Exchangeable random measures[1]

Tim Austin

Courant Institute, New York University, New York, NY 10012, USA. E-mail: tim@cims.nyu.edu

Received 26 April 2013; revised 6 September 2013; accepted 19 September 2013

Abstract. Let A be a standard Borel space, and consider the space $A^{\mathbb{N}^{(k)}}$ of A-valued arrays indexed by all size-k subsets of $\mathbb{N}$. This paper concerns random measures on such a space whose laws are invariant under the natural action of permutations of $\mathbb{N}$. The main result is a representation theorem for such "exchangeable" random measures, obtained using the classical representation theorems for exchangeable arrays due to de Finetti, Hoover, Aldous and Kallenberg.

After proving this representation, two applications of exchangeable random measures are given. The first is a short new proof of the Dovbysh–Sudakov Representation Theorem for exchangeable positive semi-definite matrices. The second is in the formulation of a natural class of limit objects for dilute mean-field spin glass models, retaining more information than just the limiting Gram–de Finetti matrix used in the study of the Sherrington–Kirkpatrick model.

Résumé. Soit A un espace de Borel standard, et soit $A^{\mathbb{N}^{(k)}}$ l'ensemble des tableaux à valeurs dans A indexés par les sous-ensembles de $\mathbb{N}$ de taille k. On s'intéresse aux mesures aléatoires sur un tel espace dont la loi est invariante par l'action naturelle des permutations de $\mathbb{N}$. Le résultat principal est une représentation de ces mesures aléatoires « échangeables », obtenue à partir des théorèmes de représentation classiques de de Finetti, Hoover, Aldous et Kallenberg pour des tableaux échangeables.

Après avoir prouvé cette représentation, on en donne deux applications. La première est une nouvelle courte preuve du théorème de représentation de Dovbysh–Sudakov pour des matrices définies semi-positives échangeables. La seconde concerne la formulation d'une classe naturelle d'objets limites pour des modèles de champ moyen dilués pour des verres de spins qui capture plus d'information que la seule matrice limite de Gram–de Finetti qui est notamment utilisée dans l'étude du modèle de Sherrington–Kirkpatrick.

1. Introduction

The theory of exchangeable arrays of random variables emerged in work of Hoover [13,14], Aldous [1–3] and Kallenberg [15,16], and amounts to a significant generalization of the classical de Finetti–Hewitt–Savage Theorem on exchangeable sequences. The heart of the theory is a collection of representation theorems for general such arrays, which then beget more specialized representation results such as the Dovbysh–Sudakov Theorem for exchangeable PSD matrices.

This note will consider the related setting of random measures on spaces of arrays, where now the laws of those random measures are assumed invariant under the relevant group action. Intuitively, this introduces an "extra layer of randomness." In order to introduce these formally, let $[n] := \{1, 2, \ldots, n\}$ for $n \in \mathbb{N}$, let $S_{\mathbb{N}} = \bigcup_{n \geq 1} S_{[n]}$ be the group

---

of all permutations of $\mathbb{N}$ which fix all but finitely many elements, and consider a measurable action $T\colon S_{\mathbb{N}} \curvearrowright E$ on a standard Borel space E. In full, this is a measurable function

$$T : S_{\mathbb{N}} \times E \longrightarrow E : (\sigma, x) \mapsto T^{\sigma} x$$

such that

$$T^{\mathrm{id}_{\mathbb{N}}} = \mathrm{id}_E \quad \text{and} \quad T^{\sigma_1} T^{\sigma_2} x = T^{\sigma_2 \sigma_1} x \quad \forall \sigma_1, \sigma_2, x.$$

As is standard, if $\mu \in \Pr E$ and $\sigma \in S_{\mathbb{N}}$ then $T_*^{\sigma} \mu$ denotes the image measure of μ under T^{σ}.

Definition 1.1. *If E is a standard Borel space and $T\colon S_{\mathbb{N}} \curvearrowright E$ is a measurable action, then an* exchangeable random measure *(ERM) on (E, T) is a random variable $\boldsymbol{\mu}$ taking values in $\Pr E$ such that*

$$\boldsymbol{\mu} \overset{\mathrm{law}}{=} T_*^{\sigma} \boldsymbol{\mu} \quad \forall \sigma \in S_{\mathbb{N}};$$

that is,

$$\big(\boldsymbol{\mu}(A)\big)_{A \subseteq_{\mathrm{Borel}} E} \overset{\mathrm{law}}{=} \big(\boldsymbol{\mu}\{x : \ T^{\sigma} x \in A\}\big)_{A \subseteq_{\mathrm{Borel}} E} \quad \forall \sigma \in S_{\mathbb{N}}.$$

These are essentially what ergodic theorists call "quasi-factors" ([11], Chapter 8). We will study these for the group actions that underlie the theory of exchangeable arrays. Given a standard Borel space A and $k \in \mathbb{N}$, the space of k-*dimensional arrays valued in A* is $A^{\mathbb{N}^{(k)}}$, where $\mathbb{N}^{(k)}$ denotes the set of size-k subsets of $\mathbb{N}$. An element of such a space of arrays will often be denoted by $(x_e)_{|e|=k}$ or similarly. (In the following, one could focus instead on arrays indexed by ordered k-tuples, but we have chosen the symmetric case as it is a little simpler and arises more often in applications.) The group $S_{\mathbb{N}}$ acts on $A^{\mathbb{N}^{(k)}}$ by permuting coordinates in the obvious manner:

$$T^{\sigma}\big((x_e)_{|e|=k}\big) = (x_{\sigma(e)})_{|e|=k},$$

where $\sigma(e) = \{\sigma(i) : \ i \in e\}$. Slightly more generally, our main results will also allow Cartesian products of such actions over finitely many different k. Thus, our arrays will usually be indexed by the family $\mathbb{N}^{(\leq k)}$ of subsets of $\mathbb{N}$ of size at most k for some fixed k.

Examples. (1) *If an exchangeable random measure $\boldsymbol{\mu}$ on (E, T) is deterministic, then its constant value must itself be invariant under the action T. In case $E = A^{\mathbb{N}^{(k)}}$ with the action above, this means $\boldsymbol{\mu}$ is almost surely the law of an exchangeable A-valued, $\mathbb{N}^{(k)}$-indexed array.*

(2) *On the other hand, if μ is a T-invariant measure for any action (E, T), then another way to obtain an exchangeable random measure from it is to let*

$$\boldsymbol{\mu} := \delta_X,$$

where X is a random element of E with law μ, and δ_X is the Dirac mass at X.

(3) *In case $E = A^{\mathbb{N}^{(k)}}$ with the action above, example (2) fits into a more general family as follows. The space of probability measures $\Pr A$ is also standard Borel with the Borel structure generated by evaluation of measures on Borel sets. Suppose $(\boldsymbol{\lambda}_e)_{|e|=k}$ is an exchangeable array of $(\Pr A)$-valued random variables, and now let*

$$\boldsymbol{\mu} = \bigotimes_{|e|=k} \boldsymbol{\lambda}_e.$$

This class of examples will feature again later. Such an example is called an exchangeable random product measure *(ERPM).*

(4) *It is also easy to exhibit an ERM which is not ERPM. For example, let* $\boldsymbol{\Pi} = (A, B)$ *be a uniform random bipartition of* $\mathbb{N}$ *(this is obviously exchangeable), and having chosen* $\boldsymbol{\Pi}$ *let* $\boldsymbol{\mu} \in \Pr\{0, 1\}^{\mathbb{N}^{(2)}}$ *be the probability which has two atoms of mass* $\frac{1}{2}$ *on the points*

$$1_{A^{\mathbb{N}^{(2)}}} \quad and \quad 1_{B^{\mathbb{N}^{(2)}}}.$$

(5) *Lastly, given a measurable family of exchangeable random measures* $\boldsymbol{\mu}_t$ *indexed by a parameter* $t \in [0, 1)$, *we may average over this parameter to obtain a mixture of these exchangeable random measures*:

$$\boldsymbol{\mu} = \int_0^1 \boldsymbol{\mu}_t \, \mathrm{d}t.$$

This is clearly still exchangeable.

The main result of this paper characterizes all ERMs on spaces of arrays. To motivate it, we next recall the Representation Theorem for exchangeable arrays themselves. This requires some more notation.

First, for any set S we let $\mathcal{P}S$ denote the power set of S.

Next, suppose that $B_0, B_1, \ldots, B_k$ and A are standard Borel spaces. A Borel function

$$f : B_0 \times B_1^k \times B_2^{[k]^{(2)}} \times \cdots \times B_k = \prod_{i \leq k} B_i^{[k]^{(i)}} \longrightarrow A$$

is *middle-symmetric* if

$$f\big(x, (x_i)_{i \in [k]}, (x_a)_{a \in [k]^{(2)}}, \ldots, x_{[k]}\big) = f\big(x, (x_{\sigma(i)})_{i \in [k]}, (x_{\sigma(a)})_{a \in [k]^{(2)}}, \ldots, x_{[k]}\big)$$

for all $\sigma \in S_{[k]}$.

Given standard Borel spaces $B_0, B_1, \ldots, B_k$ and $A_0, A_1, \ldots, A_k$, and middle-symmetric Borel functions

$$f_i : \prod_{j \leq i} B_j^{[i]^{(j)}} \longrightarrow A_i, \quad i = 0, 1, \ldots, k,$$

we will write $\widehat{f}$ for the function

$$\prod_{i \leq k} B_i^{[k]^{(i)}} \longrightarrow \prod_{i \leq k} A_i^{[k]^{(i)}} : (x_e)_{e \subseteq [k]} \mapsto \big(f_{|e|}\big((x_a)_{a \subseteq e}\big)\big)_{e \subseteq [k]},$$

which combines all of the f_i.

The tuple $(f_0, \ldots, f_k)$ is referred to as a *skew-product tuple*, and the associated function $\widehat{f}$ as a function of *skew-product type*; clearly the latter determines the former uniquely.

Example. *If $k = 2$, then a function of skew-product type* $[0, 1)^{\mathcal{P}[2]} \longrightarrow [0, 1)$ *takes the form*

$$\widehat{f}(x, x_1, x_2, x_{12}) = \big(f_0(x), f_1(x, x_1), f_1(x, x_2), f_2(x, x_1, x_2, x_{12})\big).$$

It is easily checked that if $\widehat{f}$ and $\widehat{g}$ are functions of skew-product type for the same k, then so is $\widehat{g} \circ \widehat{f}$. In terms of $(f_0, \ldots, f_k)$ and $(g_0, \ldots, g_k)$ this composition corresponds to the skew-product tuple

$$h_i\big((x_a)_{a \subseteq [i]}\big) := g_i\big(\big(f_{|a|}\big((x_b)_{b \subseteq a}\big)\big)_{a \subseteq [i]}\big), \quad i = 0, 1, \ldots, k.$$

Slightly abusively, we will also write $\widehat{f}$ for the related function

$$\prod_{i \leq k} B_i^{\mathbb{N}^{(i)}} \longrightarrow \prod_{i \leq k} A_i^{\mathbb{N}^{(i)}} : (x_e)_{|e| \leq k} \mapsto \big(f_{|e|}\big((x_a)_{a \subseteq e}\big)\big)_{|e| \leq k},$$

which also determines $(f_0, \ldots, f_k)$ uniquely.

Theorem 1.2 (Representation Theorem for Exchangeable Arrays; Theorem 7.22 in [18]). *Suppose that* A_0, $A_1, \ldots, A_k$ *are standard Borel spaces and that* $(X_e)_{|e| \leq k}$ *is an exchangeable random array of r.v.s with each* X_e *valued in* $A_{|e|}$. *Then there are middle-symmetric Borel functions*

$$f_i : [0, 1)^{\mathcal{P}[i]} \longrightarrow A_i, \quad i = 0, 1, \ldots, k,$$

such that

$$(X_e)_{|e| \leq k} \overset{\text{law}}{=} \left(f_{|e|}\left((U_a)_{a \subseteq e}\right)\right)_{|e| \leq k} \overset{\text{dfn}}{=} \widehat{f}\left((U_e)_{|e| \leq k}\right),$$

where $(U_e)_{|e| \leq k}$ *is an i.i.d. family of* $U[0, 1)$*-r.v.s.*

The companion Equivalence Theorem, which addresses the non-uniqueness of the representing function $\widehat{f}$, will be recalled later.

To produce a random measure, the idea will simply be to use directing functions f_i that depend on two sources of randomness, and then condition on one of them.

Theorem A. *Suppose that* $\boldsymbol{\mu}$ *is an ERM on* $A_0 \times \cdots \times A_k^{\mathbb{N}^{(k)}}$. *Then there are middle-symmetric Borel functions*

$$f_i : \left([0, 1) \times [0, 1)\right)^{\mathcal{P}[i]} \longrightarrow A_i$$

such that

$$\boldsymbol{\mu}(\cdot) \overset{\text{law}}{=} \mathsf{P}\left(\widehat{f}\left((U_e, V_e)_{|e| \leq k}\right) \in \cdot \,|(U_e)_{|e| \leq k}\right),$$

where U_e *and* V_e *for* $e \subseteq \mathbb{N}, |e| \leq k$ *are all i.i.d.* $\sim U[0, 1)$. *On the right-hand side, this is a measure-valued random variable as a function of the r.v.s* $(U_e)_{|e| \leq k}$.

We will find that after some manipulation of the problem, Theorem A can be deduced from the Representation Theorem and Equivalence Theorems for exchangeable random arrays themselves.

The proof of Theorem A can be considerably simplified when $k = 1$, so we will first prove that case separately. In that case, the structure given by Theorem A is essentially a combination of examples (3) and (5) above. To see this, we reformulate the result as follows.

Given a standard Borel space A, let $B([0, 1), \Pr A)$ denote the space of Lebesgue-a.e. equivalence classes of measurable functions $[0, 1) \longrightarrow \Pr A$. Then $B([0, 1), \Pr A)$ has a natural measurable structure generated by the functionals

$$f \mapsto \int_0^1 \phi(t) f(t, B) \, dt$$

corresponding to all $\phi \in L^\infty[0, 1)$ and Borel subsets $B \subseteq A$. This measurable structure is also standard Borel: for instance, if one realizes A as a Borel subset of a compact metric space, then the above becomes the Borel structure of the topology of convergence in probability on $B([0, 1), \Pr A)$, which is Polish.

Theorem B. *If* $\boldsymbol{\mu}$ *is an ERM on* $A^{\mathbb{N}}$, *then there is an exchangeable sequence of r.v.s* $(\lambda_i)_{i \in \mathbb{N}}$ *taking values in* $B([0, 1), \Pr A)$ *such that*

$$\boldsymbol{\mu}(\cdot) \overset{\text{law}}{=} \int_0^1 \left(\bigotimes_{i \in \mathbb{N}} \lambda_i(t, \cdot) \right) dt.$$

So when $k = 1$, every ERM is a mixture of ERPMs.

With the structure given by Theorem B, one may next apply the de Finetti–Hewitt–Savage Theorem to the sequence λ_i to obtain a random measure $\boldsymbol{\gamma}$ on $B([0, 1), \Pr A)$ such that λ_i is obtained by first choosing $\boldsymbol{\gamma}$ and then choosing λ_i

i.i.d. with law $\boldsymbol{\gamma}$. We write Samp($\boldsymbol{\gamma}$) for the ERM obtained by this procedure, and refer to $\boldsymbol{\gamma}$ as a *directing random measure* for $\boldsymbol{\mu}$.

After proving Theorems A and B, we offer a couple of applications of the case $k = 1$. These applications can also be given higher-dimensional extensions using the cases $k \geq 2$, but those extensions seem less natural. The reader interested only in the applications need not read the proof of the general case of Theorem A.

The first application is a new proof of the classical Dovbysh–Sudakov Theorem:

Dovbysh–Sudakov Theorem. *Suppose* $(R_{ij})_{i,j\in\mathbb{N}}$ *is a random matrix which is a.s. positive semi-definite, and is exchangeable in the sense that*

$$(R_{\sigma(i)\sigma(j)})_{i,j} \overset{\text{law}}{=} (R_{ij})_{i,j} \quad \forall \sigma \in S_{\mathbb{N}}.$$

Then there are a separable real Hilbert space $\mathfrak{H}$ *and an exchangeable sequence* $(\xi_i, a_i)_{i\in\mathbb{N}}$ *of random variables valued in* $\mathfrak{H} \times [0, \infty)$ *such that*

$$(R_{ij})_{i,j} \overset{\text{law}}{=} \big(\langle \xi_i, \xi_j \rangle + \delta_{ij}a_i\big)_{i,j},$$

where δ_{ij} *is the Kronecker delta.*

This first appeared in [9], and more complete accounts were given in [12] and [21]. The proofs of Hestir and Panchenko start with the Aldous–Hoover Representation Theorem, which treats $(R_{ij})_{i,j}$ as a general two-dimensional exchangeable array. They then require several further steps to show that the PSD assumption implies a simplification of that general Aldous–Hoover representation into the form promised above. On the other hand, we will find that if one simply interprets $(R_{ij})_{i,j}$ as the covariance matrix of an exchangeable random measure, then one can read off the Dovbysh–Sudakov Theorem from Theorem B, which in turn does not require the Aldous–Hoover Theorem.

Our second application is to the study of certain mean-field spin glass models, and particularly Viana and Bray's dilute version of the Sherrington–Kirkpatrick model [25]. In the case of the original Sherrington–Kirkpatrick model a great deal has now been proven, much of it relying on the notions of "random overlap structures" and their directing random Hilbert space measures: see, for instance, Panchenko's monograph [22]. The analogous theory for dilute models is less advanced. In this note we will simply sketch how the main conjecture of Replica Symmetry Breaking can be formulated quite neatly in terms of limits of exchangeable random measures, translating from the earlier works [20,23]. We will not recall most of the spin glass theory behind this conjecture, but will refer the reader to those references for more background.

2. The replica trick

The key to Theorem A is the simple observation that the law of a random measure on some space E can be equivalently described by the law of a random sequence in E, obtained by first sampling and quenching that random measure, and then sampling i.i.d. from it. This idea is standard in the more general setting of representing quasi-factors in ergodic theory ([11], Chapter 8). In a sense, it is an abstract version of the "replica trick" from the statistical physics of spin glasses ([19]). In physics, the phrase "replica trick" usually refers to the calculation of the sequence of moments of the (random) partition function of a random Gibbs measure, which is then fed into an ansatz for guessing more about the law of the partition function, such as the expected free energy. This resembles our "replica trick" insofar as computing a moment of the random partition function amounts to computing the partition function for the law of several i.i.d. samples from the random Gibbs measure.

Before we proceed, first observe that, since any standard Borel space is isomorphic to a Borel subset of a compact metric space, we may replace the spaces $A_0, \ldots, A_k$ with such enveloping compact spaces in Theorems A and B, and so assume these spaces are themselves compact. We will make this assumption throughout the proofs of those theorems, although some non-compact examples will re-appear later in the applications.

Proposition 2.1 (Replica trick). *If* $\boldsymbol{\mu}$ *is an ERM on* $\prod_{i\leq k} A_i^{\mathbb{N}^{(i)}}$, *then there are auxiliary standard Borel spaces* $\overline{A}_0$, $\overline{A}_1, \ldots, \overline{A}_k$ *and an exchangeable array* $(Y_e, X_e)_{|e|\leq k}$ *of random variables such that*

- *each (Y_e, X_e) takes values in $\overline{A}_{|e|} \times A_{|e|}$, and*
- *one has*

$$\boldsymbol{\mu}(\cdot) \overset{\text{law}}{=} \mathsf{P}\big((X_e)_{|e| \le k} \in \cdot | (Y_e)_{|e| \le k}\big).$$

Proof. After enlarging the background probability space if necessary, we may couple the random variable $\boldsymbol{\mu}$ with a doubly-indexed family of random variables

$$\big((X_{i,e})_{i \in \mathbb{N}, e \in \mathbb{N}^{(\le k)}}, (X_e)_{e \in \mathbb{N}^{(\le k)}}\big), \tag{1}$$

all taking values in one of the A_i's, as follows:

- first, sample the random measure $\boldsymbol{\mu}$ itself;
- then, choose the sub-families $(X_e)_{|e| \le k}$, $(X_{1,e})_{|e| \le k}$, $(X_{2,e})_{|e| \le k}$, ... independently with law $\boldsymbol{\mu}$.

In notation, this coupling is defined by

$$\mathsf{P}\big((X_e)_{|e| \le k} \in \mathrm{d}\mathbf{a}, (X_{1,e})_{|e| \le k} \in \mathrm{d}\mathbf{a}_1, (X_{2,e})_{|e| \le k} \in \mathrm{d}\mathbf{a}_2, \dots | \boldsymbol{\mu}\big)$$
$$= \boldsymbol{\mu}(\mathrm{d}\mathbf{a}) \cdot \boldsymbol{\mu}(\mathrm{d}\mathbf{a}_1) \cdot \boldsymbol{\mu}(\mathrm{d}\mathbf{a}_2) \cdot \dots \cdot.$$

Having done this, let $\overline{A}_i := A_i^{\mathbb{N}}$ and let $Y_e := (X_{j,e})_{j \in \mathbb{N}} \in \overline{A}_{|e|}$ for each $e \in \mathbb{N}^{(\le k)}$. The exchangeability of $\boldsymbol{\mu}$ implies that the joint distribution of the family (1) is invariant under applying elements of $S_{\mathbb{N}}$ to the indexing sets e, and hence that the process $(Y_e, X_e)_{|e| \le k}$ is exchangeable.

On the other hand, since we assume each A_i is compact, so is $\prod_{i \le k} A_i^{\mathbb{N}^{(i)}}$, and now the Law or Large Numbers shows that in the above process one has the a.s. convergence of empirical measures

$$\frac{1}{N} \sum_{n=1}^{N} \delta_{(X_{n,e})_{|e| \le k}} \longrightarrow \boldsymbol{\mu}$$

in the vague topology on $\Pr \prod_{i \le k} A_i^{\mathbb{N}^{(i)}}$.

Therefore in the process

$$\big((Y_e)_{|e| \le k}, (X_e)_{|e| \le k}, \boldsymbol{\mu}\big)$$

the family of r.v.s $(Y_e)_{|e| \le k}$ determine $\boldsymbol{\mu}$ a.s., whereas conditionally on $\boldsymbol{\mu}$ the family $(Y_e)_{|e| \le k}$ becomes independent from $(X_e)_{|e| \le k}$. This implies that

$$\mathsf{P}\big((X_e)_{|e| \le k} \in \cdot | (Y_e)_{|e| \le k}\big) = \mathsf{P}\big((X_e)_{|e| \le k} \in \cdot | \boldsymbol{\mu}\big) = \boldsymbol{\mu}(\cdot) \quad \text{a.s.},$$

as required. $\qquad\square$

3. Proofs in one dimension

3.1. *Some preliminaries*

We will repeatedly need the following standard tool from measure-theoretic probability. See, for instance, the slightly-stronger Theorem 6.10 in [17].

Lemma 3.1 (Noise-Outsourcing Lemma). *Suppose that A and B are standard Borel spaces and that (X, Y) is an $(A \times B)$-valued r.v. Then, possibly after enlarging the background probability space, there are a r.v. $U \sim \mathrm{U}[0, 1]$ coupled with X and Y and a Borel function $f : A \times [0, 1) \longrightarrow Y$ such that U is independent from X and*

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

Of course, the function f in this lemma is highly non-unique. The degenerate case in which X is deterministic is still important: it reduces to the assertion that for any standard Borel probability space (B, ν) there is a Borel function $f : [0, 1) \longrightarrow B$ such that $f(U) \sim \nu$ when $U \sim \mathrm{U}[0, 1)$.

Finally, let us recall the full de Finetti–Hewitt–Savage Theorem for the case $k = 1$, which is rather stronger than just the case $k = 1$ of Theorem 1.2. The following is the combination of Proposition 1.4 and Corollaries 1.5 and 1.6 in [18].

Theorem 3.2. *Suppose A is a compact metric space and $(X_n)_n$ is an exchangeable sequence of A-valued r.v.s. Then the sequence of empirical distributions*

$$W_N := \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n} \in \mathrm{Pr}\, A$$

converges a.s. to a $(\mathrm{Pr}\, A)$-*valued r.v. W which has the following properties*:

  (i) *W is a.s. a function of $(X_n)_n$,*
 (ii) *W generates the tail σ-algebra of $(X_n)_n$ up to μ-negligible sets;*
(iii) *the r.v.s X_n are conditionally i.i.d. given W;*
(iv) *if Z is any other r.v. on the same probability space such that*

$$(Z, X_1, X_2, \ldots) \stackrel{\mathrm{law}}{=} (Z, X_{\sigma(1)}, X_{\sigma(2)}, \ldots) \quad \forall \sigma \in S_{\mathbb{N}},$$

  then Z is conditionally independent from $(X_n)_n$ over W.

3.2. *Proofs in one dimension*

Proof of Theorem A in one dimension. Suppose μ is an ERM on $A^{\mathbb{N}}$ and let $(Y_n, X_n)_n$ be a process as given by Proposition 2.1.

We next apply Theorem 3.2 twice: first to the sequence $(Y_n)_n$, to obtain a r.v. W taking values in $E := \mathrm{Pr}\, \overline{A}$; and secondly to $(Y_n, X_n)_n$, to obtain a r.v. Z taking values in $F := \mathrm{Pr}(\overline{A} \times A)$. From their definitions as limits of empirical distributions, W is almost surely a function of Z. On the other hand, property (iv) of Theorem 3.2 gives that Z is conditionally independent from $(Y_n)_n$ over W.

Now pick an $n \in \mathbb{N}$. By Lemma 3.1, there is a Borel function $f_1 : E \times F \times \overline{A} \times [0, 1) \longrightarrow A$ such that

$$(W, Z, Y_n, X_n) \stackrel{\mathrm{law}}{=} \big(W, Z, Y_n, f_1(W, Z, Y_n, V_n)\big),$$

where $(V_n)_n$ are i.i.d. $\sim \mathrm{U}[0, 1)$ and are independent from (W, Z, Y_n). Moreover, this same f_1 works for every n, by exchangeability. It follows that in fact

$$\big(W, Z, (Y_n, X_n)_{n \in \mathbb{N}}\big) \stackrel{\mathrm{law}}{=} \big(W, Z, \big(Y_n, f_1(W, Z, Y_n, V_n)\big)_{n \in \mathbb{N}}\big), \tag{2}$$

because both sides have the same marginals for individual n, and both sides are conditionally i.i.d. over (W, Z), so all finite-dimensional marginals agree.

Next, another appeal to Lemma 3.1 gives a Borel function $g : E \times [0, 1) \longrightarrow F$ such that

$$(W, Z) \stackrel{\mathrm{law}}{=} \big(W, g(W, V)\big)$$

with a new independent $V \sim \mathrm{U}[0, 1)$. This implies that

$$\big(W, Z, (Y_n)_{n \in \mathbb{N}}\big) \stackrel{\mathrm{law}}{=} \big(W, g(W, V), (Y_n)_{n \in \mathbb{N}}\big),$$

because Z is conditionally independent from $(Y_n)_n$ over W, so again all finite-dimensional marginals agree. Combining this with (2) gives

$$\big(W, Z, (Y_n, X_n)_{n \in \mathbb{N}}\big) \stackrel{\mathrm{law}}{=} \big(W, g(W, V), \big(Y_n, f_2(W, Y_n, V, V_n)\big)_{n \in \mathbb{N}}\big),$$

where

$$f_2(w, y, v, v') := f_1(w, g(w, v), y, v').$$

It follows that

$$\mathsf{P}((X_n)_n \in \cdot | (Y_n)_n) = \mathsf{P}\big((f_2(W, Y_n, V, V_n))_n \in \cdot | (Y_n)_n\big).$$

Finally, we may apply de Finetti's Theorem again, this time to $(Y_n)_n$, to obtain Borel functions $h_1 : [0, 1) \longrightarrow E$ and $h_2 : [0, 1)^2 \longrightarrow \overline{A}$ such that

$$\big(W, (Y_n)_{n \in \mathbb{N}}\big) \overset{\text{law}}{=} \big(h_1(U), \big(h_2(U, U_n)\big)_{n \in \mathbb{N}}\big),$$

where U and $(U_n)_n$ are i.i.d. $\sim \mathrm{U}[0, 1)$ r.v.s, independent of everything else. Letting

$$f(u, u', v, v') := f_2\big(h_1(u), h_2(u, u'), v, v'\big),$$

substituting for $(W, (Y_n)_n)$ in the above gives

$$\mathsf{P}\big((X_n)_n \in \cdot | (Y_n)_n\big) \overset{\text{law}}{=} \mathsf{P}\big((f(U, U_n, V, V_n))_n \in \cdot | U, (U_n)_n\big),$$

as required. $\qquad\square$

Proof of Theorem B. By the Law of Iterated Conditional Expectation, the representation obtained above may be re-written as

$$\begin{aligned}
\boldsymbol{\mu}(\cdot) &\overset{\text{law}}{=} \mathsf{E}\big(\mathsf{P}\big((f(U, U_n, V, V_n))_n \in \cdot | U, (U_n)_n, V\big) | U, (U_n)_n\big) \\
&= \mathsf{E}\bigg(\bigotimes_{n \in \mathbb{N}} \mathsf{P}\big(f(U, U_n, V, V_n) \in \cdot | U, U_n, V\big) | U, (U_n)_n\bigg) \\
&= \int_0^1 \bigg(\bigotimes_{n \in \mathbb{N}} \lambda_n(t, \cdot)\bigg) \, \mathrm{d}t,
\end{aligned}$$

where

$$\lambda_n(t, \cdot) := \mathsf{P}\big(f(U, U_n, t, V_n) \in \cdot | U, U_n\big).$$

As functions of (U, U_n) for each n, these form an exchangeable sequence of r.v.s taking values in $B([0, 1), \Pr A)$, so the proof is complete. $\qquad\square$

3.3. *Relation to row–column exchangeability*

A relative of exchangeability for a two-dimensional random array $(X_{i,n})_{(i,n) \in \mathbb{N}^2}$ is *row–column exchangeability*, which asserts that

$$(X_{\sigma(i), \tau(n)})_{i,n} \overset{\text{law}}{=} (X_{i,n})_{i,n} \quad \forall \sigma, \tau \in S_{\mathbb{N}}.$$

Since σ and τ may be chosen separately, this is a rather stronger symmetry than ordinary two-dimensional exchangeability. Here, too, there is a representation theorem due to Aldous and Hoover, and also a version in arbitrary dimensions due to Kallenberg, who calls such arrays "separately exchangeable."

Theorem 3.3 (Corollary 7.23 in [18]). *If $(X_{i,n})_{i,n}$ is an A-valued row–column exchangeable array then there is a Borel function $[0, 1)^4 \longrightarrow A$ such that*

$$(X_{i,n})_{i,n} \overset{\text{law}}{=} \left(f(Z, U_i, V_n, W_{i,n}) \right)_{i,n},$$

where Z, U_i for $i \in \mathbb{N}$, V_n for $n \in \mathbb{N}$ and $W_{i,n}$ for $(i, n) \in \mathbb{N}^2$ are i.i.d. $\sim \mathrm{U}[0, 1)$.

An alternative proof of Theorem B can be given via Theorem 3.3. One begins with the construction of the two-dimensional random array $(X_{i,n})_{i,n}$ as in the proof of Proposition 2.1 (where the sets e have become singletons n). Since this array is row–column exchangeable, the Representation Theorem gives

$$(X_{i,n})_{i,n} \overset{\text{law}}{=} \left(f(U, U_i, V_n, W_{i,n}) \right)_{i,n}$$

for some Borel directing function $f : [0, 1)^4 \longrightarrow A$, where U, U_i for $i \in \mathbb{N}$, V_n for $n \in \mathbb{N}$ and $W_{i,n}$ for $i, n \in \mathbb{N}$ are i.i.d. $\sim \mathrm{U}[0, 1)$. One can now read off a directing random measure $\gamma(U)$ on $B([0, 1), \mathrm{Pr}\, A)$, a function of $U \sim \mathrm{U}[0, 1)$, in the following two steps:

- first, for each fixed U and U' one obtains an element $\lambda(U, U') \in B([0, 1), \mathrm{Pr}\, A)$ according to

$$\lambda(U, U')(t, \cdot) = \mathrm{P}_W \left(f(U, U', t, W) \in \cdot \right), \quad W \sim \mathrm{U}[0, 1);$$

- second, $\gamma(U)$ is the distribution of $\lambda(U, U')$ where $U' \sim \mathrm{U}[0, 1)$.

On the other hand, a couple of simple applications of the Noise-Outsourcing Lemma show that any directing random measure γ on $B([0, 1), \mathrm{Pr}\, A)$ can be represented this way, so this gives a bijective correspondence

$$\left\{ \text{directing random measures on } B([0, 1), \mathrm{Pr}\, A) \text{ up to equivalence} \right\}$$

$$\leftrightarrow \left\{ \text{directing functions } [0, 1)^4 \longrightarrow A \text{ up to equivalence} \right\},$$

where "up to equivalence" refers to the possibility that different directing random measures or directing functions may give rise to the same row–column exchangeable array.

This approach is the basis of the paper [20], to be discussed later. It is quick, but at the expense of assuming Theorem 3.3.

On the other hand, our approach to Theorem B does not use any exchangeability theory in dimensions greater than one. Moreover, one can reverse the idea above to give a fairly quick proof of Theorem 3.3 using Theorem B.

Proof of Theorem 3.3 from Theorem B. First let $(X_{i,n})_{i,n} = ((X_{i,n})_i)_n$, thought of as an exchangeable sequence of $A^{\mathbb{N}}$-valued r.v.s. By the de Finetti–Hewitt–Savage Theorem applied to the exchangeability in n, its law is a mixture of product measures; equivalently, there is a $(\mathrm{Pr}\, A^{\mathbb{N}})$-valued r.v. $\boldsymbol{\mu}$ such that

$$\mathrm{law}\left(((X_{i,n})_i)_n \right) = \mathsf{E}\left(\boldsymbol{\mu}^{\otimes \mathbb{N}} \right).$$

On the other hand, for any $\sigma \in S_{\mathbb{N}}$ the exchangeability in i gives

$$\mathsf{E}\left(\boldsymbol{\mu}^{\otimes \mathbb{N}} \right) = \mathrm{law}\left(((X_{i,n})_i)_n \right) = \mathrm{law}\left(((X_{\sigma(i),n})_i)_n \right) = \mathsf{E}\left((T_*^\sigma \boldsymbol{\mu})^{\otimes \mathbb{N}} \right),$$

where $T^\sigma : A^{\mathbb{N}} \longrightarrow A^{\mathbb{N}}$ is the corresponding coordinate-permuting transformation. By the uniqueness of the de Finetti–Hewitt–Savage decomposition, this implies

$$\boldsymbol{\mu} \overset{\text{law}}{=} T_*^\sigma \boldsymbol{\mu} \quad \forall \sigma \in S_{\mathbb{N}},$$

so $\boldsymbol{\mu}$ is an ERM. Therefore Theorem B gives

$$\mathrm{law}\left(((X_{i,n})_i)_n \right) = \mathsf{E}\left(\int_0^1 \bigotimes_{i \in \mathbb{N}} \lambda_i(t, \cdot)\, \mathrm{d}t \right)^{\otimes \mathbb{N}}$$

for some exchangeable random sequence $(\lambda_i)_i$ taking values in $B([0, 1), \Pr A)$.

Next, applying the de Finetti–Hewitt–Savage Theorem to the sequence $(\lambda_i)_i$ itself gives a Borel function $F : [0, 1)^2 \longrightarrow B([0, 1), \Pr A)$ such that the above becomes

$$
\mathrm{law}\big(((X_{i,n})_i)_n\big) = \mathsf{E}_{Z,(U_i)_i}\bigg(\int_0^1 \bigotimes_{i \in \mathbb{N}} F(Z, U_i, t, \cdot)\,\mathrm{d}t\bigg)^{\otimes \mathbb{N}}
$$

$$
= \mathsf{E}_{Z,(U_i)_i,(V_n)_n} \bigotimes_{(i,n) \in \mathbb{N}^2} F(Z, U_i, V_n, \cdot), \tag{3}
$$

where Z, U_i for $i \in \mathbb{N}$ and V_n for $n \in \mathbb{N}$ are i.i.d. $\sim \mathrm{U}[0, 1)$, and in the second equality we have simply changed notation from "$\int_0^1 \cdot \,\mathrm{d}t$" to "$\mathsf{E}_{V_n}$."

Finally, by Lemma 3.1 there is a Borel function $f : [0, 1)^4 \longrightarrow A$ such that

$$
\mathsf{P}\big(f(Z, U, V, W) \in \cdot\,|Z, U, V\big) = F(Z, U, V, \cdot) \quad \text{a.s.}
$$

when Z, U, V, W are i.i.d. $\sim \mathrm{U}[0, 1)$, and now the right-hand side of (3) becomes $\mathrm{law}((f(U, U_i, V_n, W_{i,n})_i)_n)$, as required. $\qquad \square$

4. Proof in higher dimensions

Not all parts of Theorem 3.2 generalize to higher-dimensional arrays, and instead we must make a more careful argument using the Equivalence Theorem 4.1 below.

4.1. *Some more preliminaries*

The Equivalence Theorem characterizes when two functions direct the same process in the setting of Theorem 1.2. Its formulation needs the following notion. Let $C = [0, 1)^r$ and $D = [0, 1)^s$ for some $r, s \in \mathbb{N}$. Then a skew-product tuple $(f_0, \dots, f_k)$ in which each $f_i : \prod_{j \leq i} C^{[i]^{(j)}} \longrightarrow D$ for each i gives rise to a skew-product-type function $\widehat{f} : C^{\mathcal{P}[k]} \longrightarrow D^{\mathcal{P}[k]}$, which is a map between Euclidean cubes. We will write that the skew-product tuple is *Lebesgue-measure-preserving* if for all $i = 0, 1, \dots, k$ and all $(x_a)_{a \subsetneq [i]} \in C^{\mathcal{P}[i]\setminus[i]}$, one has

$$
U \sim \mathrm{U}(C) \quad \Longrightarrow \quad f_i\big((x_a)_{a \subsetneq [i]}, U\big) \sim \mathrm{U}(D).
$$

This implies, in particular, that $\widehat{f}$ pushes Lebesgue measure on $C^{\mathcal{P}[k]}$ to Lebesgue measure on $D^{\mathcal{P}[k]}$ (although these cubes have different dimensions if $r \neq s$). However, the assertion that the skew-product tuple is Lebesgue-measure-preserving can be strictly stronger than this, in case the functions f_i are not all injective.

The Equivalence Theorem is as follows.

Theorem 4.1 (Equivalence Theorem for directing functions; Theorem 7.28 in [18]). *If $\widehat{f}, \widehat{f}' : [0, 1)^{\mathcal{P}[k]} \longrightarrow A_0 \times \cdots \times A_i^{[k]^{(i)}} \times \cdots \times A_k$ are functions of skew-product type such that*

$$
\big(f_{|e|}((U_a)_{a \subseteq e})\big)_{|e| \leq k} \overset{\mathrm{law}}{=} \big(f'_{|e|}((U_a)_{a \subseteq e})\big)_{|e| \leq k},
$$

then there are functions $\widehat{G}, \widehat{G}' : [0, 1)^{\mathcal{P}[k]} \longrightarrow [0, 1)^{\mathcal{P}[k]}$ of skew-product type, whose skew-product tuples are Lebesgue-measure-preserving, and which make the following diagram commute:

$$[0, 1)^{\mathcal{P}[k]}$$

$$A_0 \times A_1^k \times \cdots \times A_k$$
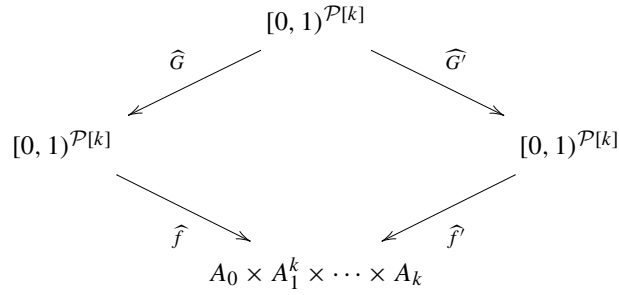
In connection with this theorem, we will also need the following "factorization" result.

Corollary 4.2. *Let* $U_{\subseteq[k]} = (U_e)_{e \subseteq [k]}$ *and* $V_{\subseteq[k]}$ *be independent uniform r.v.s valued in* $[0, 1)^{\mathcal{P}[k]}$. *If*

$$G : [0, 1)^{\mathcal{P}[k]} \longrightarrow [0, 1)^{\mathcal{P}[k]}$$

is a function of skew-product type whose skew-product tuple is Lebesgue-measure-preserving, then there is another function

$$H : \big([0, 1) \times [0, 1)\big)^{\mathcal{P}[k]} \longrightarrow [0, 1)^{\mathcal{P}[k]}$$

of skew-product type, whose skew-product tuple is Lebesgue-measure-preserving, and such that

$$U_{\subseteq[k]} = G\big(H(U_{\subseteq[k]}, V_{\subseteq[k]})\big) \quad a.s.$$

Another way to express this is that the maps in the following diagram come from Lebesgue-measure-preserving skew-product tuples and a.s. commute:
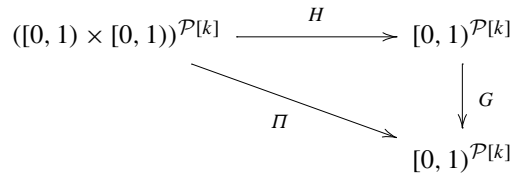
where

$$\Pi\big((x_e, y_e)_{e \subseteq [k]}\big) = (x_e)_{e \subseteq [k]}$$

is the obvious projection.

Geometrically, the intuition here is that G is "almost onto" (since its image measure is Lebesgue), and that as a result one can represent it as the projection map Π after using H to "straighten out the fibres."

Proof of Corollary 4.2. Let G be defined by the skew-product tuple $(G_0, \ldots, G_k)$. We must construct the skew-product tuple $(H_0, \ldots, H_k)$ that defines H. In terms of these tuples, our requirement is that

$$G_i\big(\big(H_{|e|}\big((U_a, V_a)_{a \subseteq e}\big)\big)_{e \subseteq [i]}\big) = U_{[i]} \quad \text{a.s. } \forall i = 0, 1, \ldots, k. \tag{4}$$

When $i = 0$ this simplifies to

$$G_0\big(H_0(U_0, V_0)\big) = U_0 \quad \text{a.s.}$$

We can obtain such an H_0 from the Noise-Outsourcing Lemma 3.1 as follows. Let Z_0 be a U[0, 1)-r.v. and let $X_0 := G_0(Z_0)$, so this is also $\sim$ U[0, 1). Applying Lemma 3.1 to the pair (X_0, Z_0) gives a Borel function $H_0 : [0, 1) \times [0, 1) \longrightarrow [0, 1)$ such that

$$(X_0, Z_0) = \big(X_0, H_0(X_0, Y_0)\big) \quad \text{a.s.}$$

for some $Y_0 \sim U[0, 1)$ independent from X_0. Since $X_0 = G_0(Z_0)$, applying G_0 to the second coordinates here gives

$$X_0 = G_0\big(H_0(X_0, Y_0)\big) \quad \text{a.s.}$$

The general case now follows by induction on i. Suppose that $i \geq 1$, let Y_e for $e \subsetneq [i]$ and Z_e for $e \subseteq [i]$ be i.i.d. $\sim U[0, 1)$; define $X_e := G_{|e|}((Z_a)_{a \subseteq e})$ for all $e \subseteq [i]$; and assume that $H_0, \ldots, H_{i-1}$ have already been constructed such that $Z_e := H_{|e|}((X_a, Y_a)_{a \subseteq e})$ for each $e \subsetneq [i]$. Applying Lemma 3.1 again gives a Borel function $H_i : ([0, 1) \times [0, 1))^{\mathcal{P}[i]} \longrightarrow [0, 1)$ and a r.v. $Y_{[i]} \sim U[0, 1)$ such that

$$\big((X_e)_{e \subseteq [i]}, (Y_e)_{e \subsetneq [i]}, Z_{[i]}\big) = \big((X_e)_{e \subseteq [i]}, (Y_e)_{e \subsetneq [i]}, H_i\big((X_e)_{e \subseteq [i]}, (Y_e)_{e \subseteq [i]}\big)\big),$$

and as before this is equivalent to the desired equality (4). $\qquad \square$

4.2. *Completion of the proof*

We need the following enhancement of Proposition 2.1.

Lemma 4.3. *If μ is an ERM on $\prod_{i \leq k} A_i^{\mathbb{N}^{(i)}}$, then there is an exchangeable array $(U_e, X_e)_{|e| \leq k}$ such that*

- $(U_e)_{|e| \leq k}$ *are i.i.d.* $\sim U[0, 1)$,
- *each X_e takes values in $A_{|e|}$,*
- *and one has*

$$\mu(\cdot) \overset{\text{law}}{=} P\big((X_e)_{|e| \leq k} \in \cdot | (U_e)_{|e| \leq k}\big).$$

Proof. Let $(Y_e, X_e)_{|e| \leq k}$ be the process given by Proposition 2.1, with each (Y_e, X_e) taking values in $\overline{A}_{|e|} \times A_{|e|}$. By the Representation Theorem 1.2 applied to $(Y_e)_{|e| \leq k}$, there is a function $\widehat{f} : [0, 1)^{\mathcal{P}[k]} \longrightarrow \prod_{i \leq k} \overline{A}_i^{[k]^{(i)}}$ of skew-product type such that

$$(Y_e)_{|e| \leq k} \overset{\text{law}}{=} \widehat{f}\big((U_e)_{|e| \leq k}\big),$$

where $(U_e)_{|e| \leq k}$ is an i.i.d. $\sim U[0, 1)$ array.

Now consider the coupling $(U_e, X_e)_{|e| \leq k}$ whose law is the relatively independent product over the condition $(Y_e)_{|e| \leq k} = \widehat{f}((U_e)_{|e| \leq k})$:

$$P\big((U_e)_e \in d\mathbf{u}, (X_e)_e \in d\mathbf{a}\big)$$
$$= P\big((U_e)_e \in d\mathbf{u}\big) \cdot P\big((X_e)_e \in d\mathbf{a} | (Y_e)_e = \widehat{f}(\mathbf{u})\big). \tag{5}$$

This is exchangeable and has the three desired properties. The exchangeability follows because both factors on the right-hand side of (5) are invariant under the action of $S_{\mathbb{N}}$ on the indexing set e, by the exchangeability of $(U_e)_e$ and $(Y_e, X_e)_e$. The first two of the properties listed are obvious, and the third follows from Proposition 2.1 because the above relative product formula gives

$$P\big((X_e)_e \in d\mathbf{a} | (U_e)_e = \mathbf{u}\big) = P\big((X_e)_e \in d\mathbf{a} | (Y_e)_e = \widehat{f}(\mathbf{u})\big),$$

and we conditioned on the equality $(Y_e)_e = \widehat{f}((U_e)_e)$. $\qquad \square$

Proof of Theorem A. Let the process $(U_e, X_e)_{|e| \leq k}$ be as in the preceding corollary. Applying Theorem 1.2 to this whole process gives functions $\widehat{g} : [0, 1)^{\mathcal{P}[k]} \longrightarrow [0, 1)^{\mathcal{P}[k]}$ and $\widehat{h} : [0, 1)^{\mathcal{P}[k]} \longrightarrow \prod_{i \leq k} A_i^{[k]^{(i)}}$ of skew-product type such that

$$\big((U_e)_{|e| \leq k}, (X_e)_{|e| \leq k}\big) \overset{\text{law}}{=} \big(\widehat{g}\big((U'_e)_{|e| \leq k}\big), \widehat{h}\big((U'_e)_{|e| \leq k}\big)\big), \tag{6}$$

where again $(U'_e)_{|e|\leq k}$ are i.i.d. $\sim \mathrm{U}[0,1]$.

For the first coordinates, this reads

$$(U_e)_{|e|\leq k} \overset{\text{law}}{=} \widehat{g}\big((U'_e)_{|e|\leq k}\big).$$

Since both input and output are i.i.d. U[0, 1] arrays, we may apply the Equivalence Theorem 4.1 to this equality of laws: it gives functions $G, G' : [0,1)^{\mathcal{P}[k]} \longrightarrow [0,1)^{\mathcal{P}[k]}$ of skew-product type whose skew-product tuples are Lebesgue-measure-preserving and which make the following diagram commute:
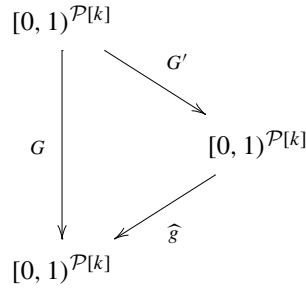
(Note that this seems almost unnecessary, since $\widehat{g}$ already sends Lebesgue measure on $[0,1)^{\mathcal{P}[k]}$ to itself. However, we will need the structure of Lebesgue-measure-preserving skew-product tuples, which need not follow in case $\widehat{g}$ is not injective: see the remarks immediately preceding Theorem 4.1.)

Now applying Corollary 4.2 to G, one obtains a function $H : ([0,1) \times [0,1))^{\mathcal{P}[k]} \longrightarrow [0,1)^{\mathcal{P}[k]}$ of skew-product type, given by a Lebesgue-measure-preserving skew-product tuple, and such that the above commutative diagram can be enlarged to
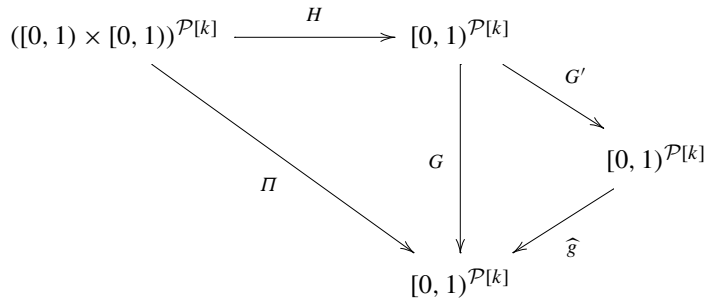
Let $(V_e)_{|e|\leq k}$ be another collection of i.i.d. U[0, 1]-r.v.s independent from $(U_e)_{|e|\leq k}$, and let $\widehat{f} := \widehat{h} \circ G' \circ H$. Then the above diagram implies that:

- on the one hand,

$$(U_e)_{|e|\leq k} = \widehat{g}\big(G'\big(H\big((U_e, V_e)_{|e|\leq k}\big)\big)\big) \quad \text{a.s.,}$$

- and on the other, $G'(H((U_e, V_e)_{|e|\leq k}))$ is an i.i.d. array of U[0, 1]-r.v.s, and so

$$\big(\widehat{g}\big((U'_e)_{|e|\leq k}\big), \widehat{h}\big((U'_e)_{|e|\leq k}\big)\big)$$

$$\overset{\text{law}}{=} \big(\widehat{g}\big(G'\big(H\big((U_e, V_e)_{|e|\leq k}\big)\big)\big), \widehat{h}\big(G'\big(H\big((U_e, V_e)_{|e|\leq k}\big)\big)\big)\big).$$

Combining (6) with these two facts now gives

$$(U_e, X_e)_{|e|\leq k} \overset{\text{law}}{=} \big(\widehat{g}\big(G'\big(H\big((U_e, V_e)_{|e|\leq k}\big)\big)\big), \widehat{h}\big(G'\big(H\big((U_e, V_e)_{|e|\leq k}\big)\big)\big)\big)_{|e|\leq k}$$

$$\overset{\text{law}}{=} \big((U_e)_{|e|\leq k}, \widehat{f}\big((U_e, V_e)_{|e|\leq k}\big)\big),$$

and conditioning both sides of this on $(U_e)_{|e|\leq k}$ gives

$$\mathsf{P}\big((X_e)_{|e|\leq k}\in\cdot|(U_e)_{|e|\leq k}\big)=\mathsf{P}\big(\widehat{f}\big((U_e,V_e)_{|e|\leq k}\big)\in\cdot|(U_e)_{|e|\leq k}\big),$$

as required. $\qquad\square$

5. Relation to Dovbysh–Sudakov Theorem

Proof of Dovbysh–Sudakov Theorem. The trick to this is the standard one-to-one correspondence

$$\big\{\text{PSD }(\mathbb{N}\times\mathbb{N})\text{-matrices}\big\}\leftrightarrow\big\{\text{Gaussian measures on }\mathbb{R}^{\mathbb{N}}\big\}$$

in which a Gaussian measure is identified with its variance–covariance matrix. (This is elementary for finite PSD matrices, and then the infinite case follows by the Daniell–Kolmogorov Theorem: see [17], Theorem 6.14.) Because Gaussian measures are uniquely determined by their variance–covariance matrices, this correspondence intertwines the two permutations actions of $\mathbb{N}$, so from $(R_{ij})_{i,j}$ we may construct an ERM $\boldsymbol{\mu}$ on $\mathbb{R}^{\mathbb{N}}$ which is almost surely Gaussian, and such that

$$R_{ij}=\int_{\mathbb{R}^{\mathbb{N}}}x_ix_j\boldsymbol{\mu}\big(\mathrm{d}(x_n)_{n\in\mathbb{N}}\big)\quad\text{a.s.}$$

Now Theorem B gives a representation

$$\boldsymbol{\mu}\stackrel{\text{law}}{=}\int_0^1\bigotimes_i\boldsymbol{\lambda}_i(t,\cdot)\,\mathrm{d}t$$

with $(\boldsymbol{\lambda}_i)_i$ drawn from some exchangeable sequence taking values in $B([0,1),\mathrm{Pr}\,\mathbb{R})$. Substituting this above gives

$$R_{ii}\stackrel{\text{law}}{=}\int_0^1\int_{\mathbb{R}}x^2\boldsymbol{\lambda}_i(t,\mathrm{d}x)\,\mathrm{d}t$$

and

$$R_{ij}\stackrel{\text{law}}{=}\int_0^1\left(\int_{\mathbb{R}}x\boldsymbol{\lambda}_i(t,\mathrm{d}x)\right)\left(\int_{\mathbb{R}}x\boldsymbol{\lambda}_j(t,\mathrm{d}x)\right)\mathrm{d}t\quad\text{when }i\neq j.$$

Letting

$$\mathfrak{H}=L^2\big([0,1),\mathrm{d}t\big),$$

$$\xi_i(t)=\int_{\mathbb{R}}x\boldsymbol{\lambda}_i(t,\mathrm{d}x)$$

and

$$a_i=\int_0^1\left(\int_{\mathbb{R}}x^2\boldsymbol{\lambda}_i(t,\mathrm{d}x)-\left(\int_{\mathbb{R}}x\boldsymbol{\lambda}_i(t,\mathrm{d}x)\right)^2\right)\mathrm{d}t,$$

this is the desired representation.

(Note that ξ_i must be in $\mathfrak{H}$ a.s. because

$$\int_0^1\xi_i(t)^2\,\mathrm{d}t=\int_0^1\left(\int_{\mathbb{R}}x\boldsymbol{\lambda}_i(t,\mathrm{d}x)\right)^2\mathrm{d}t\leq\int_0^1\int_{\mathbb{R}}x^2\boldsymbol{\lambda}_i(t,\mathrm{d}x)\,\mathrm{d}t\stackrel{\text{law}}{=}R_{ii},$$

which is finite a.s.) $\qquad\square$

6. Limiting behaviour of the Viana–Bray model

Our second, and much more tentative, application for ERMs is to the study of the Viana–Bray (VB) model [25]. This is the basic "dilute" mean-field spin glass model. On the configuration space $\{-1, 1\}^N$, it is given by the random Hamiltonian

$$H_N(\sigma) = \sum_{k=1}^{M} J_k \sigma_{i_k} \sigma_{j_k}, \tag{7}$$

where:

- M is a Poisson r.v. with mean αN (the thermodynamic limit is be taken with α fixed);
- $i_1, j_1, i_2, j_2, \ldots$ are indices from $[N]$ chosen uniformly and independently at random;
- and $J_1, J_2, \ldots$ are i.i.d. symmetric $\mathbb{R}$-valued r.v.s with some given distribution, often taken to be uniform ± 1.

(There are many essentially equivalent variants of this model, but this popular version will do here.) From a quenched choice (that is, a fixed sample) of the random function H_N, the objects of interest are the resulting Gibbs measure

$$\gamma_{\beta,N}\{\sigma\} = \frac{1}{Z_N(\beta)} \exp\big(-\beta H_N(\sigma)\big),$$

the partition function

$$Z_N(\beta) = \sum_{\sigma} \exp\big(-\beta H_N(\sigma)\big)$$

and the expected specific free energy

$$F_N(\beta) = \frac{1}{N} \mathsf{E} \log Z_N(\beta), \tag{8}$$

where the expectation is over the random function H_N. We will sometimes drop the subscript "β" or "N" in the sequel.

This is a relative of the older Sherrington–Kirkpatrick (SK) model [19], in which all pairs of spins ij interact according to independent random coefficients $g_{ij} \sim \mathrm{N}(0, 1/N)$. The rigorous study of the SK model has become quite advanced in recent years; we will not credit all of the important contributions, but refer the reader to the books [22,24] and the many references given there. By contrast, most properties of the VB model remain conjectural.

A key tool in the study of the SK model is the use of random measures on Hilbert space as a kind of "limit object" for the random Gibbs measures $\gamma_{\beta,N}$ as $N \longrightarrow \infty$. Viewing $\frac{1}{\sqrt{N}}\{-1, 1\}^N$ as a subset of ℓ_2^N, $\gamma_{\beta,N}$ is itself a random Hilbert space measure, and the appropriate notion of convergence is convergence in distribution of the Gram–de Finetti matrices obtained by sampling. Having taken a limit in this sense, a limit object in the form of a random measure on Hilbert space is provided by the Dovbysh–Sudakov Theorem. This use of exchangeability and limit objects originates in works of Arguin [4] and Arguin and Aizenman [5], with a precedent in the study of classical mean-field models in the work [10] of Fannes, Spohn and Verbeure. It is explained in more detail in [22].

The key point for this use of random Hilbert-space measures is that the main properties of the SK model, such as the free energy, really depend only on the covariances among the random variables $H(\sigma)$, and hence on this Hilbert space structure. This is no longer true for the VB model, so a more refined tool is needed. One possibility has been explored in [20], and before that physicists and mathematicians had already worked with the related notion of "multi-overlap structures" (see, e.g., [7,8], and also [23], although the latter does not use that terminology).

Here we will simply propose exchangeable random measures as a fairly intuitive equivalent formalism, and compare it with two predecessors from the literature: the weighting schemes used by Panchenko and Talagrand in [23], and Panchenko's use of directing functions in [20]. After introducing our notion of "limit object," we will give a fairly brisk summary of the translations between these formalisms; the calculations are all routine. We will restrict attention to the Viana–Bray model as above for simplicity, but the discussion could easily be extended to a more general class of dilute models, as in [20,23].

6.1. *Basic idea*

If $\gamma_{\beta,N}$ is as above, then it defines an ERM μ by sampling: first quench the random measure $\gamma_{\beta,N}$; then select samples (called "replicas") $\sigma^1, \sigma^2, \ldots \in \{-1, 1\}^N$ i.i.d. $\sim \gamma_N$; and finally use these to define μ as a mixture of delta masses:

$$\mu = \frac{1}{N} \sum_{n=1}^{N} \delta_{(\sigma_n^1, \sigma_n^2, \ldots)}. \tag{9}$$

Identifying ± 1 with the extreme points of $\mathrm{Pr}\{-1, 1\}$, this is clearly a mixture of ERPMs of the kind considered previously. Let $\mathrm{Samp}(\gamma_{\beta,N})$ be the law of μ.

It now makes sense to say that $\gamma_{\beta,N}$ *sampling converges* to some random probability measure γ_β on $B([0, 1),$ $\mathrm{Pr}\{-1, 1\})$ if $\mathrm{Samp}(\gamma_{\beta,N})$ converges to $\mathrm{Samp}(\gamma_\beta)$ for the vague topology on $\mathrm{Pr}(\mathrm{Pr}\{-1, 1\}^{\mathbb{N}})$. This last space is compact, and the laws of exchangeable random measures clearly comprise a further subspace which is closed for the vague topology (since invariance under any given continuous transformation of $\{-1, 1\}^{\mathbb{N}}$ is a closed property). Therefore one can always at least take subsequential limits of $(\mathrm{Samp}(\gamma_{\beta,N}))_N$, and now Theorem B promises the existence of some γ_β that represents the limiting ERM (although it is unique only up to equivalence).

This idea generalizes the more classical use of Gram–de Finetti matrices and their limits recalled above. Starting from the SK model, the associated Gram–de Finetti matrix is obtained by sampling and then quenching the Gibbs measure, and then sampling from that Gibbs measure a sequence of states in $\{-1, 1\}^N$ and computing their inner products as elements of ℓ_2^N, normalized by N (that is, their "overlaps"). Comparing with the random measure μ in (9), this Gram–de Finetti matrix may be recovered as simply the (random) matrix of covariances of the different coordinates in $\{-1, 1\}^{\mathbb{N}}$ under this (random) measure.

6.2. *Comparison with weighting schemes and directing functions*

In [23] the authors do not introduce a notion of limits as such for the random measures $\gamma_{\beta,N}$, but they do formulate their most general results (Section 3 of that paper) in terms of some data that they call a "weighting scheme." This consists of:

- a sequence of $\mathbb{R}$-valued r.v.s $(X_k)_k$, and a family $((X_k^{i,j})_k)_{i,j}$ of i.i.d. copies of this sequence indexed by $(i, j) \in \mathbb{N}^2$;
- and, independently of these, a $[0, 1]$-valued random sequence of weights $(v_k)_k$ such that $\sum_k v_k = 1$.

These data appear in an upper-bound formula for the free energy which will be recalled below. They can be encapsulated in a certain directing random measure γ on $B([0, 1), \mathrm{Pr}\{-1, 1\})$ as follows. First, identifying elements of $\mathrm{Pr}\{-1, 1\}$ with their expectations gives

$$B\big([0, 1), \mathrm{Pr}\{-1, 1\}\big) = B\big([0, 1), [-1, 1]\big).$$

Now let $\Phi(x) := \mathrm{e}^x/(\mathrm{e}^x + \mathrm{e}^{-x})$. Applying Lemma 3.1, we may find a sequence $(f_k)_k$ in $B([0, 1), [-1, 1])$ such that

$$\big(\Phi(X_k)\big)_k \stackrel{\text{law}}{=} \big(f_k(U)\big)_k \quad \text{when } U \sim \mathrm{U}[0, 1]. \tag{10}$$

To finish, let γ be the atomic random measure

$$\gamma = \sum_{k \geq 1} v_k \delta_{f_k}, \tag{11}$$

so the randomness of γ is derived from the random choice of the weights v_k.

Clearly one could find many other ways to convert a weighting scheme into an ERM, but this translation is appropriate because it gives the correct correspondence between upper-bound formulae for the free energy to be recalled below.

On the other hand, in [20] Panchenko does introduce a family of limit objects, closely related to our use of limiting ERMs. Given the random Gibbs measure $\gamma_{\beta,N}$ on $\{-1,1\}^N$, he draws independent replicas $\sigma^1, \sigma^2, \ldots$ from it and then considers the joint distribution of the whole $(N \times \infty)$-indexed, $\{-1,1\}$-valued random array

$$\left(\sigma_n^\ell\right)_{1 \leq n \leq N, \ell \geq 1}.$$

Whereas we used these replicas to form an empirical measure which is an ERM, Panchenko chooses an arbitrary extension of this to a two-dimensional infinite random array. Letting $N \longrightarrow \infty$, if one considers a subsequence of the γ_N for which these joint distributions converge, then in the limit one obtains a random $\{-1,1\}$-valued array which is row–column exchangeable. Applying Theorem 3.3, this array has the same law as

$$\left(\sigma(U, U_n, V_\ell, W_{n\ell})\right)_{n,\ell \geq 1}$$

for some measurable function $\sigma : [0,1)^4 \longrightarrow \{-1,1\}$, where U, U_n for $n \geq 1$, V_ℓ for $\ell \geq 1$ and $W_{n\ell}$ for $n, \ell \geq 1$ are i.i.d. $\sim U[0,1)$.

Panchenko then uses σ itself as his limit object for the sequence $(\gamma_{\beta,N})_N$. The equivalence between this formalism and the use of directing random measures on $B([0,1), \Pr\{-1,1\})$ is just the equivalence between our Theorem B and Theorem 3.3 described in Section 3.3 above.

6.3. *Formula for the limiting free energy*

A central result of [20] is a formula for the asymptotic expected free energy of models such as (7) in terms of a functional of the directing functions introduced above: see [20], Theorem 2. For the VB model itself the result is as follows.

Theorem 6.1 (Free energy formula). *As $N \longrightarrow \infty$, the expected specific free energy from (8) satisfies*

$$\lim_{N \longrightarrow \infty} F_N = \inf_\sigma \mathcal{P}(\sigma),$$

where for $\sigma : [0,1)^4 \longrightarrow \{-1,1\}$ we have

$$\mathcal{P}(\sigma) := \log 2 + \mathsf{E}^{(1)} \log \mathsf{E}^{(2)} \left(\cosh \beta \sum_{i=1}^{K_1} J_i \sigma(W, U, V_i, X_i) \right)$$

$$- \mathsf{E}^{(1)} \log \mathsf{E}^{(2)} \left(\exp \beta \sum_{i=1}^{K_2} J_i \sigma(W, U, V_i, X_i) \sigma\left(W, U, V_i', X_i'\right) \right),$$

where:

- *all the r.v.s $W, U, V_1, V_2, \ldots, V_1', V_2', \ldots, X_1, X_2, \ldots, X_1', X_2', \ldots$ are i.i.d. $\sim U[0,1)$,*
- *K_1 is an independent Poisson r.v. of mean 2α,*
- *K_2 is an independent Poisson r.v. of mean α,*
- *and the coefficients J_i are chosen independently from the same distribution as before,*

and where

$$\mathsf{E}^{(1)} = \textit{expectation over } W, K_1, K_2, (V_i)_i, \left(V_i'\right)_i \textit{ and } (J_i)_i$$

and

$$\mathsf{E}^{(2)} = \textit{expectation over } U, (X_i)_i \textit{ and } \left(X_i'\right)_i.$$

If $\boldsymbol{\gamma}$ is the random directing measure on $B([0, 1), \Pr\{-1, 1\})$ that corresponds to σ, then the above formula may easily be recast in terms of $\boldsymbol{\gamma}$: it is

$$\log 2 + \mathsf{E} \log \int_B \sum_{\varepsilon_1, \ldots, \varepsilon_{K_1} = \pm 1} \prod_{i=1}^{K_1} f\big(V_i, \{\varepsilon_i\}\big) \left(\cosh \beta \sum_{i=1}^{K_1} J_i \varepsilon_i\right) \boldsymbol{\gamma}(\mathrm{d}f)$$

$$- \mathsf{E} \log \int_B \sum_{\substack{\varepsilon_1, \ldots, \varepsilon_{K_2} = \pm 1 \\ \varepsilon'_1, \ldots, \varepsilon'_{K_2} = \pm 1}} \prod_{i=1}^{K_2} f\big(V_i, \{\varepsilon_i\}\big) f\big(V'_i, \{\varepsilon'_i\}\big) \left(\exp \beta \sum_{i=1}^{K_2} J_i \varepsilon_i \varepsilon'_i\right) \boldsymbol{\gamma}(\mathrm{d}f),$$

where

$$B = B\big([0, 1), \Pr\{-1, 1\}\big),$$

and where E is now the expectation over all the random data γ, K_1, K_2, $(V_i)_i$, $(V'_i)_i$ and $(J_i)_i$.

Another elementary (but tedious) calculation shows that under the correspondence (11) this coincides with the upper-bound expression that appears in [23]: the right-hand side of inequality (3.3) in that paper. It is for the sake of this calculation that one uses the function Φ to define f_k in (10).

Remark. *In* [20] *Panchenko also shows that the quantity above is unchanged if one instead takes the infimum only over those directing functions σ that satisfy an analog of the Aizenman–Contucci stability under cavity dynamics. This modification could also easily be formulated in terms of random directing functions, but we omit it for the sake of brevity.*

6.4. *The analog of ultrametricity*

After the general formalism of Section 3 of [23], Sections 4 and 5 of that paper propose a special class of weighting scheme objects that correspond to the physicists' notion of "replica-symmetry breaking," and conjecture that these give the correct expression for the limiting free energy. Following the prescriptions of the preceding subsections, we can translate this conjecture into a proposal for a class of limiting random directing measures which adapt the classical Parisi ultrametricity ansatz [22] to the setting of dilute models. As before, the necessary calculations are simple but tedious, so we omit the details. Some discussion along these lines is given in [20] for the SK model, rather than for dilute models.

The key objects seem to be the following. Suppose that T is a discrete rooted tree with all leaves at a fixed finite distance from the root. (The discussion that follows can certainly be extended to more general trees, but we omit that here.) Let $*$ be the root and ∂T the set of leaves. Also, let Σ be the Borel σ-algebra of $[0, 1)$. We formulate the following on $[0, 1)$, but it clearly makes sense on any probability space.

Definition 6.2. *A* branching filtration on $([0, 1), \Sigma, \mathrm{Leb})$ indexed by T *is a family of σ-subalgebras $(\Sigma_t)_{t \in T}$ such that*

- $t \le t' \implies \Sigma_t \subseteq \Sigma_{t'}$;
- *for any $t_0, \ldots, t_m$, the σ-algebra Σ_{t_0} is conditionally independent from $\Sigma_{t_1} \vee \cdots \vee \Sigma_{t_m}$ over Σ_s where $s = (t_0 \wedge t_1) \vee (t_0 \wedge t_2) \vee \cdots \vee (t_0 \wedge t_m)$, the closest vertex of T to t_0 which is a common ancestor of t_0 and some other t_i.*

By analogy with ordinary filtrations, the branching filtration is complete *if every Σ_t is complete for Lebesgue measure.*

Given a branching filtration $\boldsymbol{\Sigma} = (\Sigma_t)_{t \in T}$, a branchingale adapted to $\boldsymbol{\Sigma}$ *is a family of integrable $\mathbb{R}$-valued functions $(f_t)_{t \in T}$ on $[0, 1)$ such that*

- f_t *is Σ_t-measurable*;
- $t \le t' \implies f_t = \mathsf{E}(f_{t'} | \Sigma_t)$.

Observe that in this case every root-leaf path $*v_1 v_2 \cdots v_d$ *gives a martingale* $(f_*, f_{v_1}, \ldots, f_{v_r})$ *adapted to the filtration* $(\Sigma_*, \Sigma_{v_1}, \ldots, \Sigma_{v_d})$; *we call the branchingale* homogeneous *if every root-leaf path gives a martingale with the same distribution.*

Sometimes we refer to the whole collection $(f_t, \Sigma_t)_{t \in T}$ as a branchingale.

Remark. *Of course, stochastic processes indexed by trees have been studied before, but I have not been able to find a reference for precisely this notion. Much of the literature concerns tree-indexed Markov processes, as in* [6], *but I do not see why the r.v.s f_v that we will use should have the Markov property.*

Definition 6.3. *A subset* $Y \subseteq B([0, 1), [-1, 1])$ *is* hierarchically distributed *if it equals* $\{f_v : v \in \partial T\}$ *for some* homogeneous *branchingale* $(f_t, \Sigma_t)_{t \in T}$. *The minimal depth of T in such a representation is the* depth *of the set Y.*

Now a simple calculation shows that under the correspondence (11), the special weighting schemes used to formulate the r-step replica-symmetry breaking bound in [23], Section 5, correspond to random measures γ which are a.s. supported on hierarchically distributed sets of depth r, and with the weights given by a Derrida–Ruelle probability cascade that follows the indexing tree.

To be more specific, in their work they consider r.v.s X_t indexed by the leaves t of a tree T of depth r and infinite branching, and specify their joint distribution by constructing a larger family of random variables

$$\left(\eta^{(0)}, \eta^{(1)}_{t_1}, \eta^{(2)}_{t_1 t_2}, \ldots, \eta^{(r-1)}_{t_1 t_2 \cdots t_{r-1}}, \eta^{(r)}_{t_1 t_2 \cdots t_r} \right)$$

indexed by all downwards paths from the root in T, where:

- $\eta^{(r)}_{t_1 \cdots t_r} = X_{t_r}$ for each leaf $t_r \in \partial T$,
- for a shorter path $t_1 t_2 \cdots t_s$, $0 \le s \le r - 1$, the r.v. $\eta^{(s)}_{t_1 t_2 \cdots t_s}$ takes values in the space

$$\underbrace{\Pr\big(\Pr\big(\cdots \Pr(\mathbb{R})\big)\big)}_{r-s},$$

- and for each $t_1 t_2 \cdots t_s$ with $s \le r - 1$, the r.v.s $\eta^{(s+1)}_{t_1 t_2 \cdots t_s t}$ indexed by all the children t of t_s are chosen independently from $\eta^{(s)}_{t_1 t_2 \cdots t_s}$, and similarly the random variables at all further children along distinct ancestral lines are conditionally independent.

Such a structure arises from a homogeneous branchingale $(f_t, \Sigma_t)_{t \in T}$ for which $0 < f_t < 1$ a.s. as follows. Let $\eta^{(r-1)}_{t_1 \cdots t_{r-1}}$ be the conditional distribution of $\Phi^{-1} \circ f_{t_r}$ on $\Sigma_{t_{r-1}}$ for any child t_r of t_{r-1}, where $\Phi(x) = e^x / (e^x + e^{-x})$ as before, and the condition $0 < f_{t_r} < 1$ ensures that this composition is defined a.s. Now let $\eta^{(r-2)}_{t_1 \cdots t_{r-2}}$ be the conditional distribution of $\eta^{(r-1)}_{t_1 \cdots t_{r-1}}$ on $\Sigma_{t_{r-2}}$, and so on. These are then related to the functions f_t themselves in that f_{t_s} is obtained from $\eta^{(s)}_{t_1 \cdots t_s}$ by applying Φ and then taking barycentres $r - s$ times. If one starts instead from the r.v.s $\eta^{(s)}_{t_1 \cdots t_s}$ as above, another simple (but lengthy) iterated appeal to Lemma 3.1 produces a homogeneous branchingale that gives rise to it.

Thus, the natural analog of the Parisi ultrametricity ansatz for the Viana–Bray model seems to be that in the infimum of Theorem 6.1, if one formulates the right-hand side in terms of directing random measures, it is enough to consider directing random measures that are a.s. supported on hierarchically distributed subsets of $B([0, 1), \Pr\{-1, 1\})$.

Acknowledgements

References

[1] D. J. Aldous. Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** (4) (1981) 581–598. MR0637937

[2] D. J. Aldous. On exchangeability and conditional independence. In *Exchangeability in Probability and Statistics* (*Rome*, *1981*) 165–170. North-Holland, Amsterdam, 1982. MR0675972

[3] D. J. Aldous. Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII—1983* 1–198. *Lecture Notes in Math.* **1117**. Springer, Berlin, 1985. MR0883646

[4] L.-P. Arguin. A remark on the infinite-volume Gibbs measures of spin glasses. *J. Math. Phys.* **49** (12) (2008) 125204, 8. MR2484335

[5] L.-P. Arguin and M. Aizenman. On the structure of quasi-stationary competing particle systems. *Ann. Probab.* **37** (3) (2009) 1080–1113. MR2537550

[6] I. Benjamini and Y. Peres. Markov chains indexed by trees. *Ann. Probab.* **22** (1) (1994) 219–243. MR1258875

[7] L. De Sanctis. Random multi-overlap structures and cavity fields in diluted spin glasses. *J. Stat. Phys.* **117** (2004) 785–799. MR2107895

[8] L. De Sanctis and S. Franz. Self-averaging identities for random spin systems. In *Spin Glasses: Statics and Dynamics* 123–142. *Progr. Probab.* **62**. Birkhäuser, Basel, 2009. MR2761987

[9] L. N. Dovbysh and V. N. Sudakov. Gram–de Finetti matrices. In *Problems of the Theory of Probability Distribution VII* 77–86, 238, 244–245. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **119**. Nauka, Leningrad, 1982. MR0666087

[10] M. Fannes, H. Spohn and A. Verbeure. Equilibrium states for mean field models. *J. Math. Phys.* **21** (2) (1980) 355–358. MR0558480

[11] E. Glasner. *Ergodic Theory via Joinings*. Amer. Math. Soc., Providence, RI, 2003. MR1958753

[12] K. Hestir. The Aldous representation theorem and weakly exchangeable non-negative definite arrays. Ph.D. dissertation, Statistic Dept., Univ. California, Berkeley, 1986. MR2635653

[13] D. N. Hoover. Relations on probability spaces and arrays of random variables. Unpublished manuscript, 1979.

[14] D. N. Hoover. Row–columns exchangeability and a generalized model for exchangeability. In *Exchangeability in Probability and Statistics (Rome, 1981)* 281–291. North-Holland, Amsterdam, 1982. MR0675982

[15] O. Kallenberg. On the representation theorem for exchangeable arrays. *J. Multivariate Anal.* **30** (1) (1989) 137–154. MR1003713

[16] O. Kallenberg. Symmetries on random arrays and set-indexed processes. *J. Theoret. Probab.* **5** (4) (1992) 727–765. MR1182678

[17] O. Kallenberg. *Foundations of Modern Probability*, 2nd edition. *Probability and Its Applications (New York)*. Springer, New York, 2002. MR1876169

[18] O. Kallenberg. *Probabilistic Symmetries and Invariance Principles*. *Probability and Its Applications (New York)*. Springer, New York, 2005. MR2161313

[19] M. Mézard, G. Parisi and M. A. Virasoro. *Spin Glass Theory and Beyond*. *World Scientific Lecture Notes in Physics* **9**. World Scientific Publishing, Teaneck, NJ, 1987. MR1026102

[20] D. Panchenko. Spin glass models from the point of view of spin distributions. *Ann. Probab.* **41** (2013) 1315–1361. MR3098679

[21] D. Panchenko. On the Dovbysh–Sudakov representation result. *Electron. Comm. Probab.* **15** (2010) 330–338. MR2679002

[22] D. Panchenko. *The Sherrington–Kirkpatrick Model*. *Springer Monographs in Mathematics*. Springer, Berlin, 2013. MR3052333

[23] D. Panchenko and M. Talagrand. Bounds for diluted mean-fields spin glass models. *Probab. Theory Related Fields* **130** (3) (2004) 319–336. MR2095932

[24] M. Talagrand. *Spin Glasses: A Challenge for Mathematicians. Cavity and Mean Field Models*. 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] **46**. Springer, Berlin, 2003. MR1993891

[25] L. Viana and A. Bray. Phase diagrams for dilute spin glasses. *J. Phys. C: Solid State Phys.* **18** (15) (1985) 3037–3052.