\exp\Big(-\frac{t^2}{2K^2}\Big).$$

Our main application of this is to the free energy $\frac{1}{N}\log Z_N(\beta)$, whose expectation is $F_N(\beta)$ (equation (11)). In the case of the SK model with Hamiltonian (10), a little calculus gives

$$\frac{\partial F_N(\beta)}{\partial g_{i,j}} = \frac{-\beta}{N^{3/2}}\langle \sigma_i \sigma_j \rangle_{\beta,N}$$

(regarding $g_{i,j}$ as just a variable here, rather than a sample from a Gaussian), where we recall that the notation '$\langle - \rangle_{\beta,N}$' refers to an average over any number of independently-drawn samples from $\gamma_{\beta,N}$; see the discussion around equation (9). From this one easily deduces that $F_N(\beta)$ is $(\beta/\sqrt{N})$-Lipschitz as a function of $(g_{i,j})_{i,j}$. Using this in Proposition 13.1, we conclude that

$$\mathsf{P}\Big(\Big|F_N(\beta) - \frac{1}{N}\log Z_N(\beta)\Big| \geq t\Big) \leq 2\exp\Big(-\frac{Nt^2}{2\beta}\Big). \tag{12}$$

Similarly, in the case of the REM we may write $H(\sigma) = \sqrt{N}g_\sigma$, where each g_σ will be drawn independently from a standard Gaussian. Therefore in this case one obtains

$$\frac{\partial F_N(\beta)}{\partial g_{i,j}} = \frac{-\beta}{N^{1/2}}\langle \delta_\sigma \rangle_{\beta,N},$$

which again gives a Lipschitz constant of at most $\beta/\sqrt{N}$, and so the same concentration inequality (12).

The second tool we introduce here is Gaussian integration by parts. This simple piece of calculus has become ubiquitous in the study of spin glass models, as well as many others in statistical physics.

Proposition 13.2 (Gaussian integration by parts) *If g is a centred Gaussian r.v. and $F : \mathbb{R} \longrightarrow \mathbb{R}$ is smooth and of moderate growth at ∞ (polynomial growth is fine), then*

$$\mathsf{E}gF(g) = \mathsf{E}(g^2)\mathsf{E}F'(g).$$

□

This follows by a basic integration by parts using the density of the Gaussian, since

$$x\mathrm{e}^{-x^2/2} = \frac{\mathrm{d}}{\mathrm{d}x}(-\mathrm{e}^{-x^2/2}).$$

In the study of spin glasses, Gaussian integration by parts is often used as part of 'Gaussian interpolation'. Suppose g and g' are M-dimensional Gaussian r.v.s with different variance-covariance matrices, $F : \mathbb{R}^M \longrightarrow \mathbb{R}$ is a continuously differentiable function which does not grow too fast near ∞, and one wishes to compare the expectations $\mathsf{E}F(g)$ and $\mathsf{E}F(g')$. To use the interpolation method, one chooses an interpolating family of M-dimensional Gaussian r.v.s g_t with $t \in [0,1]$

such that $g_0 \overset{\mathrm{d}}{=} g$ and $g_1 \overset{\mathrm{d}}{=} g'$, and then uses Gaussian integration by parts to evaluate (or at least estimate)

$$\frac{\partial}{\partial t}\mathsf{E}F(g_t).$$

Of course, there are many possibly ways to choose the interpolating family, and often the success of the method depends on a very clever choice; this is why Talagrand calls it the 'smart path method'. One of its earliest successes was the following basic result of Guerra and Toninelli; see [GT02] for the smart choice of path.

Lemma 13.3 *In the Sherrington-Kirkpatrick model at a fixed value of β,*

$$\mathsf{E}\log Z_{N+M} \leq \mathsf{E}\log Z_N + \mathsf{E}\log Z_M.$$

$\square$

Using Fekete's Lemma, this has the crucial consequence that $\lim_{N\longrightarrow\infty} F_N(\beta)$ exists for all β.

14 The Aizenman-Sims-Starr scheme and the Parisi formula

Suppose that $\gamma_{\beta,N}$ is known to sampling-converge to some limiting random measure γ_β. Building on a calculational method of Guerra [Gue03], Aizenman, Sims and Starr [ASS07] showed how this γ_β then provides a formula for the limiting free energy $F(\beta)$. This insight made it possible to put Parisi's original predictions about the SK model (see below) into a more general mathematical framework, and is the basis for an approach to the Parisi formula via understanding the geometry of γ_β.

The first idea here is to write

$$F_N = \frac{1}{N}\mathsf{E}\log Z_N = \frac{1}{N}\sum_{i=0}^{N-1} A_i$$

with $A_i = \mathsf{E}\log Z_{i+1} - \mathsf{E}\log Z_i$. If one can show that these quantities A_i tend to a constant, then of course F_N will also tend to that constant. In a sense, this realizes F_N as the 'logarithmic increment' in the growing sequence of partition functions Z_N.

So now let us try to compare $\mathsf{E}\log Z_N$ with $\mathsf{E}\log Z_{N+1}$. First, identify $\{-1,1\}^{N+1}$ with $\{-1,1\}^N \times \{-1,1\}$, and for $(\sigma,\varepsilon) \in \{-1,1\}^N \times \{-1,1\}$ write

$$H_{N+1}(s,\varepsilon) = H'_N(\sigma) + \varepsilon z_N(\sigma) + \frac{1}{\sqrt{N+1}} g_{(N+1)(N+1)},$$

where

$$H'_N(\sigma) = \frac{1}{\sqrt{N+1}} \sum_{i,j=1}^N g_{ij}\sigma_i\sigma_j$$

and

$$z_N(\sigma) = \frac{1}{\sqrt{N+1}} \sum_{i=1}^N (g_{i(N+1)} + g_{(N+1)i})\sigma_i. \tag{13}$$

It is easy to show that the last term in this decomposition of H_{N+1} makes asymptotically negligible contribution to F_N, so we now ignore it. Next, $H'_N(\sigma)$ is almost the same as $H_N(\sigma)$: only the coefficient is slightly wrong. Since a sum of two independent Gaussians is still Gaussians, as random variables we can correct this with a small extra term:

$$H_N(\sigma) = H'_N(\sigma) + y_N(\sigma),$$

where

$$y_N(\sigma) = \frac{1}{\sqrt{N(N+1)}} \sum_{i,j=1}^N g'_{ij}\sigma_i\sigma_j, \tag{14}$$

where the g'_{ij} are new independent standard Gaussians.

Notice that if we condition on g_{ij} for $i,j \leq N$, then (13) and (14) define two two further independent Gaussian processes on $\{-1,1\}^N$ with covariances given by

$$\mathrm{Cov}(z_N(\sigma), z_N(\sigma')) = 2\frac{N}{N+1}(\sigma\cdot\sigma')$$
$$\text{and} \quad \mathrm{Cov}(y_N(\sigma), y_N(\sigma')) = \frac{N}{N+1}(\sigma\cdot\sigma')^2. \tag{15}$$

With a little analysis one can show that the multiplicative factor of $\frac{N}{N+1}$ is also unimportant here.

Having set up this notation, our desired comparison becomes

$$\begin{aligned}&\mathsf{E}\log Z_{N+1} - \mathsf{E}\log Z_N\\ &= \mathsf{E}\log\sum_\sigma\sum_\varepsilon \exp(-\beta(H'_N(\sigma)+\varepsilon z_N(\sigma))) - \mathsf{E}\log\sum_\sigma \exp(-\beta(H'_N(\sigma)+y_N(\sigma))).\end{aligned}$$

Letting $Z'_N := \sum_\sigma \exp(-\beta H'_N(\sigma))$, we may add and subtract $\mathsf{E}\log Z'_N$ in the above to obtain

$$\begin{aligned}&\mathsf{E}\log Z_{N+1} - \mathsf{E}\log Z_N\\ &= \mathsf{E}\log\frac{1}{Z'_N(\sigma)}\sum_\sigma\sum_\varepsilon \exp(-(\beta H'_N(\sigma) + \beta\varepsilon z_n(\sigma)))\\ &\qquad\qquad -\mathsf{E}\log\frac{1}{Z'_N(\sigma)}\sum_\sigma \exp(-(\beta H'_N(\sigma) + \beta y_N(\sigma)))\\ &= \mathsf{E}\log\int_{\{-1,1\}^N}\sum_{\varepsilon=\pm1}\exp(-\beta\varepsilon z_N(\sigma))\,\gamma'_{\beta,N}(\mathrm{d}\sigma)\\ &\qquad\qquad -\mathsf{E}\log\int_{\{-1,1\}^N}\exp(-\beta y_N(\sigma))\,\gamma'_{\beta,N}(\mathrm{d}\sigma)\\ &= \mathsf{E}\log\int_{\{-1,1\}^N} 2\cosh(-\beta z_N(\sigma))\,\gamma'_{\beta,N}(\mathrm{d}\sigma)\\ &\qquad\qquad -\mathsf{E}\log\int_{\{-1,1\}^N}\exp(-\beta y_N(\sigma))\,\gamma'_{\beta,N}(\mathrm{d}\sigma),\end{aligned}$$

where $\gamma'_{\beta,N}$ is the random Gibbs measure on $\{-1,1\}^N$ corresponding to the Hamiltonian H'_N (which is not conceptually different from the Gibbs measure for H_N), and where the expectation is in both $\gamma'_{\beta,N}$ and the independent random variables $z_N(\sigma)$, $y_N(\sigma)$.

Importantly, one can (at least formally) make sense of this last expression for *any* random Hilbert space measure γ. Suppose γ is such a measure, say on the unit ball B of a Hilbert space. We need independent Gaussian random linear functionals

$$z : \operatorname{spt}\gamma \longrightarrow \mathbb{R} \quad \text{and} \quad y : \operatorname{spt}\gamma \longrightarrow \mathbb{R}$$

with covariances as in (15) (ignoring the factor of $\frac{N}{N+1}$):

$$\operatorname{Cov}(z(\xi), z(\xi')) = 2(\xi\cdot\xi'), \quad \operatorname{Cov}(y(\xi), y(\xi')) = (\xi\cdot\xi')^2.$$

Provided γ and its support are not too irregular, one can construct such random functionals using the theory of Gaussian Hilbert spaces, essentially uniquely. In terms of these one may now write down the analog of the above expression:

$$\mathsf{E}' \log \int_B 2\cosh(-\beta z(\xi))\,\gamma(\mathrm{d}\xi) - \mathsf{E}' \log \int_B \exp(-\beta y(\xi))\,\gamma(\mathrm{d}\xi),$$

where E' denotes expectation in all of the random data γ, z and y (ignoring issues of integrability here). Since the laws of z and y are determined by their covariances, this quantity is really a functional of the law of the random measure γ. We will write it as $\Phi(\mathrm{law}\,\gamma)$.

The usefulness of this is as follows: if we can describe a sampling-limit random measure γ_β for $\gamma'_{\beta,N}$, then at least heuristically it should follow that A_N tends to the limiting value $\Phi(\mathrm{law}\,\gamma_\beta)$. This isn't quite immediate, since one must prove that the sampling-convergence $\gamma'_{\beta,N} \longrightarrow \gamma_\beta$ is strong enough to imply the convergence of these Φ-values. However, that continuity can be proved using more machinery from Gaussian processes: it is implied by the following, which is Theorem 1.3 in [Panar].

Theorem 14.1 *For each $\varepsilon > 0$, there are $n \geq 1$ and a continuous function $F_\varepsilon : [-1,1]^{n^2} \longrightarrow \mathbb{R}$ such that*

$$\Big|\Phi(\mathrm{law}\,\gamma) - \mathsf{E}_\gamma \int_{B^n} F_\varepsilon\big((\xi^i \cdot \xi^j)_{i,j \leq n}\big)\,\gamma^{\otimes N}(\mathrm{d}\xi^1, \dots, \mathrm{d}\xi^N)\Big| \leq \varepsilon$$

for all random measures γ on B. □

Thus, if we knew the sampling convergence of $\gamma'_{\beta,N}$ to γ_β, it would follow that

$$\lim_{N \longrightarrow \infty} F_N(\beta) = \Phi(\mathrm{law}\,\gamma_\beta).$$

This is the precise sense in which a good enough understanding of the asymptotic structure of $\gamma_{\beta,N}$ (or, to be precise, the very-similar $\gamma'_{\beta,N}$) would give the asymptotic value of $F_N(\beta)$.

Unfortunately, this convergence is not known. However, recent work of Panchenko has yielded results almost as good: with some tweaking, the possible subsequential sampling-limits of the sequence $(\gamma'_{\beta,N})_{N \geq 1}$ have been restricted to a very precise family called the Ruelle Probability Cascades. This restriction is enough to express $\lim_{N \longrightarrow \infty} F_N(\beta)$ in terms of a rather more concrete variational problem. If we temporarily hide some important technical issues, an overview of the argument is as follows. For a proper treatment of these and related ideas, see [ASS07, AC, Pan12].

Let $\mathcal{M}$ denote the space of all laws of random probability measures on B, and let $\mathcal{M}_{\mathrm{lim}} \subseteq \mathcal{M}$ be the set of subsequential limits of the sequence $(\gamma'_{\beta,N})_{N\geq 1}$ (all taken up to random orthogonal rotations, in view of Proposition 8.3). The definition of A_N certainly gives

$$\lim_{N\longrightarrow\infty} F_N(\beta) \geq \liminf_{N\longrightarrow\infty} A_N,$$

so in view of the above reasoning this implies

$$\lim_{N\longrightarrow\infty} F_N(\beta) \geq \inf_{\mathrm{law}\,\gamma\in\mathcal{M}_{\mathrm{lim}}} \Phi(\mathrm{law}\,\gamma). \tag{16}$$

On the other hand, for any $M \geq 1$, similar steps to the above allow one to express the difference

$$\mathsf{E}\log Z_{N+M} - \mathsf{E}\log Z_N$$

as a (slightly more complicated) functional Φ_M applied to the random Gibbs measure $\gamma''_{\beta,M,N}$ with Hamiltonian

$$H''_{N,M}(\sigma) = \frac{1}{\sqrt{N+M}}\sum_{i,j=1}^{N} g_{ij}\sigma_i\sigma_j.$$

This new measure $\gamma''_{\beta,M,N}$ should still be very close to $\gamma'_{\beta,N}$ if $N \gg M$.

For this functional, Aizenman, Sims and Starr showed in [ASS07] that one can prove the inequality

$$\Phi_M(\mathrm{law}\,\gamma''_{\beta,N,M}) \leq \Phi_M(\mathrm{law}\,\gamma)$$

for any other random measure γ on B. Their proof uses a clever interpolation method based on Gaussian integration by parts, abstracted from a crucial earlier insight of Guerra [Gue03], which will not be explained here.

For a certain special subclass of laws $\mathcal{M}_{\mathrm{ss}} \subseteq \mathcal{M}$ referred to as 'stochastically stable' (which will not be defined here), one has $\Phi_M = M\Phi$, and hence if $\mathrm{law}\,\gamma \in \mathcal{M}_{\mathrm{ss}}$ then

$$\frac{1}{M}\big(\mathsf{E}\log Z_{N+M} - \mathsf{E}\log Z_N\big) \leq \Phi(\mathrm{law}\,\gamma).$$

Note this seems to go the other way from (16).

However, the Ruelle Probability Cascades *are* all stochastically stable, so Panchenko's results essentially give $\mathcal{M}_{\mathrm{lim}} \subseteq \mathcal{M}_{\mathrm{ss}}$. Therefore we may combine the above inequalities to deduce

$$\lim_{N\longrightarrow\infty} F_N(\beta) = \inf_{\mathrm{law}\,\gamma\in\mathcal{M}_{\mathrm{lim}}} \Phi(\mathrm{law}\,\gamma),$$

where this infimum really runs only over the Ruelle Probability Cascades. For those special random measures, the functional Φ may be written out rather more explicitly, giving the famous Parisi formula:

Theorem 14.2 (The Parisi formula, formulated as in [ASS07]) *As $N \longrightarrow \infty$, the random quantities $F_N(\beta)$ converge in probability to the deterministic quantity*

$$F(\beta) := \inf_{\varphi} \mathcal{P}(\varphi, \beta),$$

where the infimum is taken over all right continuous non-decreasing functions $[0,1] \longrightarrow [0,1]$, where

$$\mathcal{P}(\varphi, \beta) := \ln(2) + f(0,0;\varphi) - \frac{\beta^2}{2} \int_0^1 q\varphi(q)\,\mathrm{d}q,$$

and where $f(q,y;\varphi)$ for $(q,y) \in [0,1]^2$ is the solution to the PDE

$$\frac{\partial f}{\partial q} + \frac{1}{2}\Big(\frac{\partial^2 f}{\partial y^2} + \varphi(q)\Big(\frac{\partial f}{\partial y}\Big)^2\Big) = 0$$

subject to the boundary condition

$$f(1, y; \varphi) \equiv \ln(\cosh(\beta y)).$$

□

This extraordinary conclusion was contained among Parisi's original predictions for this model. Before Panchenko was able to complete the program outlined above, the Parisi formula was first proved by Talagrand in [Tal06] using a different set of very subtle estimates (still mostly obtained from the Gaussian interpolation method). However, that earlier proof of Talagrand does not give so much information on the structure of the Gibbs measures, and I will not discussed it further here.

The Parisi formula still looks very complicated. What is important to understand, however, is that neither the PDE nor the variational problem over φ that are involved in it is too difficult to approximate numerically, and so this gives a relatively 'simple' way to estimate $\lim_{N\longrightarrow\infty} F_N(\beta)$, and hence also $\frac{1}{N}\max_\sigma(-H_N(\sigma))$. By contrast, directly estimating this maximum for a computer simulation of the random variables H is prohibitively difficult for even moderately large N. (On the other hand, it is still largely open to understand rigorously how continuous is the functional $\mathcal{P}(\varphi, \beta)$ in its argument φ.)

Several important technical points have been ignored in the above discussion. Perhaps the most serious is that Panchenko does not prove that all limits of the sequence $(\gamma'_{\beta,N})_{N\geq 1}$ are Ruelle Probability Cascades for the Sherrington-Kirkpatrick Hamiltonian itself. Rather, he proves that this holds 'typically' in a small neighbourhood of that Hamiltonian for an infinite-dimensional family of perturbations of it. This will be explained in a little more detail below in connexion with the Ghirlanda-Guerra identities, which are the principal ingredient in Panchenko's proof. The key point is that one can find such perturbations such that:

- on the one hand, they do satisfy all of these identities, so that Panchenko's argument gives the Ruelle-Probability-Cascade result;
- but on the other, they are asymptotically close enough to the unperturbed Sherrington-Kirkpatrick Hamiltonian that their specific free energies have the same leading-order behaviour in N, so that they still give the correct evaluation of $\lim_{N\longrightarrow\infty} F_N(\beta)$ for the Sherrington-Kirkpatrick model itself.

The last two sections of these notes will offer a rough discussion of the Ghirlanda-Guerra identities and their key geometric consequence: ultrametricity.

15 The Ghirlanda-Guerra identities and ultrametricity

The Parisi ansatz

In addition to his formula, Parisi also predicted the salient asymptotic features of the structure of the measures $\gamma_{\beta,N}$ as $N \longrightarrow \infty$. This structure is now known to obtain for certain perturbations of the SK model, as mentioned above.

To introduce these, let $H^{\mathrm{pert}} : \{-1,1\}^N \longrightarrow \mathbb{R}$ be a random function, independent from H, of the form

$$H^{\mathrm{pert}}(\sigma) = \sum_{p\geq 1} 2^{-p} x_p \frac{1}{N^{p/2}} \sum_{i_1,\dots,i_p} g_{i_1,\dots,i_p} \sigma_{i_1}\cdots\sigma_{i_p} \tag{17}$$

for some choice of $(x_p)_{p\geq 1} \in [0,1]^{\mathbb{N}}$, and where all the coefficients $g_{i_1,\dots,i_p}$ are independent standard Gaussians. Using this, form the combined Hamiltonian

$$H^{\mathrm{comb}}(\sigma) = H(\sigma) + s_N H^{\mathrm{pert}}(\sigma)$$

for some sequence of coefficients s_N.

Of course, the rather off-putting formula in (17) need not be seen as a natural model in its own right[4], but as a convenient choice of a very general function which provides many extra parameters that we can tweak as needed. Note, for instance, that the spin-flip symmetry is broken if there are nonzero terms for any odd p.

Now, the point is that if the coefficients s_N are small enough then one can show that this perturbation has only a higher-order effect on the free energy: to be precise,

$$\frac{s_N}{\sqrt{N}} \longrightarrow 0 \quad \Longrightarrow \quad |F_N^{\mathrm{pert}}(\beta) - F_N(\beta)| \longrightarrow 0 \quad \forall \beta, x_1, x_2, \ldots.$$

Therefore, if we can evaluate the asymptotic behaviour of $F_N^{\mathrm{pert}}(\beta)$ as $N \longrightarrow \infty$ for such s_N, this still answers the first main question about the SK model. On the other hand, it turns out that for a generic choice of the coefficients x_p, all the unwanted symmetry is broken, and the resulting Gibbs measures have a very special structure.

This is described by the 'Ruelle Probability Cascades', which we will not introduce carefully here, but their important qualitative features are given in the following theorem:

Theorem 15.1 (The Parisi ansatz) *For almost every $(x_1, x_2, \ldots) \in [0,1]^{\mathbb{N}}$, every subsequential limit γ of the random measures $\gamma_{\beta,N}$ has the following properties:*

- *γ is supported on the sphere $\{\xi \in \ell_2 : \|\xi\| = q^*(\beta)\}$ for some non-random $q^*(\beta) \in [0,1]$,*
- *(Talagrand's positivity principle) if ξ_1, ξ_2 are drawn independently from γ, then $\xi_1 \cdot \xi_2 \geq 0$ a.s.*
- *the support of γ is an ultrametric subset Y of the radius-$q_*(\beta)$ sphere.*

The deepest and most surprising part of this result is that the random measure γ is supported on an ultrametric subset of B, and it turns out that once this is known, the rest of the structure can be deduced fairly directly. This was known as the 'Parisi ultrametricity conjecture', and was the last piece of the above theorem to fall into place in the recent work [Pan13].

[4] Although it has been studied as such; it is called the **mixed p-spin model**.

Recall that a metric space (Y, d_Y) is **ultrametric** if the triangle inequality may be strengthened to

$$d_Y(y_1, y_3) \leq \max\{d_Y(y_1, y_2), d_Y(y_2, y_3)\} \quad \forall y_1, y_2, y_3 \in Y.$$

If Y is contained in a sphere of constant radius q^* in a Hilbert space $\mathfrak{H}$, as in the case above, then this ultrametric inequality implies a very explicit 'heirarchical' structure. If we assume also that the distances between points of Y assume only finitely many different values, it may be described as follows. There are

- a rooted tree T of constant depth d, say,
- a sequence of values
$$0 = q_0 < q_1 < q_2 < \ldots < q_d = q^*,$$
- and pairwise-orthogonal vectors $\xi_{uv} \in \mathfrak{H}$ for every edge uv of T such that $\|\xi_{uv}\| = \sqrt{q_i^2 - q_{i-1}^2}$ if uv connects levels $i-1$ and i of T,

such that Y is the image of the set of leaves ∂T under following map $\varphi : \partial T \longrightarrow \mathfrak{H}$:

$$\varphi(v) = \xi_{v_0 v_1} + \xi_{v_1 v_2} + \cdots + \xi_{v_{d-1} v},$$

where $v_0 v_1 \cdots v_{d-1} v$ is the path from the root to v in T. Now Y is determined up to isometry by T and the lengths q_i. If Y has infinitely many possible inter-point distances, then one needs a slightly more complicated version of this picture. See [Pan12] for a more careful discussion of ultrametricity.

It is easy to see that the full Parisi ansatz cannot hold for the Hamiltonian (10) by itself. Whatever the values of $g_{i,j}$, that Hamiltonian is always invariant under the 'spin-flip' symmetry $(\sigma_i)_i \mapsto (-\sigma_i)_i$, from which it follows easily that any non-trivial limit random measure would violate Talagrand's positivity principle. This spin-flip symmetry is actually obscuring some other structure of importance, and so one must at least perturb the model so far as to break this symmetry, and then try to understand the resulting perturbed Gibbs measures. This situation would be very similar to how the symmetric Gibbs measures for the low-temperature Ising model on $\mathbb{Z}^2$ should be understood as a convex combination of two asymmetric Gibbs measures

So some perturbation to the SK Hamiltonian is needed for the Parisi ansatz, but it is still open whether one really needs the whole infinite-dimensional family introduced above.

From concentration results to the Ghirlanda-Guerra identities

The point of embarkation for obtaining the Ghirlanda-Guerra identities for the SK model is a very basic principle concerning Gibbs measures. It can also be illustrated on the REM. Suppose now that $H : \{-1,1\}^N \longrightarrow \mathbb{R}$ is the random Hamiltonian in either of these models, and form the resulting family of Gibbs measures

$$\gamma_\beta\{\sigma\} = \frac{\exp(-\beta H(\sigma))}{Z(\beta)}.$$

Let $\Phi(\beta) = \log Z(\beta)$, so $F_N(\beta) = \mathsf{E}\frac{1}{N}\Phi(\beta)$.

Now, on the one hand, applying Hölder's inequality to $Z(\beta)$ with $\beta_1, \beta_2 \geq 0$ and $0 \leq t \leq 1$ gives

$$\begin{aligned} Z(t\beta_1 + (1-t)\beta_2) &= \sum_\sigma \mathrm{e}^{-t\beta_1 H(\sigma)}\mathrm{e}^{-(1-t)\beta_2 H(\sigma)} \\ &\leq \Big(\sum_\sigma \mathrm{e}^{-\beta_1 H(\sigma)}\Big)^t \Big(\sum_\sigma \mathrm{e}^{-\beta_2 H(\sigma)}\Big)^{1-t}, \end{aligned}$$

hence convexity:

$$\Phi(t\beta_1 + (1-t)\beta_2) \leq t\Phi(\beta_1) + (1-t)\Phi(\beta_2).$$

On the other, basic calculus gives

$$\Phi'(\beta) = \frac{Z'(\beta)}{Z(\beta)} = \int_{\{-1,1\}^N} (-H(\sigma))\,\gamma_\beta(\mathrm{d}\sigma) \qquad \Longrightarrow \qquad |\Phi'(\beta)| \leq \|H\|_\infty.$$

Another differentiation gives

$$\begin{aligned} \Phi''(\beta) &= \frac{Z''(\beta)}{Z(\beta)} - \frac{Z'(\beta)^2}{Z(\beta)^2} \\ &= \int_{\{-1,1\}^N} H(\sigma)^2\,\gamma_\beta(\mathrm{d}\sigma) - \Big(\int_{\{-1,1\}^N} H(\sigma)\,\gamma_\beta(\mathrm{d}\sigma)\Big)^2 \\ &= \mathrm{Var}_{\gamma_\beta}(H). \end{aligned}$$

With only this in hand, one concludes that for any interval $[a,b] \subseteq [0,\infty)$,

$$\int_a^b \mathrm{Var}_{\gamma_\beta}(H)\,\mathrm{d}\beta = \int_a^b \Phi''(\beta)\,\mathrm{d}\beta = \Phi'(b) - \Phi'(a) \leq 2\|H\|_\infty.$$

This inequality has remarkable consequences in case H already takes large values: if $\|H\|_\infty$ is large, it tells us that $\mathrm{Var}_{\gamma_\beta}(H)$ is not much larger than $\|H\|_\infty$ for most values of $\beta \in [a,b]$. Therefore, one expects the fluctuations of H to be typically $\mathrm{O}(\sqrt{\|H\|_\infty})$ (where 'typically' refers to γ_β). On the other hand, if H is not too irregular then one often finds that $|H|$ itself typically takes values comparable to $\|H\|_\infty$, so that *its fluctuations are much smaller than its typical values*. This applies in the case of the SK model and REM, because there we expect $H(\sigma)$ to have values of order N for most σ, and one can show that its maximum is typically not too much larger than this (see [Tal03, Proposition 1.1.3]).

Now recall that in either of the models of interest, H is a centred Gaussian random field on $\{-1,1\}^N$, and that an appeal to the concentration inequality of Proposition 13.1 gives (12). This tells us that the random function $\frac{1}{N}\Phi_N(\beta)$ is very close to the deterministic function $F_N(\beta)$ as $N \longrightarrow \infty$. Since these are also convex functions, one can turn this into an approximation between their derivatives. Working out the details of these estimates in these particular models, the upshot of this is the estimate

$$\int_a^b \mathsf{E}\Big\langle\Big|\frac{H(\sigma)}{N} - \mathsf{E}\Big\langle\frac{H(\sigma)}{N}\Big\rangle_{\beta,N}\Big|\Big\rangle_{\beta,N}\,\mathrm{d}\beta = \mathrm{O}\Big(\frac{1}{N^{1/4}}\Big).$$

This is explained more carefully as Theorem 2.12.1 in [Tal03].

Now let $\nu = \nu_{\beta,N}$ be the (deterministic) measure $\mathsf{E}\langle - \rangle_{\beta,N}$. The above implies that for any fixed interval $[a,b]$, for most $\beta \in [a,b]$ the quantity $H/N : \{-1,1\}^N \longrightarrow \mathbb{R}$ must be very highly concentrated under the measure $\nu_{\beta,N}$. Using this, with a little care one can extract a β in any chosen interval and a subsequence of these measures such that for any functions $f_N : (\{-1,1\}^N)^m \longrightarrow [-1,1]$ one must have

$$\nu\Big(f_N(\sigma^1,\dots,\sigma^m)\frac{H(\sigma^1)}{N}\Big) - \nu\big(f_N(\sigma^1,\dots,\sigma^m)\big)\nu\Big(\frac{H(\sigma^1)}{N}\Big) \longrightarrow 0, \qquad (18)$$

uniformly in the choice of f_N.

Applying Guassian integration by parts to this apparently simple phenomenon has far-reaching consequences. On the one hand, for any function f we find that

$$\begin{aligned}&\nu\Big(f(\sigma^1,\dots,\sigma^m)\frac{H(\sigma^1)}{N}\Big)\\&= -\beta\Big(\sum_{\ell=1}^n \nu\big(f(\sigma^1,\dots,\sigma^m)(\sigma^1\cdot\sigma^\ell)\big) - m\nu\big(f(\sigma^1,\dots,\sigma^m)(\sigma^1\cdot\sigma^{m+1})\big)\Big),\end{aligned}$$

where

$$\sigma^i \cdot \sigma^j := \frac{1}{N}\mathrm{Cov}(H(\sigma^i), H(\sigma^j)).$$

This is obtained by applying Proposition 13.2 for each tuple $(\sigma^1, \dots, \sigma^m)$ separately to the function

$$F((H(\sigma))_\sigma) = f(\sigma^1, \dots, \sigma^m) \cdot \gamma_{\beta,N}^{\otimes m}(\sigma^1, \dots, \sigma^m)$$
$$= f(\sigma^1, \dots, \sigma^m) \cdot \frac{\mathrm{e}^{-\beta H(\sigma^1)-\beta H(\sigma^2)-\cdots-\beta H(\sigma^m)}}{Z(\beta)^m}.$$

Since $(H(\sigma))_\sigma$ is a centred Gaussian process, we can perform the integration by parts in the Gaussian r.v. $H(\sigma^1)$ with the orthogonal Gaussian process held fixed.

Similarly one can compute that

$$\nu\Big(\frac{H(\sigma^1)}{N}\Big) = -\beta\big(\nu(\sigma^1 \cdot \sigma^1) - \nu(\sigma^1 \cdot \sigma^2)\big)$$

(where in both the SK model and the REM the quantity $\sigma^1 \cdot \sigma^1$ is actually constant, i.e. the same for every σ^1).

Substituting these into (18) and taking a subsequential limit gives

$$\nu\big(f(\sigma^1, \dots, \sigma^m)(\sigma^1 \cdot \sigma^{m+1})\big)$$
$$= \frac{1}{m}\Big(\nu(\sigma^1 \cdot \sigma^2)\nu(f) + \sum_{\ell=2}^{m} \nu(f(\sigma^1, \dots, \sigma^m)(\sigma^1 \cdot \sigma^\ell))\Big),$$

where now $\nu := \mathsf{E}\langle - \rangle$ refers to the subsequential sampling-limit random measure γ. These are the **Ghirlanda-Guerra identities**.

In fact, these are only the first in a large family of identities. If γ is a random Hilbert space measure and $\nu = \mathsf{E}\langle - \rangle$ as before, then γ satisfies the **extended Ghirlanda-Guerra identities** if

$$\nu\big(f(\sigma^1, \dots, \sigma^m)(\sigma^1 \cdot \sigma^{m+1})^p\big)$$
$$= \frac{1}{m}\Big(\nu((\sigma^1 \cdot \sigma^2)^p)\nu(f) + \sum_{\ell=2}^{m} \nu(f(\sigma^1, \dots, \sigma^m)(\sigma^1 \cdot \sigma^\ell)^p)\Big).$$

for all bounded continuous functions f and all $p \geq 1$. Equivalently, this asserts that if one first chooses γ itself at random, and then chooses $\sigma^1, \sigma^2, \dots$ independently at random from γ, then conditionally on $\sigma^1, \dots, \sigma^m$ the inner product $\sigma^1 \cdot \sigma^{m+1}$ has distribution

$$\frac{1}{m}\Big(\mathrm{law}(\sigma^1 \cdot \sigma^2) + \sum_{\ell=2}^{m} \delta_{(\sigma^1 \cdot \sigma^\ell)}\Big).$$

When these extended identities are satisfied, one can show that they give all the desired control over the structure of γ. One needs the large family of perturbations to the SK Hamiltonian that were introduced previously in order to find parameter values at which all of these identities hold simultaneously; we will not explain this further here, but see [Panar, Chapter 3].

16 Obtaining consequences from Ghirlanda-Guerra

Ultrametricity and the Ruelle Probability Cascades

The heart of Panchenko's breakthrough [Pan13] is a proof that the extended Ghirlanda-Guerra identities imply the ultrametricity part of Theorem 15.1. That proof is difficult and a little long, so we will not broach it here, except to report a simple geometric feature of independent interest. To prove ultrametricity, Panchenko actually shows that the Ghirlanda-Guerra identities imply the following property for the support of the limiting Gibbs measure:

Proposition 16.1 (See proof of Theorem 2.13 in [Panar]) *Suppose that $Y \subseteq \mathfrak{H}$ is a closed subset of a Hilbert space with the following property:*

> *If $\xi_1, \xi_2, \dots, \xi_m \in Y$ are points such that*
>
> $$\|\xi_m - \xi_{m-1}\| = \min\{\|\xi_m - \xi_i\| : \ i \le m-1\},$$
>
> *then there are 'duplicates' $\xi'_1, \xi'_2, \dots, \xi'_m \in Y$ and also $\xi''_m \in Y$ such that*
>
> $$\|\xi'_j - \xi'_i\| = \|\xi_j - \xi_i\| \quad \forall i, j \le m,$$
> $$\|\xi''_m - \xi'_i\| = \|\xi_m - \xi_i\| \quad \forall i \le m-1$$
>
> *and*
>
> $$\|\xi'_m - \xi''_m\| = \|\xi_m - \xi_{m-1}\|.$$

Then Y is ultrametric.

The proof of this rests on a careful application of the Cauchy-Schwartz inequality to the average of a large sequence of such duplicates. If one starts with a non-ultrametric triangle, one can produce a distance that must be negative, and hence a contradiction. Note that the above condition is certainly not *necessary* for ultrametric subsets of a Hilbert space: for example, it cannot be satisfied by any finite ultrametric subset.

Once ultrametricity is known, it remains to describe Y exactly in terms of a tree T and distances q_i (or some version of these data for general ultrametrics), as discussed at the beginning of the previous section; and then to describe the distribution of the random measure γ_β supported on Y. The structure of Ruelle Probability Cascades finally appears in the latter step, and is also deduced from the Ghirlanda-Guerra identities once ultrametricity is known. We will not explain these carefully here (again, [Pan12, Panar] give good introductions), but to give some of the flavour we will discuss the analogous problem in the much simpler, toy situation of the REM.

Solving the REM

Assume we know that the limiting random probability measure γ of the REM satisfies the extended Ghirlanda-Guerra identities. For the REM the quantities $\sigma \cdot \sigma'$ between different states can take only two values, since for this model

$$\sigma \cdot \sigma' := \frac{1}{N}\mathrm{Cov}(H(\sigma), H(\sigma')) = \delta_{\sigma,\sigma'}$$

(so $\sigma \cdot \sigma'$ is not now the inner product coming from regarding $\{-1,1\}^N$ as a subset of ℓ_2^N). This property clearly persists for the limiting measure γ, so it follows that the random measure γ is a.s. supported on a sequence of orthogonal elements of its auxiliary Hilbert space $\mathfrak{H}$. However, this means that in this case the extended Ghirlanda-Guerra identities reduce to the following principle:

> If we choose γ at random and then choose elements ξ_1, ξ_2, ... i.i.d. from γ, , then having chosen $\xi_1, \dots, \xi_m$, the probability that $\xi_{m+1} = \xi_1$ (i.e., that $\xi_1 \cdot \xi_{m+1} = 1$, not 0) is
>
> $$\frac{1}{m}((\text{number of } \ell \in \{2, \dots, m\} \text{ s.t. } \sigma^\ell = \sigma^1) + p)$$

where p is the overall probability that two vectors ξ and ξ' drawn in this process will be equal.

Considering only the process that determines whether ξ_{m+1} agrees with one of $\xi_1, \ldots, \xi_m$ or is distinct from all of them, this reveals a random partition of $\mathbb{N}$ as m increases, and now we recognize it: provided $0 < p < 1$, it is the Chinese Restaurant Process with parameter $\alpha = 1 - p$ (recall formula (7)).

Therefore, in the case of the REM, provided it turns out that $0 < \alpha < 1$, the random weights of the limiting random measure γ follow the random mass partition $\mathrm{PD}(\alpha, 0)$ with this α. A separate analysis can now be given to show that $\alpha \in (0,1)$ when $\beta > 2\sqrt{\log 2}$, and then

$$\alpha = 2\sqrt{\log 2}/\beta$$

(see Chapter 1 of [Tal03]). On the other hand, when $\beta \leq 2\sqrt{\log 2}$ (corresponding to high temperature in the physical interpretation), it works out that $p = 1$, $\alpha = 0$, and the limiting probability measure γ simply collapses to a Dirac mass at 0.

References

[AC] L.-P. Arguin and S. Chatterjee. Random overlap structures: properties and applications to spin glasses. Preprint, available online at `arXiv.org`: 1011.1823.

[Ald] David J. Aldous. More uses of exchangeability: representations of complex random structures. to appear in *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*.

[Ald81] David J. Aldous. Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.*, 11(4):581–598, 1981.

[Ald82] David J. Aldous. On exchangeability and conditional independence. In *Exchangeability in probability and statistics (Rome, 1981)*, pages 165–170. North-Holland, Amsterdam, 1982.

[Ald85] David J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985.

[ASS07] Michael Aizenman, Robert Sims, and Shannon L. Starr. Mean-field spin glass models from the cavity-ROSt perspective. In *Prospects in mathematical physics*, volume 437 of *Contemp. Math.*, pages 1–30. Amer. Math. Soc., Providence, RI, 2007.

[Aus08] Tim Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probability Surveys*, (5):80–145, 2008.

[Ber06] Jean Bertoin. *Random fragmentation and coagulation processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

[CH06] Philippe Carmona and Yueyun Hu. Universality in Sherrington-Kirkpatrick's spin glass model. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(2):215–222, 2006.

[Con90] John B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.

[Der81] Bernard Derrida. Random-energy model: an exactly solvable model of disordered systems. *Phys. Rev. B (3)*, 24(5):2613–2626, 1981.

[DF80] P. Diaconis and D. Freedman. Finite exchangeable sequences. *Ann. Probab.*, 8(4):745–764, 1980.

[DJ07] Persi Diaconis and Svante Janson. Graph limits and exchangeable random graphs. Preprint; available online at `arXiv.org`: math.PR math.CO/0712.2749, 2007.

[DS82] L. N. Dovbysh and V. N. Sudakov. Gram-de Finetti matrices. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 119:77–86, 238, 244–245, 1982. Problems of the theory of probability distribution, VII.

[FT85] D. H. Fremlin and M. Talagrand. Subgraphs of random graphs. *Trans. Amer. Math. Soc.*, 291(2):551–582, 1985.

[Gro99] Mikhael Gromov. *Metric Structures for Riemannian and Non-Riemannian Spaces*. Birkhäuser, Boston, 1999.

[GT02] Francesco Guerra and Fabio Lucio Toninelli. The thermodynamic limit in mean field spin glass models. *Comm. Math. Phys.*, 230(1):71–79, 2002.

[Gue03] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.*, 233(1):1–12, 2003.

[Hes86] K. Hestir. The Aldous representation theorem and weakly exchangeable non-negative definite arrays. Ph.D. dissertation, Statistic Dept., Univ. of California, Berkeley, 1986.

[Hoo79] David N. Hoover. Relations on probability spaces and arrays of random variables. 1979.

[Hoo82] David N. Hoover. Row-columns exchangeability and a generalized model for exchangeability. In *Exchangeability in probability and statistics (Rome, 1981)*, pages 281–291, Amsterdam, 1982. North-Holland.

[Kal89] Olav Kallenberg. On the representation theorem for exchangeable arrays. *J. Multivariate Anal.*, 30(1):137–154, 1989.

[Kal92] Olav Kallenberg. Symmetries on random arrays and set-indexed processes. *J. Theoret. Probab.*, 5(4):727–765, 1992.

[Kal02] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[Kal05] Olav Kallenberg. *Probabilistic symmetries and invariance principles*. Probability and its Applications (New York). Springer, New York, 2005.

[Kin78a] J. F. C. Kingman. The representation of partition structures. *J. London Math. Soc. (2)*, 18(2):374–380, 1978.

[Kin78b] J. F. C. Kingman. Uses of exchangeability. *Ann. Probability*, 6(2):183–197, 1978.

[Led01] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[LS06] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[MPV87] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific Publishing Co. Inc., Teaneck, NJ, 1987.

[Pan10] Dmitry Panchenko. On the Dovbysh-Sudakov representation result. *Elec. Commun. in Probab.*, 15:330–338, 2010.

[Pan12] Dmitry Panchenko. The Sherrington-Kirkpatrick model: an overview. *J. Stat. Phys.*, 149(2):632–383, 2012.

[Pan13] Dmitry Panchenko. The Parisi ultrametricity conjecture. *Ann. of Math. (2)*, 177(1):383–393, 2013.

[Panar] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer Monographs in Mathematics. To appear.

[Pit95] Jim Pitman. Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields*, 102(2):145–158, 1995.

[PY97] Jim Pitman and Marc Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25(2):855–900, 1997.

[Tal03] Michel Talagrand. *Spin glasses: a challenge for mathematicians*, volume 46 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, 2003. Cavity and mean field models.

[Tal06] Michel Talagrand. The Parisi formula. *Ann. of Math. (2)*, 163(1):221–263, 2006.

[Tal11] Michel Talagrand. *Mean field models for spin glasses. Volume I*, volume 54 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, 2011. Basic examples.

Courant Institute of Mathematical Sciences
New York University
New York, NY 10012, USA
tim@cims.nyu.edu
http://www.cims.nyu.edu/˜tim