Sur une version finiment échangeable du théorème de Hewitt et Savage

Guillaume Carlier [a]

Travail commun avec Gero Friesecke et Daniela Vögler (TUM Münich).
Séminaire de Mathématiques Appliquées, Collège de France, 26 novembre 2021.

[a]CEREMADE, Université Paris Dauphine et MOKAPLAN (Inria-Dauphine).

Introduction

Let $N \geq 2$ and $(Z_1, \ldots, Z_N)$ be N random variables (with values in a metrizable, complete and separable space X), it is said to be **exchangeable** if for every permutation $\sigma \in \mathcal{S}_N$, the law of $(Z_{\sigma(1)}, \ldots, Z_{\sigma(N)})$ is the same as that of $(Z_1, \ldots, Z_N)$.

Of course, i.i.d. samples are exchangeable but the converse is false: $Z_i = Z$, $Z_i = G(Y_i, U)$ with the Y_i i.i.d. and independent from U (common noise), urns without replacement...

A sequence of random variables $(Z_i)_{i\geq 1}$, is exchangeable whenever permuting finitely many coordinates does not change its law.

De Finetti in the 1930's (in the case of binary variables) and Hewitt and Savage in the 1950's showed that laws of (infinite) exhangeable sequences are mixtures (convex combinations) of independent laws. Roughly speaking

$$\mathcal{L}(Z) = \int_{\mathcal{P}(X)} \lambda^{\otimes\infty} \mathrm{d}\alpha(\lambda)$$

for some (actually unique) $\alpha \in \mathcal{P}(\mathcal{P}(X))$. Interpretation: the Z_i's are independent conditionally on some common noise.

Such a result cannot be true for finitely exchangeable families, e.g. $N = 2$, $X = \{0, 1\}$, if the law of (Z_1, Z_2) is $\mu = \frac{1}{2}(\delta_{(1,0)} + \delta_{(0,1)})$, if we had

$$\mu = \int_0^1 ((1-t)\delta_0 + t\delta_1) \otimes ((1-t)\delta_0 + t\delta_1)\mathrm{d}\alpha(t)$$

with $\mu(\{(0,0)\}) = \mu(\{(1,1)\}) = 0$, this would require

$$0 = \int_0^1 t^2 d\alpha(t) = \int_0^1 (1-t)^2 d\alpha(t)$$

which is clearly impossible.

The Diaconis-Freedman bound: if μ is the law of a finite exchangeable sequence $(Z_1, \ldots, Z_N)$ there is an $\alpha \in \mathcal{P}(\mathcal{P}(X))$ such that

$$\|\mu - \int_{\mathcal{P}(X)} \lambda^{\otimes N} \mathrm{d}\alpha(\lambda)\|_{\mathrm{TV}} = O(\frac{1}{N}).$$

Very much studied, Diaconis, Aldous, Kallenberg, just to name a few.

Exact representations for finite exchangeable laws are also known:

- in terms of mixtures of laws of urn sequences, Kerns and Székely, Kallenberg,
- in terms of "mixtures" of independent laws but with a signed α, Jaynes, Kerns and Székely, Janson, Konstantopoulos and Yuan

Goal in the sequel: give a DeFinetti-Hewitt-Savage like representation for exchangeable laws on X^N (or their k-point marginals, $k \leq N$) in terms of mixtures of some explicit polynomials of measures with correlated correction terms. Equivalent to the approach with laws of urns, but more explicit and emphasizes the structure of extreme points through a certain polynomial bijective parametrization of these extreme measures.

Motivations from optimal transport (but there are other ones, MFGs see Pierre-Louis Lions' lectures at Collège de France in 2007). Optimal transport with several marginals and a symmetric cost. Typical case: Coulomb cost, minimize

$$\int_{\mathbf{R}^{3N}} \sum_{i \neq j} \frac{1}{|x_i - x_j|} \mathrm{d}\gamma(x_1, \ldots, x_N)$$

among (symmetric without loss of generality) measures γ with first marginal ρ (given).

Has received a lot of attention since the papers of Buttazzo-De Pascale-Gori-Giorgi (2012) and Cotar-Friesecke-Klüppelberg (2013).

Motivated by the fact that the semiclassical limit (strictly correlated regime) of the Hohenberg-Kohn functional in density functional theory (DFT) is at least formally (for rigorous justifications see Cotar, Friesecke, Klüppelberg, (2018), Lewin (2018)). The value of the above problem (suitably rescaled i.e. divided by $\binom{N}{2}$) reads

$$C_N(\rho) := \inf_{\mu} \int_{\mathbf{R}^6} \frac{1}{|x-y|} \mathrm{d}\mu(x,y)$$

constrained by the fact that μ is the 2-point marginal of a probability on γ on $(\mathbf{R}^3)^N$, symmetric and having ρ as first marginal.

When $N = 2$ (Buttazzo-De Pascale-Gori-Giorgi), if $\rho \in L^1$, the solution is given by a map (essentially involutive) as in Brenier's theorem, $y = T(x) = x + \frac{\nabla\varphi(x)}{|\nabla\varphi(x)|^{3/2}}$ (where φ solves the dual problem), the position of the second electron is a deterministic function of the position of the first one.

When $N = \infty$ (Cotar-Friesecke-Pass), using Hewitt and Savage, what remains from the symmetry condition is just that $\mu = \int_{\mathcal{P}(\mathbf{R}^3)} (\lambda \otimes \lambda) \mathrm{d}\alpha(\lambda)$ for some α so

$$C_\infty(\rho) = \inf_{\alpha : \int \lambda d\alpha(\lambda) = \rho} \left\{ \int_{\mathcal{P}(\mathbf{R}^3)} \int_{\mathbf{R}^6} \frac{1}{|x-y|} \, \mathrm{d}\lambda(x) \, \mathrm{d}\lambda(x) \mathrm{d}\alpha(\lambda) \right\}$$

but (the Coulomb kernel has a positive Fourier Transform)

$$\lambda \mapsto \int_{\mathbf{R}^6} \frac{1}{|x-y|} \, \mathrm{d}\lambda(x) \, \mathrm{d}\lambda(y) \text{ is convex}$$

So

$$C_\infty(\rho) = \int_{\mathbf{R}^6} \frac{1}{|x-y|}\ \mathrm{d}\rho(x)\ \mathrm{d}\rho(y)$$

and $\mu = \rho \otimes \rho$ is the optimal plan. It is optimal to draw x and y independently according to ρ.

This is in (extremely) sharp contrast with the case $N = 2$ where the optimal plan is supported by the graph of a transport map.

What happens for $N = 100$ or $N = 500$ (and more general costs with k-body interactions, $2 \leq k \leq N$...)?

Outline

① Some special measure-valued polynomials

② A representation à la Hewitt and Savage, extremal symmetric laws

③ Symmetric multi-marginal optimal transport

Measure valued polynomials

X is a complete and separable metric space, $\mathcal{P}(X)$ the set of Borel probability measures on X. We fix $N \geq 2$ and $k \in \{1, \ldots, N\}$, probability measures on X^k will be called k-plans. Given $\gamma \in \mathcal{P}(X^N)$ we denote by $M_k\gamma$ the k-point-marginal of γ, i.e.,

$$(M_k\gamma)(A) := \gamma(A \times X^{N-k}) \text{ for every Borel subset } A \text{ of } X^k \quad (1)$$

(with the convention $M_N\gamma = \gamma$).

For $\gamma \in \mathcal{P}(X^N)$ and $\sigma \in \mathcal{S}_N$, the measure $\gamma^\sigma \in \mathcal{P}(X^N)$ is defined by

$$\int_{X^N} \varphi \mathrm{d}\gamma^\sigma = \int_{X^N} \varphi(x_{\sigma(1)}, \ldots, x_{\sigma(N)}) \mathrm{d}\gamma(x_1, \ldots, x_N)$$

for every test-function $\varphi \in C_b(X^N)$. A measure $\gamma \in \mathcal{P}(X^N)$ is called symmetric if $\gamma = \gamma^\sigma$ for every $\sigma \in \mathcal{S}_N$. If $\gamma \in \mathcal{P}(X^N)$ is arbitrary, its symmetrization $S_N\gamma$ is given by

$$S_N\gamma := \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \gamma^\sigma. \tag{2}$$

The set of symmetric N-plans is denoted by $\mathcal{P}_{\mathrm{sym}}(X^N)$:

$$\mathcal{P}_{\mathrm{sym}}(X^N) := S_N(\mathcal{P}(X^N)) = \{\gamma \in \mathcal{P}(X^N) \ : \ \gamma = S_N\gamma\}. \tag{3}$$

N-representable k-plans:

For $N \geq 2$ and $k \in \{1, ..., N\}$, a k-plan $\mu_k \in \mathcal{P}(X^k)$ is said to be N-representable if it is the k-point marginal of a symmetric N-plan, that is to say if there exists $\gamma \in \mathcal{P}_{\text{sym}}(X^N)$ such that $\mu_k = M_k\gamma$. We denote by $\mathcal{P}_{N-\text{rep}}(X^k)$ the set of N-representable k-plans, i.e.:

$$\mathcal{P}_{N-\text{rep}}(X^k) = M_k S_N(\mathcal{P}(X^N)) = M_k(\mathcal{P}_{\text{sym}}(X^N)).$$

In other words, a symmetric N-plan $\gamma \in \mathcal{P}_{\text{sym}}(X^N)$ is the law of a finite exchangeable random sequence $(Z_1, \ldots, Z_N)$ with values in X^N, whereas $\mu_k = M_k\gamma \in \mathcal{P}_{N-\text{rep}}(X^k)$ is the law of its first k-components $(Z_1, \ldots, Z_k)$.

Computations from scratch: start with a Dirac mass. Let $x = (x_1, \dots, x_N) \in X^N$, $2 \le k \le N$ and

$$\mu_k := M_k S_N \delta_x \tag{4}$$

we claim that μ_k is a polynomial of degree k in

$$\lambda := M_1 S_N \delta_x = M_1 \mu_k = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}$$

Of course $\mu_1 = \lambda$, and for $k = 2$

$$\mu_2 = \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \delta_{(x_{\sigma(1)}, x_{\sigma(2)})} = \frac{1}{N(N-1)} \sum_{i \neq j} \delta_{(x_i, x_j)}$$

$$= \frac{N}{N-1} \lambda^{\otimes 2} - \frac{1}{N-1} \mathrm{id}^{\otimes 2}{}_{\#} \lambda$$

where $\lambda^{\otimes 2} = \lambda \otimes \lambda$ (quadratic) and $\mathrm{id}^{\otimes 2}{}_{\#}\lambda = (\mathrm{id}, \mathrm{id})_{\#}\lambda$ (linear).

A similar computation gives by simple inclusion exclusion

$$\mu_3 = \frac{1}{N(N-1)(N-2)} \sum_{i_1,i_2,i_3 \text{pairwise distinct}} \delta_{(x_{i_1},x_{i_2},x_{i_3})}$$

$$= \frac{N^2}{(N-1)(N-2)} \left[\lambda^{\otimes 3} - \frac{3}{N} S_3\Big((\mathrm{id}_{\#}^{\otimes 2}\,\lambda) \otimes \lambda\Big) + \frac{2}{N^2}\,\mathrm{id}_{\#}^{\otimes 3}\,\lambda\right]$$

General polynomial ansatz: develop μ_k on a basis of terms which are tensor products of measures of the form $\lambda^{\otimes p}$ and $\mathrm{id}_{\#}^{\otimes q}\,\lambda$, where

$$\int_{X^q} \varphi \,\mathrm{d}(\mathrm{id}_{\#}^{\otimes q}\,\lambda) = \int_X \varphi(x,\ldots,x)\mathrm{d}\lambda(x)$$

for every $\varphi \in C_b(X^q)$.

For $4 \leq k \leq N$, combinatorics become trickier because there can be many ways (even up to symmetrization) a term of degree $l \leq k$ may arise: e.g. if $k = 4$ both terms $(\mathrm{id}_{\#}^{\otimes 3}\lambda) \otimes \lambda$ and $(\mathrm{id}_{\#}^{\otimes 2}\lambda) \otimes (\mathrm{id}_{\#}^{\otimes 2}\lambda)$ have degree 2. Quite different, for $\varphi \in C_b(X^4)$

$$\int_{X^4} \varphi((\mathrm{id}_{\#}^{\otimes 3}\lambda) \otimes \lambda) = \int_{X^2} \varphi(x, x, x, y)\mathrm{d}\lambda(x)\mathrm{d}\lambda(y)$$

whereas

$$\int_{X^4} \varphi((\mathrm{id}_{\#}^{\otimes 2}\lambda) \otimes (\mathrm{id}_{\#}^{\otimes 2}\lambda)) = \int_{X^2} \varphi(x, x, y, y)\mathrm{d}\lambda(x)\mathrm{d}\lambda(y).$$

Easy to see that the ansatz is justified, recursive formula, for $1 \leq k \leq N-1$,

$$\mu_{k+1} = \frac{N}{N-k}\mu_k \otimes \lambda - \frac{1}{N-k}\sum_{j=1}^{k} R_{j\,\#}\mu_k \tag{5}$$

where $R_j : X^k \to X^{k+1}$ is given by $R_j(z_1, \ldots, z_k) := (z_1, \ldots, z_k, z_j)$. Closed form formula for μ_k as a polynomial of degree k function of λ requires some considerations on partitions.

Let $j \geq 1$, a partition of j of length n is a vector $\mathbf{p} = (p_1, ..., p_n) \in \mathbf{N}^{*n}$ such that $\sum_{i=1}^{n} p_i = j$, $p_1 \geq \ldots \geq p_n$. For any partition $\mathbf{p}$ we denote its length by $n(\mathbf{p})$; e.g. partitions of 4

$$1+1+1+1$$

$$2+1+1$$

$$2+2$$

$$3+1$$

$$4.$$

It will be convenient to view a partition $\mathbf{p}$ as a map from the set of its component indices to $\mathbf{N}^*$; the range $\operatorname{Ran}\mathbf{p}$ of this map is the set of values taken by the components, and $|\mathbf{p}^{-1}(q)|$ denotes the number of components with value q. For example, for $\mathbf{p} = (2, 1, 1)$ and $q = 1$, $|\mathbf{p}^{-1}(q)| = 2$.

Theorem 1 *Let $N \geq 2$, and $2 \leq k \leq N$, for $x \in X^N$, then $\mu_k := M_k S_N \delta_x$ can be written in terms of $\lambda = N^{-1} \sum_{i=1}^{N} x_i$ as*

$$\mu_k = \frac{N^{k-1}}{\prod_{i=1}^{k-1}(N-i)} \left[\lambda^{\otimes k} + \sum_{j=1}^{k-1} \frac{(-1)^j}{N^j} S_k P_j^{(k)}(\lambda) \right] =: F_{N,k}(\lambda) \tag{6}$$

$$P_j^{(k)}(\lambda) = \sum_{\substack{\mathbf{p}=(p_1,\ldots p_{n(\mathbf{p})}) partition \\ of\ j with\ j+n(\mathbf{p}) \leq k}} d_{\mathbf{p}}^{(k)} \ \mathrm{id}_{\#}^{\otimes(p_1+1)} \lambda \otimes \ldots \otimes \mathrm{id}_{\#}^{\otimes(p_{n(\mathbf{p})}+1)} \lambda \otimes \lambda^{\otimes(k-j-n(\mathbf{p}))} \tag{7}$$

$$d_{\mathbf{p}}^{(k)} = \frac{k!}{(k-j-n(\mathbf{p}))!} \prod_{i=1}^{n(\mathbf{p})} \frac{1}{p_i+1} \prod_{q \in Ran\,\mathbf{p}} \frac{1}{(|\mathbf{p}^{-1}(q)|)!} \quad . \tag{8}$$

For instance

$$
\begin{aligned}
F_{N,5}(\lambda) = & \frac{N^5(N-5)!}{N!}\Bigg[\lambda^{\otimes 5} - \frac{10}{N} S_5\, \mathrm{id}_{\#}^{\otimes 2}\lambda \otimes \lambda^{\otimes 3} \\
& + \frac{20\, S_5\, \mathrm{id}_{\#}^{\otimes 3}\lambda \otimes \lambda^{\otimes 2} + 15\, S_5\, \mathrm{id}_{\#}^{\otimes 2}\lambda \otimes \mathrm{id}_{\#}^{\otimes 2}\lambda \otimes \lambda}{N^2} \\
& - \frac{30\, S_5\, \mathrm{id}_{\#}^{\otimes 4}\lambda \otimes \lambda + 20\, S_5\, \mathrm{id}_{\#}^{\otimes 3}\lambda \otimes \mathrm{id}_{\#}^{\otimes 2}\lambda}{N^3} + \frac{24}{N^4}\, \mathrm{id}_{\#}^{\otimes 5}\lambda\Bigg].
\end{aligned}
$$

The leading term is the independent one $\lambda^{\otimes k}$ the next ones are correlated correction terms with $P_j^{(k)}(\lambda)$ of degree $k-j$, the coefficients satisfy the sum rule (Stirling numbers)

$$\sum_{\substack{\mathbf{p}\text{ partition of } j \\ \text{with } j+n(\mathbf{p})\leq k}} d_{\mathbf{p}}^{(k)} = \sum_{1\leq i_1<\ldots<i_j\leq k-1} i_1 \cdots i_j =: c_j^{(k)}. \tag{9}$$

Note also that $F_{N,k}(\lambda)$ can be defined for an arbitrary $\lambda \in \mathcal{P}(X)$ (not only an empirical measure) and $F_{N,k}$ is narrowly continuous.

Retaining only the mean-field term gives in (6)

$$F_{N,k}(\lambda) = \frac{N^{k-1}}{\prod_{j=1}^{k-1}(N-j)} \left(\lambda^{\otimes k} + \varepsilon_{N,k}(\lambda)\right) \text{ with } \|\varepsilon_{N,k}(\lambda)\|_{\mathrm{TV}} \leq \frac{C_k}{N} \tag{10}$$

as in Diaconis and Freedman (but with a bad constant...).

Keeping the first p correction terms ($p \in \{1, ..., k-2\}$) we have

$$F_{N,k}(\lambda) = \frac{N^{k-1}}{\prod_{j=1}^{k-1}(N-j)} \left[\lambda^{\otimes k} + \sum_{j=1}^{p} \frac{(-1)^j}{N^j} S_k \, P_j^{(k)}(\lambda) + \varepsilon_{N,k,p}(\lambda) \right]$$

$$\text{with } \|\varepsilon_{N,k,p}(\lambda)\|_{\mathrm{TV}} \leq \frac{C_k}{N^{p+1}}, \tag{11}$$

with constants C_k independent of N and p.

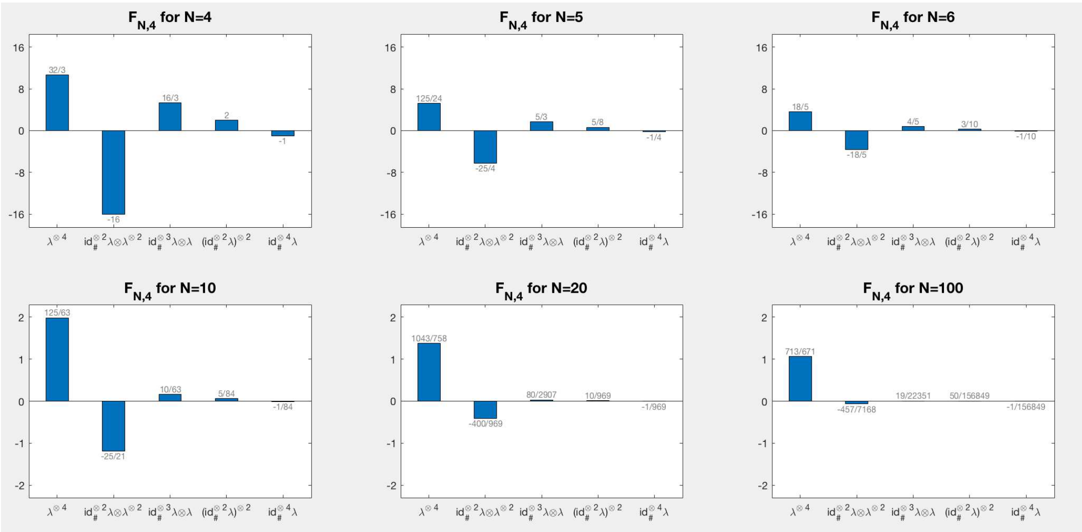

Figure 1: Coefficients of the universal polynomial $F_{N,4}$ for different N. For $N = 5$ and 6, the second (correlated) term is bigger respectively equal in absolute value to the first (independent) term; for $N = 20$ its size is about 30% that of the first term. For large N, $F_{N,4}$ converges to the independent measure $\lambda^{\otimes 4}$, but even for $N = 100$ the deviation from the latter is still visible.

Hewitt-Savage like representation

Let $\mathcal{P}_{\frac{1}{N}}(X)$ be the set of $\frac{1}{N}$-quantized probability measures on X i.e.

$$\mathcal{P}_{\frac{1}{N}}(X) = \{\frac{1}{N}\sum_{i=1}^{N}\delta_{x_i},\ (x_1,\ldots,x_N) \in X^N\} = \Lambda(X^N)$$

where Λ is the (Lipschitz for W_1) map: $X^N \to \mathcal{P}_{\frac{1}{N}}(X)$ defined by $\Lambda(x) = \frac{1}{N}\sum_{i=1}^{N}\delta_{x_i}$.

We have seen that if $\gamma = \delta_x$ then

$$\mu_N = S_N \gamma = F_{N,N}(\Lambda(x)) = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,N}(\lambda) \mathrm{d}\alpha(\lambda)$$

for

$$\alpha = \delta_{\Lambda(x)} = \Lambda_{\#}\gamma = \Lambda_{\#}\mu_N.$$

By linearity, for every $\mu_N \in \mathcal{P}_{\text{sym}}(X^N)$ i.e. $\mu_N = S_N \gamma$ with $\gamma \in \mathcal{P}(X^N)$ discrete, the same formula holds

$$\mu_N = S_N \gamma = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,N}(\lambda) \mathrm{d}\alpha(\lambda), \quad \text{with } \alpha = \Lambda_{\#}\mu_N$$

By density of discrete measures and continuity, the same holds for any symmetric measure.

We therefore have the Hewitt-Savage like representation: presentation result for N-representable k-plans.

Theorem 2 *Let $N \geq k \geq 2$. A measure $\mu_k \in \mathcal{P}(X^k)$ is N-representable if and only if there exists $\alpha \in \mathcal{P}(\mathcal{P}(X))$ such that $\alpha(\mathcal{P}_{\frac{1}{N}}(X)) = 1$ and*

$$\mu_k = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,k}(\lambda)\, d\alpha(\lambda) \tag{12}$$

where $F_{N,k}$ is defined by (6)–(8). Moreover, if $k = N$, the measure α in (12) is unique.

The case $k \leq N$ follows from the case $k = N$ since $\mu_k = M_k \mu_N$ and $F_{N,k} = M_k \circ F_{N,N}$ on $\mathcal{P}_{\frac{1}{N}}(X)$. Uniqueness for $k = N$ is by showing directly $\alpha = \Lambda_{\#} \mu_N$.

Remarks

- Letting $N \to \infty$, we recover the classical form of Hewitt and Savage, as well as the Diaconis-Freedman bound (because $F_{N,k}(\lambda) - \lambda^{\otimes k} = O_{\mathrm{TV}}(\frac{1}{N})$).
- Main differences: domain of integration consists of empirical measures only and the polynomials $F_{N,k}(\lambda)$ contain terms of degree less than k so as to account for correlated corrections.
- Natural sampling interpretation.

Corollary 1 *Let* $Z = (Z_1, \ldots, Z_N)$ *be a finitely exchangeable sequence of random variables with values in* X*, let* $\mu \in \mathcal{P}_{\mathrm{sym}}(X^N)$ *be the law of* Z *and let* $\alpha \in \mathcal{P}(\mathcal{P}(X))$ *be such that* $\alpha(\mathcal{P}_{\frac{1}{N}}(X)) = 1$ *and*

$$\mu = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,N}(\lambda)\, d\alpha(\lambda). \tag{13}$$

Let $(Z^{(\nu)})_{\nu \in \mathbf{N}}$ *be i.i.d drawn according to* μ*, and consider the* $\mathcal{P}(X)$*-valued sequence*

$$\Lambda(Z^{(\nu)}) := \frac{1}{N} \sum_{i=1}^{N} \delta_{Z_i^{(\nu)}}.$$

Then, almost surely, the empirical measure $\frac{1}{n} \sum_{\nu=1}^{n} \delta_{\Lambda(Z^{(\nu)})}$ *converges narrowly to* α *as* $n \to \infty$*.*

Extremal symmetric laws

Recall that an extreme point of a convex set C is a point of C which cannot be written as $ta + (1-t)b$ with $(t,a,b) \in (0,1) \times C^2$ unless $a = b$. An easy consequence of the Choquet representation in Theorem 2 is that the set of extreme points of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ is included in

$$\begin{aligned} E_{N,k} &:= \{F_{N,k}(\lambda) \,:\, \lambda \in \mathcal{P}_{\frac{1}{N}}(X)\} \\ &= \{M_k S_N \delta_{x_1,\ldots,x_N} \,:\, (x_1,\ldots,x_N) \in X^N\} \end{aligned} \tag{14}$$

The converse (and slightly more) is true:

Theorem 3 *Let $N \geq k \geq 2$.*

a) The set of extreme points of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ is given by the set $E_{N,k} = F_{N,k}(\mathcal{P}_{\frac{1}{N}}(X))$ defined in (14).

b) Every such extreme point is also exposed (that is for every $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$ one can find $\varphi \in C_b(X^k)$ such that $\mu \mapsto \int_{X^k} \varphi d\mu$ achieves its maximum on $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ at $F_{N,k}(\lambda)$ only).

In particular extremal N-representable k-plans form a narrowly closed set. Moreover, $\lambda \in \mathcal{P}_{\frac{1}{N}}(X) \mapsto F_{N,k}(\lambda)$ is a bijective parametrization ($M_1 F_{N,k}(\lambda) = \lambda$) of the extreme points of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$, note also that taking marginals (passing from $k+1$ to k) does not destroy extreme points.

Symmetric multi-marginal optimal transport

Given integers $2 \leq k \leq N$ and $\Phi \in C_b(X^k)$ *symmetric* (i.e., $\Phi(x_{\sigma(1)}, \dots, x_{\sigma(k)}) = \Phi(x_1, \dots, x_k)$ for every $(x_1, \dots, x_k) \in X^k$ and every permutation $\sigma \in \mathcal{S}_k$), we consider the (k-body interaction) symmetric cost c_Φ defined on X^N by

$$c_\Phi(x_1, \cdots, x_N) := \frac{1}{\binom{N}{k}} \sum_{1 \leq i_1 < i_2 \dots < i_k \leq N} \Phi(x_{i_1}, \dots, x_{i_k}). \tag{15}$$

Given $\rho \in \mathcal{P}(X)$ we are interested in the multi-marginal optimal transport problem

$$C_{N,k}(\rho) := \inf_{\gamma \in \mathcal{P}_{\mathrm{sym}}(X^N),\ M_1\gamma = \rho} \left\{ \int_{X^N} c_\Phi(x_1, \dots, x_N) \mathrm{d}\gamma(x_1, \dots, x_N) \right\}. \tag{16}$$

This can be reformulated in terms of the k-marginal $\mu_k = M_k(\gamma) \in \mathcal{P}_{N-\mathrm{rep}}(X^k)$:

$$C_{N,k}(\rho) := \inf \Big\{ \int_{X^k} \Phi \mathrm{d}\mu_k \;:\; \mu_k \in \mathcal{P}_{N-\mathrm{rep}}(X^k),\; M_1(\mu_k) = \rho \Big\}. \tag{17}$$

The de Finetti style representation from Theorem 2, together with the fact that $M_1(F_{N,k}(\lambda)) = \lambda$ for every $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$, enables us to write $C_{N,k}(\rho)$ as the infimum

$$\int_{\mathcal{P}_{\frac{1}{N}}(X)} \Big(\int_{X^k} \Phi \, \mathrm{d}F_{N,k}(\lambda) \Big) \mathrm{d}\alpha(\lambda)$$

with respect to $\alpha \in \mathcal{P}(\mathcal{P}_{\frac{1}{N}}(X))$ subject to the marginal constraint

$$\int_{\mathcal{P}_{\frac{1}{N}}(X)} \lambda \, \mathrm{d}\alpha(\lambda) = \rho. \tag{18}$$

One can observe that $\lambda \in \mathcal{P}_{\frac{1}{N}}(X) \mapsto \int_{X^k} \Phi \, \mathrm{d}F_{N,k}(\lambda)$ is a polynomial of degree k expression in the weights of the discrete measure λ, for instance if $k = 2$:

$$\int_{X^2} \Phi \, \mathrm{d}F_{N,2}(\lambda) = \frac{N}{N-1} \int_{X^2} \Phi(x,y) \mathrm{d}\lambda(x) \mathrm{d}\lambda(y) - \frac{1}{N-1} \int_X \Phi(x,x) \mathrm{d}\lambda(x)$$

and for $k = 3$:

$$\int_{X^3} \Phi \, \mathrm{d}F_{N,3}(\lambda) = \frac{N^2}{(N-1)(N-2)} \int_{X^3} \Phi(x,y,z) \mathrm{d}\lambda(x) \mathrm{d}\lambda(y) \mathrm{d}\lambda(z) - \frac{3N}{(N-1)(N-2)} \int_{X^2} \Phi(x,x,y) \mathrm{d}\lambda(x) \mathrm{d}\lambda(y) + \frac{2}{(N-1)(N-2)} \int_X \Phi(x,x,x) \mathrm{d}\lambda(x).$$

Defining the polynomial $P_{N,k}$ for every single marginal (not necessarily $\frac{1}{N}$-quantized) $\lambda \in \mathcal{P}(X)$ by

$$P_{N,k}(\lambda) := \int_{X^k} \Phi \, \mathrm{d}F_{N,k}(\lambda), \quad \text{for all } \lambda \in \mathcal{P}(X),$$

we see that the previous expression for $C_{N,k}(\rho)$ is a $\frac{1}{N}$-quantized constrained version of the convexification of the polynomial $P_{N,k}$:

$$P^{**}_{N,k}(\rho) := \inf_{\alpha \in \mathcal{P}(\mathcal{P}(X))} \left\{ \int_{\mathcal{P}(X)} P_{N,k}(\lambda) \mathrm{d}\alpha(\lambda) \ : \ \int_{\mathcal{P}(X)} \lambda \mathrm{d}\alpha(\lambda) = \rho \right\}. \tag{19}$$

In particular note that

$$C_{N,k}(\rho) \geq P^{**}_{N,k}(\rho), \ \forall \rho \in \mathcal{P}(X). \tag{20}$$

Of course

$$C_{N,k}(\rho) \leq P_k(\rho) := \int_{X^k} \Phi \mathrm{d}\rho^{\otimes k}, \ \forall \rho \in \mathcal{P}(X)$$

but since $C_{N,k}$ is convex this also gives

$$C_{N,k}(\rho) \leq P_k^{**}(\rho), \quad \text{for all } \rho \in \mathcal{P}(X). \tag{21}$$

Taking convex envelopes and using (20) we thus get, for every $\rho \in \mathcal{P}(X)$:

$$\frac{(N-k)!N^k}{N!}\Big(P_k^{**}(\rho) - \frac{C_k \|\Phi\|_\infty}{N}\Big) \leq C_{N,k}(\rho) \leq P_k^{**}(\rho). \tag{22}$$

In particular $C_{N,k}$ converges uniformly on $\mathcal{P}(X)$ to P_k^{**} as $N \to +\infty$ (but this already follows from the classical form of Hewitt-Savage).

Alternative convexification of polynomials viewpoint. Does not really break the curse of dimensionality (except if $X = \{0, 1\}$ where $C_{N,k}$ can be computed in linear in N time).