A new characterization of endogeny

Tibor Mach [*] Anja Sturm [†] Jan M. Swart [*]

January 16, 2018

Abstract

Aldous and Bandyopadhyay have shown that each solution to a recursive distributional equation (RDE) gives rise to recursive tree process (RTP), which is a sort of Markov chain in which time has a tree-like structure and in which the state of each vertex is a random function of its descendents. If the state at the root is measurable with respect to the sigma field generated by the random functions attached to all vertices, then the RTP is said to be endogenous. For RTPs defined by continuous maps, Aldous and Bandyopadhyay showed that endogeny is equivalent to bivariate uniqueness, and they asked if the continuity hypothesis can be removed. We answer this question positively. Our main tool is a higher-level RDE that through its n-th moment measures contains all n-variate RDEs. We show that this higher-level RDE has minimal and maximal fixed points with respect to the convex order, and that these coincide if and only if the corresponding RTP is endogenous.

MSC 2010. Primary: 60K35, Secondary: 60J05, 82C22, 60J80.
Keywords. Recursive tree process, endogeny.
Acknowledgement. Work sponsored by grant 16-15238S of the Czech Science Foundation (GA CR).

Contents

[*]The Czech Academy of Sciences, Institute of Information Theory and Automation. Pod vodárenskou věží 4, 18208 Praha 8, Czech Republic; swart@utia.cas.cz

[†]Institute for Mathematical Stochastics, Georg-August-Universität Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany; asturm@math.uni-goettingen.de

1 Recursive distributional equations

Let S be a Polish space and for $k \geq 0$, let S^k denote the space of all ordered sequences $(x_1, \ldots, x_k)$ of elements of S, where by definition S^0 is a set containing a single element, the empty sequence, which we denote by $(\emptyset)$. Let $\mathcal{P}(S)$ denote the space of all probability measures on S, equipped with the topology of weak convergence and its associated Borel-σ-field. It is well-known that $\mathcal{P}(S)$ is a Polish space. A measurable map $g : S^k \to S$ gives rise to a measurable map $\check{g} : \mathcal{P}(S)^k \to \mathcal{P}(S)$ defined as

$$\check{g}(\mu_1, \ldots, \mu_k) := \mu_1 \otimes \cdots \otimes \mu_k \circ g^{-1}, \tag{1.1}$$

where $\mu_1 \otimes \cdots \otimes \mu_k$ denotes product measure and the right-hand side of (1.1) is the image of this under the map g. A more probabilistic way to express (1.1) is to say that if $X_1, \ldots, X_k$ are independent random variables with laws $\mu_1, \ldots, \mu_k$, then $g(X_1, \ldots, X_k)$ has law $\check{g}(\mu_1, \ldots, \mu_k)$. In particular, we let $T_g : \mathcal{P}(S) \to \mathcal{P}(S)$ denote the map

$$T_g(\mu) := \check{g}(\mu, \ldots, \mu). \tag{1.2}$$

Note that T_g is in general nonlinear, unless $k = 1$. Let $\mathcal{G}$ be a measurable set whose elements are measurable maps $g : S^k \to S$, where $k = k_g \geq 0$ may depend on g, and let π be a probability law on $\mathcal{G}$. Then

$$T(\mu) := \int_{\mathcal{G}} \pi(\mathrm{d}g)\, T_g(\mu) \qquad \big(\mu \in \mathcal{P}(S)\big) \tag{1.3}$$

defines a map $T : \mathcal{P}(S) \to \mathcal{P}(S)$. Equations of the form

$$T(\mu) = \mu \tag{1.4}$$

are called *Recursive Distributional Equations* (RDEs). A nice collection of examples of such RDEs can be found in [AB05].

2 Recursive Tree Processes

For $d \in \mathbb{N}_+ := \{1, 2, \ldots\}$, let $\mathbb{T}^d$ denote the space of all finite words $\mathbf{i} = i_1 \cdots i_t$ $(t \geq 0)$ made up from the alphabet $\{1, \ldots, d\}$, and define $\mathbb{T}^\infty$ similarly, using the alphabet $\mathbb{N}_+$. Let $\emptyset$ denote the word of length zero. We view $\mathbb{T}^d$ as a tree with root $\emptyset$, where each vertex $\mathbf{i} \in \mathbb{T}^d$ has d children $\mathbf{i}1, \mathbf{i}2, \ldots$, and each vertex $\mathbf{i} = i_1 \cdots i_t$ except the root has precisely one ancestor $\overleftarrow{\mathbf{i}} := i_1 \cdots i_{t-1}$. We denote the length of a word $\mathbf{i} = i_1 \cdots i_t$ by $|\mathbf{i}| := t$ and set $\mathbb{T}^d_t := \{\mathbf{i} \in \mathbb{T}^d : |\mathbf{i}| < t\}$. For any subtree $\mathbb{U} \subset \mathbb{T}$, we let $\partial\mathbb{U} := \{\mathbf{i} \in \mathbb{T}^d : \overleftarrow{\mathbf{i}} \in \mathbb{U},\ \mathbf{i} \notin \mathbb{U}\}$ denote the outer boundary of $\mathbb{U}$. In particular, $\partial\mathbb{T}^d_t = \{\mathbf{i} \in \mathbb{T}^d : |\mathbf{i}| = t\}$ is the set of all vertices at distance t from the root.

As before, let $\mathcal{G}$ be a measurable set whose elements are measurable maps $g : S^k \to S$, where $k = k_g \geq 0$ may depend on g, let π be a probability law on $\mathcal{G}$, and let T be the operator in (1.3). Following [AB05], we will give a stochastic representation of the operator T and its iterates. Fix some $d \in \mathbb{N}_+ \cup \{\infty\}$ such that $k_g \leq d$ for all $g \in \mathcal{G}$, and to simplify notation write $\mathbb{T} := \mathbb{T}^d$. Let $(\gamma_\mathbf{i})_{\mathbf{i}\in\mathbb{T}}$ be an i.i.d. collection of random maps with common law π, and write $k_\mathbf{i} := k_{\gamma_\mathbf{i}}$, where as before for a map $g \in \mathcal{G}$ we let $k_g \geq 0$ denote the associated integer such that $g : S^{k_g} \to S$. Fix $t \geq 1$ and $\mu \in \mathcal{P}(S)$, and let $(X_\mathbf{i})_{\mathbf{i}\in\partial\mathbb{T}_t}$ be a collection of S-valued random variables such that

$$(X_\mathbf{i})_{\mathbf{i}\in\partial\mathbb{T}_t} \text{ are i.i.d. with common law } \mu \text{ and independent of } (\gamma_\mathbf{i})_{\mathbf{i}\in\mathbb{T}_t}. \tag{2.1}$$

Define $(X_\mathbf{i})_{\mathbf{i}\in\mathbb{T}_t}$ inductively by

$$X_\mathbf{i} = \gamma_\mathbf{i}\big(X_{\mathbf{i}1},\ldots,X_{\mathbf{i}k_\mathbf{i}}\big) \qquad (\mathbf{i}\in\mathbb{T}_t). \tag{2.2}$$

Then it is easy to see that for each $1 \leq s \leq t$,

$$(X_\mathbf{i})_{\mathbf{i}\in\partial\mathbb{T}_s} \text{ are i.i.d. with common law } T^{t-s}(\mu) \text{ and independent of } (\gamma_\mathbf{i})_{\mathbf{i}\in\mathbb{T}_s}. \tag{2.3}$$

where T^n denotes the n-th iterate of the map in (1.3). Also, $X_\emptyset$ (the state at the root) has law $T^t(\mu)$. If μ is a solution of the RDE (1.4), then, by Kolmogorov's extension theorem, there exists a collection $(\gamma_\mathbf{i}, X_\mathbf{i})_{\mathbf{i}\in\mathbb{T}}$ of random variables whose joint law is uniquely characterized by the following requirements:

$$\begin{array}{rl} \text{(i)} & (\gamma_\mathbf{i})_{\mathbf{i}\in\mathbb{T}} \text{ is an i.i.d. collection of } \mathcal{G}\text{-valued r.v.'s with common law } \pi, \\ \text{(ii)} & \text{for each } t\geq 1, \text{ the } (X_\mathbf{i})_{\mathbf{i}\in\partial\mathbb{T}_t} \text{ are i.i.d. with common law } \mu \\ & \text{and independent of } (\gamma_\mathbf{i})_{\mathbf{i}\in\mathbb{T}_t}, \\ \text{(iii)} & X_\mathbf{i} = \gamma_\mathbf{i}\big(X_{\mathbf{i}1},\ldots,X_{\mathbf{i}k_\mathbf{i}}\big) \qquad (\mathbf{i}\in\mathbb{T}). \end{array} \tag{2.4}$$

We call such a collection $(\gamma_\mathbf{i}, X_\mathbf{i})_{\mathbf{i}\in\mathbb{T}}$ a *Random Tree Process* (RTP). We can think of a RTP as a generalization of a Markov chain, where the time index set $\mathbb{T}$ has a tree structure and time flows in the direction of the root. In each step, the new value $X_\mathbf{i}$ is a function of the previous values $X_{\mathbf{i}1},\ldots,X_{\mathbf{i}k_\mathbf{i}}$ plus some independent randomness, represented by the maps $(\gamma_\mathbf{i})_{\mathbf{i}\in\mathbb{T}}$. Following [AB05, Def 7], we say that the RTP corresponding to a solution μ of the RDE (1.4) is *endogenous* if $X_\emptyset$ is measurable w.r.t. the σ-field generated by the random variables $(\gamma_\mathbf{i})_{\mathbf{i}\in\mathbb{T}}$.

3 The n-variate RDE

Let $g : S^k \to S$ with $k \geq 0$ be a measurable map and let $n \geq 1$ be an integer. We can naturally identify the space $(S^n)^k$ with the space of all $n\times k$ matrices $x = (x^j_i)^{j=1,\ldots,n}_{i=1,\ldots,k}$. We let $x^j := (x^j_1,\ldots,x^j_k)$ and $x_i = (x^1_i,\ldots,x^n_i)$ denote the rows and columns of such a matrix, respectively. With this notation, we define an *n-variate map* $g^{(n)} : (S^n)^k \to S^n$ by

$$g^{(n)}(x) := \big(g(x^1),\ldots,g(x^n)\big) \qquad \big(x\in(S^n)^k\big). \tag{3.1}$$

The map $g^{(n)}$ describes n systems that are coupled in such a way that the same map g is applied to each system. We will be interested in the *n-variate map* (compare (1.3))

$$T^{(n)}(\nu) := \int_\mathcal{G} \pi(\mathrm{d}g)\, T_{g^{(n)}}(\nu) \qquad \big(\nu\in\mathcal{P}(S^n)\big). \tag{3.2}$$

and the corresponding *n-variate RDE* (compare (1.4))

$$T^{(n)}(\nu) = \nu. \tag{3.3}$$

The maps $T^{(n)}$ are consistent in the following sense. Let $\nu\big|_{\{i_1,\ldots,i_m\}}$ denote the marginal of ν with respect to the coordinates $i_1,\ldots,i_m$, i.e., the image of ν under the projection $(x_1,\ldots,x_n)\mapsto(x_{i_1},\ldots,x_{i_m})$. Then

$$T^{(n)}(\nu)\big|_{\{i_1,\ldots,i_m\}} = T^{(m)}\big(\nu\big|_{\{i_1,\ldots,i_m\}}\big). \tag{3.4}$$

In particular, if ν solves the n-variate RDE (3.3), then its its one-dimensional marginals $\nu|_{\{m\}}$ $(1 \leq m \leq n)$ solve the RDE (1.4). For any $\mu \in \mathcal{P}(S)$, we let

$$\mathcal{P}(S^n)_\mu := \big\{\nu \in \mathcal{P}(S^n) : \nu|_{\{m\}} = \mu \ \forall 1 \leq m \leq n\big\} \tag{3.5}$$

denote the set of probability measures on S^n whose one-dimensional marginals are all equal to μ. We also let $\mathcal{P}_{\rm sym}(S^n)$ denote the space of all probability measures on S^n that are symmetric with respect to permutations of the coordinates $\{1,\ldots,n\}$, and denote $\mathcal{P}_{\rm sym}(S^n)_\mu := \mathcal{P}_{\rm sym}(S^n) \cap \mathcal{P}(S^n)_\mu$. It is easy to see that $T^{(n)}$ maps $\mathcal{P}_{\rm sym}(S^n)$ into itself. If μ solves the RDE (1.4), then $T^{(n)}$ also maps $\mathcal{P}_{\rm sym}(S^n)_\mu$ into itself.

Given a measure $\mu \in \mathcal{P}(S)$, we define $\overline{\mu}^{(n)} \in \mathcal{P}(S^n)$ by

$$\overline{\mu}^{(n)} := \mathbb{P}\big[(X,\ldots,X) \in \cdot\,\big] \quad \text{where } X \text{ has law } \mu. \tag{3.6}$$

We will prove the following theorem, which is similar to [AB05, Thm 11]. The main improvement compared to the latter is that the implication (ii)⇒(i) is shown without the additional assumption that $T^{(2)}$ is continuous with respect to weak convergence, solving Open Problem 12 of [AB05].

Theorem 1 (Endogeny and bivariate uniqueness) *Let μ be a solution to the RDE (1.4). Then the following statements are equivalent.*

(i) *The RTP corresponding to μ is endogenous.*

(ii) *The measure $\overline{\mu}^{(2)}$ is the unique fixed point of $T^{(2)}$ in the space $\mathcal{P}_{\rm sym}(S^2)_\mu$.*

(iii) $(T^{(n)})^t(\nu) \underset{t\to\infty}{\Longrightarrow} \overline{\mu}^{(n)}$ *for all $\nu \in \mathcal{P}(S^n)_\mu$.*

4 The higher-level RDE

We define a *higher-level map* $\check{T} : \mathcal{P}(\mathcal{P}(S)) \to \mathcal{P}(\mathcal{P}(S))$ by

$$\check{T}(\rho) := \int_{\mathcal{G}} \pi(\mathrm{d}g)\, T_{\check{g}}(\rho) \qquad \big(\rho \in \mathcal{P}(\mathcal{P}(S))\big), \tag{4.1}$$

where for any $g : S^k \to S$, the map $\check{g} : \mathcal{P}(S)^k \to \mathcal{P}(S)$ is defined as in (1.1). Our main tool for proving Theorem 1 is the *higher-level RDE*

$$\check{T}(\rho) = \rho. \tag{4.2}$$

A measure $\rho \in \mathcal{P}(\mathcal{P}(S))$ is the law of a random measure ξ on S. The *n-th moment measure* $\rho^{(n)}$ of such a random measure ξ is defined as

$$\rho^{(n)} := \mathbb{E}\big[\underbrace{\xi \otimes \cdots \otimes \xi}_{n \text{ times}}\big] \quad \text{where } \xi \text{ has law } \rho. \tag{4.3}$$

Here, the expectation of a random measure ξ on S is defined in the usual way, i.e., $\mathbb{E}[\xi]$ is the deterministic measure defined by $\int \phi \,\mathrm{d}\mathbb{E}[\xi] := \mathbb{E}[\int \phi \,\mathrm{d}\xi]$ for any bounded measurable $\phi : S \to \mathbb{R}$. A similar definition applies for measures on S^n. The following lemma links the higher-level map $\check{T}$ to the n-variate maps $T^{(n)}$ of the previous subsection.

Lemma 2 (Moment measures) *Let $n \geq 1$ and let $T^{(n)}$ and $\check{T}$ be defined as in (3.3) and (4.2). Then the n-th moment measure of $\check{T}(\rho)$ is given by*

$$\check{T}(\rho)^{(n)} = T^{(n)}(\rho^{(n)}) \qquad \big(\rho \in \mathcal{P}(\mathcal{P}(S))\big). \tag{4.4}$$

Lemma 2 implies in particular that if ρ solves the higher-level RDE (4.2), then its first moment measure $\rho^{(1)}$ solves the original RDE (1.4). For any $\mu \in \mathcal{P}(S)$, we let

$$\mathcal{P}(\mathcal{P}(S))_\mu := \big\{\rho \in \mathcal{P}(\mathcal{P}(S)) : \rho^{(1)} = \mu\big\} \tag{4.5}$$

denote the set of all ρ whose first moment measure is μ. Note that $\rho \in \mathcal{P}(\mathcal{P}(S))_\mu$ implies $\rho^{(n)} \in \mathcal{P}_{\rm sym}(S^n)_\mu$ for each $n \geq 1$.

We equip $\mathcal{P}(\mathcal{P}(S))$ with the *convex order*. By Theorem 11 in the appendix, two measures $\rho_1, \rho_1 \in \mathcal{P}(\mathcal{P}(S))$ are ordered in the convex order, denoted $\rho_1 \leq_{\rm cv} \rho_2$, if and only if there exists an S-valued random variable X defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and sub-σ-fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$ such that $\rho_i = \mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\mathcal{F}_i] \in \cdot\,\big]$ $(i = 1, 2)$. It is not hard to see that $\mathcal{P}(\mathcal{P}(S))_\mu$ has a minimal and maximal element w.r.t. the convex order. For any $\mu \in \mathcal{P}(S)$, let us define

$$\overline{\mu} := \mathbb{P}\big[\delta_X \in \cdot\,\big] \quad \text{where } X \text{ has law } \mu. \tag{4.6}$$

Clearly $\delta_\mu, \overline{\mu} \in \mathcal{P}(\mathcal{P}(S))_\mu$. Moreover (as will be proved in Section 6 below)

$$\delta_\mu \leq_{\rm cv} \rho \leq_{\rm cv} \overline{\mu} \quad \text{for all} \quad \rho \in \mathcal{P}(\mathcal{P}(S))_\mu. \tag{4.7}$$

In line with notation that has already been introduced in (3.6), the n-th moment measures of δ_μ and $\overline{\mu}$ are given by

$$\delta_\mu^{(n)} = \mathbb{P}\big[(X_1, \ldots, X_n) \in \cdot\,\big] \quad \text{and} \quad \overline{\mu}^{(n)} = \mathbb{P}\big[(X, \ldots, X) \in \cdot\,\big], \tag{4.8}$$

where $X_1, \ldots, X_n$ are i.i.d. with common law μ and X has law μ. The following proposition says that the higher-level RDE (4.2) has a minimal and maximal solution with respect to the convex order.

Proposition 3 (Minimal and maximal solutions) *Let μ be a solution to the RDE (1.4). Then the map $\check{T}$ maps $\mathcal{P}(\mathcal{P}(S))_\mu$ into itself and is monotone w.r.t. the convex order. There exists a unique $\underline{\mu} \in \mathcal{P}(\mathcal{P}(S))_\mu$ such that*

$$\check{T}^t(\delta_\mu) \underset{t\to\infty}{\Longrightarrow} \underline{\mu}, \tag{4.9}$$

where $\Rightarrow$ denotes weak convergence of measures on $\mathcal{P}(S)$, equipped with the topology of weak convergence. The measures $\underline{\mu}$ and $\overline{\mu}$ solve the higher-level RDE (4.2), and any $\rho \in \mathcal{P}(\mathcal{P}(S))_\mu$ that solves the higher-level RDE (4.2) must satisfy

$$\underline{\mu} \leq_{\rm cv} \rho \leq_{\rm cv} \overline{\mu}. \tag{4.10}$$

Since $\underline{\mu}$ and $\overline{\mu}$ solve the higher-level RDE (4.2), there exist RTPs corresponding to $\underline{\mu}$ and $\overline{\mu}$. The following proposition gives an explicit description of these higher-level RTPs.

Proposition 4 (Higher-level RTPs) *Let $(\gamma_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be a RTP corresponding to a solution μ of the RDE (1.4). Set*

$$\xi_{\mathbf{i}} := \mathbb{P}\big[X_{\mathbf{i}} \in \cdot \,|\, (\gamma_{\mathbf{ij}})_{\mathbf{j}\in\mathbb{T}}\big]. \tag{4.11}$$

Then $(\check{\gamma}_{\mathbf{i}}, \xi_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is a RTP corresponding to the solution $\underline{\mu}$ of the higher-level RDE (4.2). Also, $(\check{\gamma}_{\mathbf{i}}, \delta_{X_{\mathbf{i}}})_{\mathbf{i}\in\mathbb{T}}$ is a RTP corresponding to $\overline{\mu}$.

We will derive Theorem 1 from the following theorem, which is our main result.

Theorem 5 (The higher-level RDE) *Let μ be a solution to the RDE (1.4). Then the following statements are equivalent.*

(i) *The RTP corresponding to μ is endogenous.*

(ii) $\underline{\mu} = \overline{\mu}$.

(iii) $\check{T}^t(\rho) \underset{t\to\infty}{\Longrightarrow} \overline{\mu}$ *for all* $\rho \in \mathcal{P}(\mathcal{P}(S))_\mu$.

5 Proof of the main theorem

In this section, we use Lemma 2 and Propositions 3 and 4 to prove Theorems 1 and 5. We need one more lemma.

Lemma 6 (Convergence in probability) *Let $(\gamma_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an endogenous RTP corresponding to a solution μ of the RDE (1.4), and let $(Y_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an independent i.i.d. collection of S-valued random variables with common law μ. For each $t \geq 1$, set $X^t_{\mathbf{i}} := Y_{\mathbf{i}}$ ($\mathbf{i} \in \partial\mathbb{T}_t$), and define $(X^t_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_t}$ inductively by*

$$X^t_{\mathbf{i}} = \gamma_{\mathbf{i}}\big(X^t_{\mathbf{i}1}, \ldots, X^t_{\mathbf{i}k_{\mathbf{i}}}\big) \qquad (\mathbf{i} \in \mathbb{T}_t). \tag{5.1}$$

Then

$$X^t_\emptyset \underset{t\to\infty}{\longrightarrow} X_\emptyset \quad \textit{in probability.} \tag{5.2}$$

Proof The argument is basically the same as in the proof of [AB05, Thm 11 (c)], but for completeness, we give it here. Let $f, g : S \to \mathbb{R}$ be bounded and continuous and let $\mathcal{F}_t$ resp. $\mathcal{F}_\infty$ be the σ-fields generated by $(\gamma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_t}$ resp. $(\gamma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$. Since $X_\emptyset$ and $X^t_\emptyset$ are conditionally independent and identically distributed given $\mathcal{F}_t$,

$$\begin{aligned}\mathbb{E}\big[f(X_\emptyset)g(X^t_\emptyset)\big] &= \mathbb{E}\big[\mathbb{E}\big[f(X_\emptyset)\,\big|\,\mathcal{F}_t\big]\mathbb{E}\big[g(X^t_\emptyset)\,\big|\,\mathcal{F}_t\big]\big] = \mathbb{E}\big[\mathbb{E}\big[f(X_\emptyset)\,\big|\,\mathcal{F}_t\big]\mathbb{E}\big[g(X_\emptyset)\,\big|\,\mathcal{F}_t\big]\big]\\ &\underset{t\to\infty}{\longrightarrow} \mathbb{E}\big[\mathbb{E}\big[f(X_\emptyset)\,\big|\,\mathcal{F}_\infty\big]\mathbb{E}\big[g(X_\emptyset)\,\big|\,\mathcal{F}_\infty\big]\big] = \mathbb{E}\big[f(X_\emptyset)g(X_\emptyset)\big],\end{aligned} \tag{5.3}$$

where we have used martingale convergence and in the last step also endogeny. Since this holds for arbitrary f, g, we conclude that the law of $(X_\emptyset, X^t_\emptyset)$ converges weakly to the law of $(X_\emptyset, X_\emptyset)$, which implies (5.2). ∎

Proof of Theorem 5 If the RTP corresponding to μ is endogenous, then the random variable $\xi_\emptyset$ defined in (4.11) satisfies $\xi_\emptyset = \delta_{X_\emptyset}$. By Proposition 4, $\xi_\emptyset$ and $\delta_{X_\emptyset}$ have laws $\underline{\mu}$ and $\overline{\mu}$, respectively, so (i)⇒(ii). Conversely, if (i) does not hold, then $\xi_\emptyset$ is with positive probability not a delta measure, so (i)⇔(ii).

The implication (iii)⇒(ii) is immediate from the definition of $\underline{\mu}$ in (4.9). To get the converse implication, we observe that by Proposition 3, $\check{T}$ is monotone with respect to the convex order, so (4.7) implies

$$\check{T}^t(\delta_\mu) \leq_{\rm cv} \check{T}^t(\rho) \leq_{\rm cv} \check{T}^t(\overline{\mu}) \qquad (t \geq 0). \tag{5.4}$$

By Lemma 2, $\check{T}^t(\rho) \in \mathcal{P}(\mathcal{P}(S))_\mu$ for each $t \geq 0$, so by Lemma 7 in the appendix, the measures $(\check{T}^t(\rho))_{t\geq 1}$ are tight. By Proposition 3, the left-hand side of (5.4) converges weakly to $\underline{\mu}$ as

$t \to \infty$ while the right-hand side equals $\overline{\mu}$ for each t, so we obtain that any subsequential limit $\check{T}^{t_n}(\rho) \Rightarrow \rho_*$ satisfies $\underline{\mu} \leq_{\rm cv} \rho_* \leq_{\rm cv} \overline{\mu}$. In particular, this shows that (ii)⇒(iii). ∎

Proof of Theorem 1 The implication (iii)⇒(ii) is trivial. By Lemma 2 and the fact that $\underline{\mu}$ and $\overline{\mu}$ solve the higher-level RDE, we see that (ii) implies $\underline{\mu}^{(2)} = \overline{\mu}^{(2)}$. By Proposition 3, $\underline{\mu} \leq_{\rm cv} \overline{\mu}$. Now Lemma 12 from the appendix shows that $\underline{\mu}^{(2)} = \overline{\mu}^{(2)}$ and $\underline{\mu} \leq_{\rm cv} \overline{\mu}$ imply $\underline{\mu} = \overline{\mu}$, so applying Theorem 5 we obtain that (ii)⇒(i).

To complete the proof, we will show that (i)⇒(iii). Let $(\gamma_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be a RTP corresponding μ and let $(Y^1_{\mathbf{i}}, \ldots, Y^n_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an independent i.i.d. collection of S^n-valued random variables with common law ν. For each $t \geq 1$ and $1 \leq m \leq n$, set $X^{m,t}_{\mathbf{i}} := Y^m_{\mathbf{i}}$ ($\mathbf{i} \in \partial\mathbb{T}_t$), and define $(X^{m,t}_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_t}$ inductively as in (5.1). Then $(X^{1,t}_{\emptyset}, \ldots X^{n,t}_{\emptyset})$ has law $(T^{(n)})^t(\nu)$, and using endogeny, Lemma 6 tells us that

$$(X^{1,t}_{\emptyset}, \ldots X^{n,t}_{\emptyset}) \underset{t\to\infty}{\longrightarrow} (X_{\emptyset}, \ldots, X_{\emptyset}) \quad \text{in probability.} \tag{5.5}$$

Since the right-hand side has law $\overline{\mu}^{(n)}$, this completes the proof. ∎

6 Other proofs

In this section, we provide the proofs of Lemma 2 and Propositions 3 and 4, as well as formula (4.7). We start with the latter.

Proof of formula (4.7) Let ξ be a $\mathcal{P}(S)$-valued random variable with law ρ and conditional on ξ, let X be an S-valued random variable with law ξ. Let $\mathcal{F}_0$ be the trivial σ-field, let $\mathcal{F}_1$ be the σ-field generated by ξ, and let $\mathcal{F}_2$ be the σ-field generated by ξ and X. Then $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$. Since $\rho^{(1)} = \mu$, the random variable X has law μ. Now

$$\begin{aligned}
\mathbb{P}\big[\mathbb{P}[X \in \cdot \,|\, \mathcal{F}_0] \in \cdot\,\big] &= \mathbb{P}\big[\mu \in \cdot\,\big] = \delta_\mu, \\
\mathbb{P}\big[\mathbb{P}[X \in \cdot \,|\, \mathcal{F}_1] \in \cdot\,\big] &= \mathbb{P}\big[\xi \in \cdot\,\big] = \rho, \\
\mathbb{P}\big[\mathbb{P}[X \in \cdot \,|\, \mathcal{F}_2] \in \cdot\,\big] &= \mathbb{P}\big[\delta_X \in \cdot\,\big] = \overline{\mu}.
\end{aligned} \tag{6.1}$$

This proves that $\delta_\mu \leq_{\rm cv} \rho \leq_{\rm cv} \overline{\mu}$. ∎

For each $\rho \in \mathcal{P}(\mathcal{P}(S))$ we can find an S-valued random variable X defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as well as a sub-σ-field $\mathcal{H} \subset \mathcal{F}$ such that

$$\rho = \mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\mathcal{H}] \in \cdot\,\big]. \tag{6.2}$$

More generally, we can construct, on some probability space, S-valued random variables $X^1, \ldots, X^n$ that are conditionally independent given a σ-field $\mathcal{H}$, in such a way that (6.2) holds with X replaced by X^i, for all $i = 1, \ldots, n$. Then the law of $(X^1, \ldots, X^n)$ is the n-th moment measure associated with ρ, i.e.,

$$\rho^{(n)} = \mathbb{P}\big[(X^1, \ldots, X^n) \in \cdot\,\big], \tag{6.3}$$

as can be seen by writing

$$\mathbb{E}\Big[\prod_{i=1}^n f_i(X^i)\Big] = \mathbb{E}\Big[\prod_{i=1}^n \mathbb{E}[f_i(X^i)\,|\,\mathcal{H}]\Big] = \int \rho(\mathrm{d}\xi) \int_{S^n} \xi(\mathrm{d}x_1) \cdots \xi(\mathrm{d}x_n)\, f_1(x) \cdots f_n(x_n). \tag{6.4}$$

for arbitrary bounded measurable $f_i : S \to \mathbb{R}$.

Fix $\rho \in \mathcal{P}(\mathcal{P}(S))$. We wish to give a stochastic representation of the probability measure $\check{T}(\rho)$, where $\check{T}$ is defined in (4.2). We start by giving a representation of $T_{\check{g}}(\rho)$, where $g : S^k \to S$ is measurable. Let $X_1, \ldots, X_k$ be S-valued random variables and let $\mathcal{H}_1, \ldots, \mathcal{H}_k$ be σ-fields, such that $(X_1, \mathcal{H}_1), \ldots, (X_k, \mathcal{H}_k)$ are independent and

$$\rho = \mathbb{P}\big[\mathbb{P}[X_i \in \cdot\, |\mathcal{H}_i] \in \cdot\,\big] \qquad (i = 1, \ldots, k). \tag{6.5}$$

Let $\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k$ denote the σ-field generated by $\mathcal{H}_1, \ldots, \mathcal{H}_k$. We claim that

$$T_{\check{g}}(\rho) = \mathbb{P}\big[\mathbb{P}[g(X_1, \ldots, X_k) \in \cdot\, |\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k] \in \cdot\,\big]. \tag{6.6}$$

To see this, set $\xi_i := \mathbb{P}[X_i \in \cdot\, |\mathcal{H}_i]$. Then, conditional on $\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k$, the random variables $X_1, \ldots, X_k$ are independent with laws $\xi_1, \ldots, \xi_k$, respectively, and hence

$$\mathbb{P}[g(X_1, \ldots, X_k) \in \cdot\, |\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k] = \check{g}(\xi_1, \ldots, \xi_k) \quad \text{a.s.} \tag{6.7}$$

Since $\xi_1, \ldots, \xi_k$ are i.i.d. with common law ρ, (6.6) follows. Similarly, if γ is a $\mathcal{G}$-valued random variable with law π, then

$$\check{T}(\rho) = \mathbb{P}\big[\mathbb{P}[\gamma(X_1, \ldots, X_k) \in \cdot\, |\mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k] \in \cdot\,\big], \tag{6.8}$$

as can be seen by integrating (6.6) with respect to π.

Proof of Lemma 2 Let $\xi_1, \ldots, \xi_k$ be i.i.d. with common law ρ and conditional on $\xi_1, \ldots, \xi_k$, let $(X_i^j)_{i=1,\ldots,k}^{j=1,\ldots,n}$ be independent S-valued random variables such that X_i^j has law ξ_i. Let $\mathcal{H}_i$ denote the σ-field generated by ξ_i. Then $\rho = \mathbb{P}[\mathbb{P}[X_i^j \in \cdot\, |\mathcal{H}_i] \in \cdot\,]$ for each i, j, and hence, by (6.3)

$$\rho^{(n)} = \mathbb{P}\big[(X_i^1, \ldots, X_i^n) \in \cdot\,\big] \qquad (j = 1, \ldots, k). \tag{6.9}$$

Set $X_i := (X_i^1, \ldots, X_i^n)$ and $X^j := (X_1^j, \ldots, X_k^j)$. Since $X_1, \ldots, X_k$ are independent with law $\rho^{(n)}$,

$$T_{g^{(n)}}(\rho^{(n)}) = \mathbb{P}\big[g^{(n)}(X_1, \ldots, X_k) \in \cdot\,\big] = \mathbb{P}\big[\big(g(X^1), \ldots, g(X^n)\big) \in \cdot\,\big]. \tag{6.10}$$

Let $\mathcal{H} := \mathcal{H}_1 \vee \cdots \vee \mathcal{H}_k$. Then, by (6.6), $T_{\check{g}}(\rho) = \mathbb{P}\big[\mathbb{P}[g(X^j) \in \cdot\, |\mathcal{H}] \in \cdot\,\big]$. Since moreover $g(X^1), \ldots, g(X^n)$ are conditionally independent given $\mathcal{H}$, by (6.3)

$$T_{\check{g}}(\rho)^{(n)} = \mathbb{P}\big[\big(g(X^1), \ldots, g(X^n)\big) \in \cdot\,\big]. \tag{6.11}$$

Combining this with (6.10), we see that $T_{\check{g}}(\rho)^{(n)} = T_{g^{(n)}}(\rho^{(n)})$ for each $g \in \mathcal{G}$. Now (4.4) follows by integrating w.r.t. π. ∎

Proof of Propositions 3 and 4 We first show that $\check{T}$ is monotone w.r.t. the convex order. Let $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$ satisfy $\rho_1 \leq_{\text{cv}} \rho_2$. Then there exists S-valued random variables $X_1, \ldots, X_k$ and σ-fields $(\mathcal{H}_i^j)_{i=1,\ldots,k}^{j=1,2}$, such that $(X_1, \mathcal{H}_1^1, \mathcal{H}_1^2), \ldots, (X_k, \mathcal{H}_k^1, \mathcal{H}_k^2)$ are independent,

$$\rho_j = \mathbb{P}\big[\mathbb{P}[X_i \in \cdot\, |\mathcal{H}_i^j] \in \cdot\,\big] \qquad (i = 1, \ldots, k,\ j = 1, 2), \tag{6.12}$$

and $\mathcal{H}_i^1 \subset \mathcal{H}_i^2$ for all $i = 1, \ldots, k$. Let γ be an independent $\mathcal{G}$-valued random variable with law π. Then (6.8) says that

$$\check{T}(\rho_j) = \mathbb{P}\big[\mathbb{P}[\gamma(X_1, \ldots, X_k) \in \cdot\, |\mathcal{H}_1^j \vee \cdots \vee \mathcal{H}_k^j] \in \cdot\,\big] \qquad (j = 1, 2). \tag{6.13}$$

Since $\mathcal{H}_1^1 \vee \cdots \vee \mathcal{H}_k^1 \subset \mathcal{H}_1^2 \vee \cdots \vee \mathcal{H}_k^2$, this proves that $\check{T}(\rho_1) \leq_{\rm cv} \check{T}(\rho_2)$.

It will be convenient to combine the proof of the remaining statements of Proposition 3 with the proof of Proposition 4. To check (as claimed in Proposition 4) that $(\check{\gamma}_{\mathbf{i}}, \xi_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is a RTP corresponding to $\underline{\mu}$, we need to check that:

(i) The $(\check{\gamma}_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ are i.i.d.

(ii) For each $t \geq 1$, the $(\xi_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_t}$ are i.i.d. with common law $\underline{\mu}$ and independent of $(\check{\gamma}_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_t}$.

(iii) $\xi_{\mathbf{i}} = \check{\gamma}_{\mathbf{i}}\big(\xi_{\mathbf{i}1}, \ldots, \xi_{\mathbf{i}k_{\mathbf{i}}}\big)$ $(\mathbf{i} \in \mathbb{T})$.

Here (i) follows immediately from the fact that the $(\gamma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ are i.i.d. Since $\xi_{\mathbf{i}}$ depend only on $(\gamma_{\mathbf{ij}})_{\mathbf{j}\in\mathbb{T}}$, it is also clear that the $(\xi_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_t}$ are i.i.d. and independent of $(\gamma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_t}$. To see that their common law is $\underline{\mu}$, we may equivalently show that $\xi_\emptyset$ has law $\underline{\mu}$. Thus, we are left with the task to prove (iii) and

(iv) $\mathbb{P}[\xi_\emptyset \in \cdot\,] = \underline{\mu}$.

Let $\mathcal{F}^{\mathbf{i}}$ denote the σ-field generated by $(\gamma_{\mathbf{ij}})_{\mathbf{j}\in\mathbb{T}}$. Then, for any $\mathbf{i} \in \mathbb{T}$,

$$\xi_{\mathbf{i}} = \mathbb{P}[X_{\mathbf{i}} \in \cdot\,|\mathcal{F}^{\mathbf{i}}] = \mathbb{P}\big[\gamma_{\mathbf{i}}(X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}k_{\mathbf{i}}}) \in \cdot\,\big|\mathcal{F}^{\mathbf{i}}\big]. \tag{6.14}$$

Conditional on $\mathcal{F}^{\mathbf{i}}$, the random variables $X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}k_{\mathbf{i}}}$ are independent with laws $\xi_{\mathbf{i}1}, \ldots, \xi_{\mathbf{i}k_{\mathbf{i}}}$, and hence $\gamma_{\mathbf{i}}(X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}k_{\mathbf{i}}})$ has law $\check{\gamma}_{\mathbf{i}}(\xi_{\mathbf{i}1}, \ldots, \xi_{\mathbf{i}k_{\mathbf{i}}})$, proving (iii).

To prove also (iv), we first need to prove (4.9) from Proposition 3. Fix $t \geq 1$ and for $\mathbf{i} \in \mathbb{T}_t \cup \partial\mathbb{T}_t$, let $\mathcal{F}_t^{\mathbf{i}}$ denote the σ-field generated by $\{\gamma_{\mathbf{ij}} : \mathbf{j} \in \mathbb{T},\ |\mathbf{ij}| < t\}$. In particular, if $\mathbf{i} \in \partial\mathbb{T}_t$, then $\mathcal{F}_t^{\mathbf{i}}$ is the trivial σ-field. Set

$$\xi_{\mathbf{i}}^t := \mathbb{P}[X_{\mathbf{i}} \in \cdot\,|\mathcal{F}_t^{\mathbf{i}}] \qquad (\mathbf{i} \in \mathbb{T}_t \cup \partial\mathbb{T}_t). \tag{6.15}$$

In particular, $\xi_{\mathbf{i}}^t = \mu$ a.s. for $\mathbf{i} \in \partial\mathbb{T}_t$. Arguing as before, we see that

$$\xi_{\mathbf{i}}^t = \check{\gamma}_{\mathbf{i}}\big(\xi_{\mathbf{i}1}^t, \ldots, \xi_{\mathbf{i}k_{\mathbf{i}}}^t\big) \qquad (\mathbf{i} \in \mathbb{T}_t), \tag{6.16}$$

and hence

$$\check{T}^t(\delta_\mu) = \mathbb{P}[\xi_\emptyset^t \in \cdot\,]. \tag{6.17}$$

By martingale convergence,

$$\xi_\emptyset^t = \mathbb{P}[X_\emptyset \in \cdot\,|\mathcal{F}_t^\emptyset] \underset{t\to\infty}{\longrightarrow} \mathbb{P}[X_\emptyset \in \cdot\,|(\gamma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}] = \xi_\emptyset \quad \text{a.s.} \tag{6.18}$$

Combining this with (6.17), we obtain (4.9) where $\underline{\mu}$ is in fact the law of $\xi_\emptyset$, proving (iv) as well. This completes the proof that $(\check{\gamma}_{\mathbf{i}}, \xi_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is a RTP corresponding to $\underline{\mu}$.

The proof that $(\check{\gamma}_{\mathbf{i}}, \delta_{X_{\mathbf{i}}})_{\mathbf{i}\in\mathbb{T}}$ is a RTP corresponding to $\overline{\mu}$ is simpler. It is clear that (i) the $(\check{\gamma}_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ are i.i.d., and (ii) for each $t \geq 1$, the $(\delta_{X_{\mathbf{i}}})_{\mathbf{i}\in\partial\mathbb{T}_t}$ are i.i.d. with common law $\overline{\mu}$ and independent of $(\check{\gamma}_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_t}$. To prove that also (iii) $\delta_{X_{\mathbf{i}}} = \check{\gamma}_{\mathbf{i}}\big(\delta_{X_{\mathbf{i}1}}, \ldots, \delta_{X_{\mathbf{i}k_{\mathbf{i}}}}\big)$ $(\mathbf{i} \in \mathbb{T})$, it suffices to show that for any measurable $g : S^k \to S$,

$$\check{g}(\delta_{x_1}, \ldots, \delta_{x_k}) = \delta_{g(x_1,\ldots,x_k)}. \tag{6.19}$$

By definition, the left-hand side of this equation is the law of $(X_1, \ldots, X_k)$, where $X_1, \ldots, X_k$ are independent with laws $\delta_{x_1}, \ldots, \delta_{x_k}$, so the statement is obvious.

This completes the proof of Proposition 4. Moreover, since the marginal law of a RTP solves the corresponding RDE, our proof also shows that the measures $\underline{\mu}$ and $\overline{\mu}$ solve the higher-level RDE (4.2).

In view of this, to complete the proof of Proposition 3, it suffices to prove (4.10). If ρ solves the higher-level RDE (4.2), then applying $\check{T}^t$ to (4.7), using the monotonicity of $\check{T}$ with respect to the convex order, we see that $\check{T}^t(\delta_\mu) \leq_{\rm cv} \rho \leq_{\rm cv} \overline{\mu}$ for all t. Letting $t \to \infty$, (4.10) follows. ∎

A The convex order

By definition, a *G_δ-set* is a set that is a countable intersection of open sets. By [Bou58, §6 No. 1, Theorem. 1], for a metrizable space S, the following statements are equivalent.

(i) S is Polish.

(ii) There exists a metrizable compactification $\overline{S}$ of S such that S is a G_δ-subset of $\overline{S}$.

(iii) For each metrizable compactification $\overline{S}$ of S, S is a G_δ-subset of $\overline{S}$.

Moreover, a subset $S' \subset S$ of a Polish space S is Polish in the induced topology if and only if S' is a G_δ-subset of S.

Let S be a Polish space. Recall that $\mathcal{P}(S)$ denotes the space of probability measures on S, equipped with the topology of weak convergence. In what follows, we fix a metrizable compactification $\overline{S}$ of S. Then we can identify the space $\mathcal{P}(S)$ (including its topology) with the space of probability measures μ on $\overline{S}$ such that $\mu(S) = 1$. By Prohorov's theorem, $\mathcal{P}(\overline{S})$ is compact, so $\mathcal{P}(\overline{S})$ is a metrizable compactification of $\overline{S}$. Recall the definition of $\mathcal{P}(\mathcal{P}(S))_\mu$ from (4.5).

Lemma 7 (Measures with given mean) *For any $\mu \in \mathcal{P}(S)$, the space $\mathcal{P}(\mathcal{P}(S))_\mu$ is compact.*

Proof Since any $\rho \in \mathcal{P}(\mathcal{P}(\overline{S}))$ whose first moment measure is μ must be concentrated on $\mathcal{P}(S)$, we can identify $\mathcal{P}(\mathcal{P}(S))_\mu$ with the space of probability measures on $\mathcal{P}(\overline{S})$ whose first moment measure is μ. From this we see that $\mathcal{P}(\mathcal{P}(S))_\mu$ is a closed subset of $\mathcal{P}(\mathcal{P}(\overline{S}))$ and hence compact. ∎

We let $\mathcal{C}(\overline{S})$ denote the space of all continuous real functions on $\overline{S}$, equipped with the supremumnorm, and we let $B(\overline{S})$ denote the space of bounded measurable real functions on $\overline{S}$. The following fact is well-known (see, e.g., [Car00, Cor 12.11]).

Lemma 8 (Space of continuous functions) *$\mathcal{C}(\overline{S})$ is a separable Banach space.*

For each $f \in \mathcal{C}(\overline{S})$, we define an affine function $l_f \in \mathcal{C}(\mathcal{P}(\overline{S}))$ by $l_f(\mu) := \int f \, \mathrm{d}\mu$. The following lemma says that all continuous affine functions on $\mathcal{P}(\overline{S})$ are of this form.

Lemma 9 (Continuous affine functions) *A function $\phi \in \mathcal{C}(\mathcal{P}(\overline{S}))$ is affine if and only if $\phi = l_f$ for some $f \in \mathcal{C}(\overline{S})$.*

Proof Let $\phi : \mathcal{P}(\overline{S}) \to \mathbb{R}$ be affine and continuous. Since ϕ is continuous, setting $f(x) := \phi(\delta_x)$ $(x \in \overline{S})$ defines a continuous function $f : \overline{S} \to \mathbb{R}$. Since ϕ is affine, $\phi(\mu) = l_f(\mu)$ whenever μ is a finite convex combination of delta measures. Since such measures are dense in $\mathcal{P}(\overline{S})$ and ϕ is continuous, we conclude that $\phi = l_f$. ∎

Lemma 10 (Lower semi-continuous convex functions) *Let $C \subset \mathcal{C}(\overline{S})$ be convex, closed, and nonempty. Then*

$$\phi := \sup_{f \in C} l_f \tag{A.1}$$

defines a lower semi-continuous convex function $\phi : \mathcal{P}(\overline{S}) \to (-\infty, \infty]$. Conversely, each such ϕ is of the form (A.1).

Proof It is straightforward to check that (A.1) defines a lower semi-continuous convex function $\phi : \mathcal{P}(\overline{S}) \to (-\infty, \infty]$. To prove that every such function is of the form (A.1), let $\mathcal{C}(\overline{S})'$ denote the dual of the Banach space $\mathcal{C}(\overline{S})$, i.e., $\mathcal{C}(\overline{S})'$ is the space of all continuous linear forms $l : \mathcal{C}(\overline{S}) \to \mathbb{R}$. We equip $\mathcal{C}(\overline{S})'$ with the weak-$*$ topology, i.e., the weakest topology that makes the maps $l \mapsto l(f)$ continuous for all $f \in \mathcal{C}(\overline{S})$. Then $\mathcal{C}(\overline{S})'$ is a locally convex topological vector space and by the Riesz-Markov-Kakutani representation theorem, we can view $\mathcal{P}(\overline{S})$ as a convex compact metrizable subset of $\mathcal{C}(\overline{S})'$. Now any lower semi-continuous convex function $\phi : \mathcal{P}(\overline{S}) \to (-\infty, \infty]$ can be extended to $\mathcal{C}(\overline{S})'$ by putting $\phi := \infty$ on the complement of $\mathcal{P}(\overline{S})$. Applying [CV77, Thm I.3] we obtain that ϕ is the supremum of all continuous affine functions that lie below it. By Lemma 9, we can restrict ourselves to continuous affine functions of the form l_f with $f \in \mathcal{C}(\overline{S})$. It is easy to see that $\{f \in \mathcal{C}(\overline{S}) : l_f \leq \phi\}$ is closed and convex, proving that every lower semi-continuous convex function $\phi : \mathcal{P}(\overline{S}) \to (-\infty, \infty]$ is of the form (A.1). ∎

We define

$$\mathcal{C}_{\rm cv}\big(\mathcal{P}(\overline{S})\big) := \big\{\phi \in \mathcal{C}(\mathcal{P}(\overline{S})) : \phi \text{ is convex}\big\} \tag{A.2}$$

If two probability measures $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$ satisfy the equivalent conditions of the following theorem, then we say that they are ordered in the *convex order*, and we denote this as $\rho_1 \leq_{\rm cv} \rho_2$. The fact that $\leq_{\rm cv}$ defines a partial order will be proved in Lemma 13 below. The convex order can be defined more generally for $\rho_1, \rho_2 \in \mathcal{P}(C)$ where C is a convex space, but in the present paper we will only need the case $C = \mathcal{P}(\overline{S})$.

Theorem 11 (The convex order for laws of random probability measures) *Let S be a Polish space and let $\overline{S}$ be a metrizable compactification of S. Then, for $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$, the following statements are equivalent.*

(i) $\displaystyle\int \phi \, \mathrm{d}\rho_1 \leq \int \phi \, \mathrm{d}\rho_2$ *for all* $\phi \in \mathcal{C}_{\rm cv}\big(\mathcal{P}(\overline{S})\big)$.

(ii) *There exists an S-valued random variable X defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and sub-σ-fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$ such that $\rho_i = \mathbb{P}\big[\mathbb{P}[X \in \cdot \,|\mathcal{F}_i] \in \cdot\,\big]$ $(i = 1, 2)$.*

Proof For any probability kernel P on $\mathcal{P}(\overline{S})$, measure $\rho \in \mathcal{P}(\overline{S})$, and function $\phi \in \mathcal{C}(\mathcal{P}(\overline{S}))$, we define $\rho P \in \mathcal{P}(\mathcal{P}(\overline{S}))$ and $P\phi \in B(\mathcal{P}(\overline{S}))$ by

$$\rho P := \int \rho(\mathrm{d}\mu) P(\mu, \,\cdot\,) \quad \text{and} \quad Pf := \int P(\,\cdot\,, \mathrm{d}\mu) f(\mu). \tag{A.3}$$

By definition, a *dilation* is a probability kernel P such that $P l_f = l_f$ for all $f \in \mathcal{C}(\overline{S})$.

As in the proof of Lemma 10, we can view $\mathcal{P}(\overline{S})$ as a convex compact metrizable subset of the locally convex topological vector space $\mathcal{C}(\overline{S})'$. Then [Str65, Thm 2] tells us that (i) is equivalent to:

(iii) There exists a dilation P on $\mathcal{P}(\overline{S})$ such that $\rho_2 = \rho_1 P$.

To see that this implies (ii), let ξ_1, ξ_2 be $\mathcal{P}(\overline{S})$-valued random variables such that ξ_1 has law ρ_1 and the conditional law of ξ_2 given ξ_1 is given by P. Let $\mathcal{F}_1$ be the σ-field generated by ξ_1, let $\mathcal{F}_2$ be the σ-field generated by (ξ_1, ξ_2), and let X be an $\overline{S}$-valued random variable whose conditional law given $\mathcal{F}_2$ is given by ξ_2. Then

$$\mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\mathcal{F}_2] \in \cdot\,\big] = \mathbb{P}[\xi_2 \in \cdot\,] = \rho_1 P = \rho_2. \tag{A.4}$$

For $f \in \mathcal{C}(\overline{S})$ and $\mu \in \mathcal{P}(\overline{S})$, write $l_f(\mu) := \int f \,\mathrm{d}\mu$. Since P is a dilation

$$\mathbb{E}[f(X)\,|\,\mathcal{F}_1] = \mathbb{E}\big[\mathbb{E}[f(X)\,|\,\mathcal{F}_2]\,\big|\,\mathcal{F}_1\big] = \mathbb{E}\big[l_f(\xi_2)\,\big|\,\mathcal{F}_1\big] = \int P(\xi_1, \mathrm{d}\mu) l_f(\mu) = l_f(\xi_1) \tag{A.5}$$

for all $f \in \mathcal{C}(\overline{S})$, and hence

$$\mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\mathcal{F}_1] \in \cdot\,\big] = \mathbb{P}\big[\xi_1 \in \cdot\,\big] = \rho_1. \tag{A.6}$$

We note that since $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$, we have $\xi_1, \xi_2 \in \mathcal{P}(S)$ a.s. and hence $X \in S$ a.s. This proves the implication (iii)⇒(ii).

To complete the proof, it suffices to show that (ii)⇒(i). By Lemma 10, each $\phi \in \mathcal{C}_{\mathrm{cv}}(\mathcal{P}(\overline{S}))$ is of the form $\phi = \sup_{f \in C} l_f$ for some $C \subset \mathcal{C}(\overline{S})$. Then (ii) implies

$$\begin{aligned}\int \phi \,\mathrm{d}\rho_1 &= \mathbb{E}\big[\sup_{f\in C} \mathbb{E}[f(X)\,|\,\mathcal{F}_1]\big] = \mathbb{E}\big[\sup_{f\in C} \mathbb{E}\big[\mathbb{E}[f(X)\,|\,\mathcal{F}_2]\,\big|\,\mathcal{F}_1\big]\big] \\ &\leq \mathbb{E}\big[\mathbb{E}\big[\sup_{f\in C} \mathbb{E}[f(X)\,|\,\mathcal{F}_2]\,\big|\,\mathcal{F}_1\big]\big] = \mathbb{E}\big[\sup_{f\in C} \mathbb{E}[f(X)\,|\,\mathcal{F}_2]\big] = \int \phi \,\mathrm{d}\rho_2.\end{aligned} \tag{A.7}$$

∎

The n-th moment measure $\rho^{(n)}$ associated with a probability measure $\rho \in \mathcal{P}(\mathcal{P}(\overline{S}))$ has been defined in (4.3). The following lemma links the first and second moment measures to the convex order.

Lemma 12 (First and second moment measures) *Let S be a Polish space. Assume that $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$ satisfy $\rho_1 \leq_{\mathrm{cv}} \rho_2$. Then $\rho_1^{(1)} = \rho_2^{(1)}$ and*

$$\int \rho_1^{(2)}(\mathrm{d}x, \mathrm{d}y) f(x) f(y) \leq \int \rho_2^{(2)}(\mathrm{d}x, \mathrm{d}y) f(x) f(y) \qquad \big(f \in B(S)\big). \tag{A.8}$$

If $\rho_1 \leq_{\mathrm{cv}} \rho_2$ and (A.8) holds with equality for all bounded continuous $f : S \to \mathbb{R}$, then $\rho_1 = \rho_2$.

Proof By Theorem 11, there exists an $\overline{S}$-valued random variable X defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and sub-σ-fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$ such that $\rho_i = \mathbb{P}\big[\mathbb{P}[X \in \cdot\,|\mathcal{F}_i] \in \cdot\,\big]$ $(i = 1, 2)$. Since for each $f \in B(S)$

$$\int \rho_1^{(1)}(\mathrm{d}x) f(x) = \mathbb{E}\big[\mathbb{E}[f(X)\,|\mathcal{F}_1]\big] = \mathbb{E}[f(X)] = \mathbb{E}\big[\mathbb{E}[f(X)\,|\mathcal{F}_2]\big] = \int \rho_1^{(1)}(\mathrm{d}x) f(x), \tag{A.9}$$

we see that $\rho_1^{(1)} = \rho_2^{(1)}$. Fix $f \in B(S)$ and set $M_i := \mathbb{E}[f(X)\,|\mathcal{F}_i]$ $(i = 1, 2)$. Then

$$\begin{aligned}\int \rho_2^{(2)}(\mathrm{d}x, \mathrm{d}y) f(x) f(y) &= \mathbb{E}\big[\mathbb{E}[f(X)\,|\mathcal{F}_2]^2\big] = \mathbb{E}[M_2^2] \\ &= \mathbb{E}[M_1^2] + \mathbb{E}\big[(M_2 - M_1)^2\big] \geq \mathbb{E}[M_1^2] = \int \rho_1^{(2)}(\mathrm{d}x, \mathrm{d}y) f(x) f(y),\end{aligned} \tag{A.10}$$

proving (A.8). Let $\overline{S}$ be a metrizable compactification of S. If $\rho_1 \leq_{\rm cv} \rho_2$ and (A.8) holds with equality for all bounded continuous $f : S \to \mathbb{R}$, then (A.10) tells us that $M_1 = M_2$ for each $f \in \mathcal{C}(\overline{S})$, i.e.,

$$\mathbb{E}[f(X)\,|\mathcal{F}_1] = \mathbb{E}[f(X)\,|\mathcal{F}_2] \text{ a.s. for each } f \in \mathcal{C}(\overline{S}). \tag{A.11}$$

Using Lemma 8, we can choose a countable dense set $\mathcal{D} \subset \mathcal{C}(\overline{S})$. Then $\mathbb{E}[f(X)\,|\mathcal{F}_1] = \mathbb{E}[f(X)\,|\mathcal{F}_2]$ for all $f \in \mathcal{D}$ a.s. and hence $\mathbb{P}[X \in \cdot\,|\mathcal{F}_1] = \mathbb{E}[X \in \cdot\,|\mathcal{F}_2]$ a.s., proving that $\rho_1 = \rho_2$. ∎

The following lemma shows that the convex order is a partial order,

Lemma 13 (Convex functions are distribution determining) *If $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(\overline{S}))$ satisfy $\int \phi \, \mathrm{d}\rho_1 = \int \phi \, \mathrm{d}\rho_2$ for all $\phi \in \mathcal{C}_{\rm cv}(\mathcal{P}(\overline{S}))$, then $\rho_1 = \rho_2$.*

Proof For any $f \in \mathcal{C}(\overline{S})$ and $\rho \in \mathcal{P}(\mathcal{P}(\overline{S}))$,

$$\int_{\overline{S}^2} \rho^{(2)}(\mathrm{d}x, \mathrm{d}y) f(x) f(y) = \int_{\mathcal{P}(\overline{S})} \rho(\mathrm{d}\mu) \int_{\overline{S}^2} \mu(\mathrm{d}x)\mu(\mathrm{d}y) f(x) f(y) = \int_{\mathcal{P}(\overline{S})} \rho(\mathrm{d}\mu) l_f(\mu)^2. \tag{A.12}$$

Therefore, since l_f^2 is a convex function, $\int \phi \, \mathrm{d}\rho_1 = \int \phi \, \mathrm{d}\rho_2$ for all $\phi \in \mathcal{C}_{\rm cv}(\mathcal{P}(\overline{S}))$ implies equality in (A.8) and hence, by Lemma 12, $\rho_1 = \rho_2$. ∎

Acknowledgements

We thank Wolfgang Löhr and Jan Seidler for their help with Lemmas 9 and 10, respectively.

References

[AB05] D.J. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equations. *Ann. Appl. Probab.* 15(2) (2005), 1047–1110.

[Bou58] N. Bourbaki. *Éléments de Mathématique. VIII. Part. 1: Les Structures Fondamentales de l'Analyse. Livre III: Topologie Générale. Chap. 9: Utilisation des Nombres Réels en Topologie Générale. 2iéme éd.* Actualités Scientifiques et Industrielles 1045. Hermann & Cie, Paris, 1958.

[Car00] N.L. Carothers. *Real Analysis.* Cambridge University Press, 2000.

[CV77] C. Castaing and M. Valadier. *Convex Analysis and Measurable Multifunctions.* Springer, Berlin, 1977.

[Str65] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Stat.* 36 (1965), 423–439.