A Third Information-Theoretic Approach to Finite de Finetti Theorems

Mario Berta (Aachen), Lampros Gavalakis [1] (Gustave Eiffel),
Ioannis Kontoyiannis (Cambridge)

Athens, Greece | July 08, 2024

[1] L.G. has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 101034255.

Outline

De Finetti's Representation Theorem

$X_1, X_2, \ldots, X_n$ random variables in A

De Finetti's Representation Theorem

$X_1, X_2, \ldots, X_n$ random variables in A
Notation: $X_1^n = (X_1, \ldots, X_n)$

De Finetti's Representation Theorem

$X_1, X_2, \ldots, X_n$ random variables in A
Notation: $X_1^n = (X_1, \ldots, X_n)$

Exchangeability: $X_1, X_2, \ldots, X_n$ and $X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)}$ have the same distribution for every permutation π on $\{1, 2, \ldots, n\}$.

De Finetti's Representation Theorem

$X_1, X_2, \ldots, X_n$ random variables in A
Notation: $X_1^n = (X_1, \ldots, X_n)$

Exchangeability: $X_1, X_2, \ldots, X_n$ and $X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)}$ have the same distribution for every permutation π on $\{1, 2, \ldots, n\}$.

$\{X_i\}$ is *exchangeable*, if $X_1, X_2, \ldots, X_n$ are exchangeable for every $n \geq 1$

De Finetti's Representation Theorem

Theorem (de Finetti, 1930s)

$\{X_i\}$ is exchangeable if and only if it is a mixture of i.i.ds:

$$\mathbb{P}(X_1^k = x_1^k) = M_{k,\mu}(x_1^k) := \int_{\mathcal{P}} Q^k(x_1^k) d\mu(Q)$$

for some measure μ on the simplex $\mathcal{P}$

De Finetti's Representation Theorem

Theorem (de Finetti, 1930s)

$\{X_i\}$ is exchangeable if and only if it is a mixture of i.i.ds:

$$\mathbb{P}(X_1^k = x_1^k) = M_{k,\mu}(x_1^k) := \int_{\mathcal{P}} Q^k(x_1^k) d\mu(Q)$$

for some measure μ on the simplex $\mathcal{P}$

Fails if a *finite* collection of random variables $(X_1, X_2, \ldots, X_n)$ is exchangeable for some *fixed n* [Diaconis & Freedman, '80]:

De Finetti's Representation Theorem

Theorem (de Finetti, 1930s)

$\{X_i\}$ is exchangeable if and only if it is a mixture of i.i.ds:

$$\mathbb{P}(X_1^k = x_1^k) = M_{k,\mu}(x_1^k) := \int_{\mathcal{P}} Q^k(x_1^k) d\mu(Q)$$

for some measure μ on the simplex $\mathcal{P}$

Fails if a *finite* collection of random variables $(X_1, X_2, \ldots, X_n)$ is exchangeable for some *fixed n* [Diaconis & Freedman, '80]:

$$\mathbb{P}(X_1 = 1, X_2 = 0) = \mathbb{P}(X_1 = 0, X_2 = 1) = 1/2$$
$$\mathbb{P}(X_1 = 0, X_2 = 0) = \mathbb{P}(X_1 = 1, X_2 = 1) = 0$$

De Finetti's Representation Theorem

Theorem (de Finetti, 1930s)

$\{X_i\}$ is exchangeable if and only if it is a mixture of i.i.ds:

$$\mathbb{P}(X_1^k = x_1^k) = M_{k,\mu}(x_1^k) := \int_{\mathcal{P}} Q^k(x_1^k) d\mu(Q)$$

for some measure μ on the simplex $\mathcal{P}$

Fails if a *finite* collection of random variables $(X_1, X_2, \ldots, X_n)$ is exchangeable for some *fixed n* [Diaconis & Freedman, '80]:

$$\mathbb{P}(X_1 = 1, X_2 = 0) = \mathbb{P}(X_1 = 0, X_2 = 1) = 1/2$$
$$\mathbb{P}(X_1 = 0, X_2 = 0) = \mathbb{P}(X_1 = 1, X_2 = 1) = 0$$

However, $P_{X_1^k}$ is *close* to $M_{k,\mu}$

Diaconis and Freedman TV bounds

[Diaconis, Freedman '80]:

- $$\|P_k - M_{k,\mu_n}\|_{\mathrm{TV}} \leq \frac{2|A|k}{n}.$$
- $$\|P_k - M_{k,\mu_n}\|_{\mathrm{TV}} \leq \frac{k(k-1)}{2n}.$$

Diaconis and Freedman TV bounds

[Diaconis, Freedman '80]:

- $$\|P_k - M_{k,\mu_n}\|_{\mathrm{TV}} \leq \frac{2|A|k}{n}.$$
- $$\|P_k - M_{k,\mu_n}\|_{\mathrm{TV}} \leq \frac{k(k-1)}{2n}.$$

- Condition on the type $\Rightarrow \mu_n =$ law of the empirical measure

Information-Theoretic Approaches: Previous proofs

Closeness will be measured in terms of relative entropy

$$D(P||Q) = \sum_{x \in A} P(x) \log \frac{P(x)}{Q(x)}$$

Information-Theoretic Approaches: Previous proofs

Closeness will be measured in terms of relative entropy

$$D(P||Q) = \sum_{x \in A} P(x) \log \frac{P(x)}{Q(x)}$$

- Stronger by Pinsker's inequality

$$\|P - Q\|_{\mathrm{TV}} \leq \Big(2D(P\|Q)\Big)^{\frac{1}{2}}.$$

Information-Theoretic Approaches

- [G., Kontoyiannis, 2021] Binary RVs

$$D(P_k \| M_{k,\mu_n}) \leq \frac{5k^2 \log n}{n-k}$$

Information-Theoretic Approaches

- [G., Kontoyiannis, 2021] Binary RVs

$$D(P_k \| M_{k,\mu_n}) \leq \frac{5k^2 \log n}{n-k}$$

- [G., Kontoyiannis, 2022] RVs on finite alphabets

$$D(P_k \| M_{k,\mu_n}) = O\left(\left(\frac{k}{\sqrt{n}}\right)^{1/2} \log \frac{n}{k}\right)$$

Information-Theoretic Approaches

- [G., Kontoyiannis, 2021] Binary RVs

$$D(P_k \| M_{k,\mu_n}) \leq \frac{5k^2 \log n}{n-k}$$

- [G., Kontoyiannis, 2022] RVs on finite alphabets

$$D(P_k \| M_{k,\mu_n}) = O\left(\left(\frac{k}{\sqrt{n}}\right)^{1/2} \log \frac{n}{k}\right)$$

- Diaconis & Freedman: $O(\frac{k}{n})$ for the total variation distance

Information-Theoretic Approaches: New Approach

Theorem (Exchangeability and information)

Suppose X_1^n *is an exchangeable with values in a standard Borel space* $(A, \mathcal{A})$.
For every $1 \leq k \leq n-1$ *there exists a probability measure* $\mu = \mu_{k,n}$ *on the space of probability measures on* $(A, \mathcal{A})$ *s.t.*

$$D\big(P_{X_1^k} \big\| M_{k,\mu}\big) \leq \frac{1}{n-k+1} \sum_{i=1}^{k} I(X_1^{i-1}; X_k^n),$$

Information-Theoretic Approaches: New Approach

Theorem (Exchangeability and information)

Suppose X_1^n is an exchangeable with values in a standard Borel space $(A, \mathcal{A})$.
For every $1 \leq k \leq n-1$ there exists a probability measure $\mu = \mu_{k,n}$ on the space of probability measures on $(A, \mathcal{A})$ s.t.

$$D\big(P_{X_1^k} \big\| M_{k,\mu}\big) \leq \frac{1}{n-k+1} \sum_{i=1}^{k} I(X_1^{i-1}; X_k^n),$$

- For A discrete (finite or countably infinite)

$$I(X_1^k; Y) = H(X_1^k) - H(X_1^k|Y) \leq H(X_1^k) \leq kH(X_1) \leq k \log |A|$$

Explicit de Finetti-style theorems

Therefore

Corollary

If X_1^n exchangeable, discrete, then for every $1 \leq k \leq n-1$ there is $\mu = \mu_{k,n}$ s.t.

$$D\left(P_{X_1^k} \middle\| M_{k,\mu}\right) \leq \frac{k(k-1)}{2(n-k+1)} H(X_1)$$
$$\leq \frac{k(k-1)}{2(n-k+1)} \log |A|$$

Explicit de Finetti-style theorems

Therefore

Corollary

If X_1^n exchangeable, discrete, then for every $1 \leq k \leq n-1$ there is $\mu = \mu_{k,n}$ s.t.

$$D\left(P_{X_1^k} \middle\| M_{k,\mu}\right) \leq \frac{k(k-1)}{2(n-k+1)} H(X_1)$$
$$\leq \frac{k(k-1)}{2(n-k+1)} \log |A|$$

- more general

Explicit de Finetti-style theorems

Therefore

Corollary

If X_1^n exchangeable, discrete, then for every $1 \leq k \leq n-1$ there is $\mu = \mu_{k,n}$ s.t.

$$D\left(P_{X_1^k} \middle\| M_{k,\mu}\right) \leq \frac{k(k-1)}{2(n-k+1)} H(X_1)$$
$$\leq \frac{k(k-1)}{2(n-k+1)} \log |A|$$

- more general
- stronger

Explicit de Finetti-style theorems

Therefore

Corollary

If X_1^n exchangeable, discrete, then for every $1 \leq k \leq n-1$ there is $\mu = \mu_{k,n}$ s.t.

$$D\left(P_{X_1^k} \middle\| M_{k,\mu}\right) \leq \frac{k(k-1)}{2(n-k+1)} H(X_1)$$
$$\leq \frac{k(k-1)}{2(n-k+1)} \log |A|$$

- more general
- stronger

Explicit de Finetti-style theorems

Corollary (Classical de Finetti theorem for compact spaces)

Suppose $\{X_k\}$ *is exchangeable with values in a compact metrisable space* G *equipped with its Borel* σ*-algebra. If for every* k *we have*

$$I(X_1^{k-1}; X_k^n) = o(n) \quad \text{as } n \to \infty,$$

then there is μ *s.t.*

$$P_{X_1^k} = M_{k,\mu}$$

Explicit de Finetti-style theorems

Corollary (Classical de Finetti theorem for compact spaces)

Suppose $\{X_k\}$ *is exchangeable with values in a compact metrisable space* G *equipped with its Borel* σ*-algebra. If for every* k *we have*

$$I(X_1^{k-1}; X_k^n) = o(n) \quad \text{as } n \to \infty,$$

then there is μ *s.t.*

$$P_{X_1^k} = M_{k,\mu}$$

- Applies to finite-alphabet

Explicit de Finetti-style theorems

Corollary (Classical de Finetti theorem for compact spaces)

Suppose $\{X_k\}$ is exchangeable with values in a compact metrisable space G equipped with its Borel σ-algebra. If for every k we have

$$I(X_1^{k-1}; X_k^n) = o(n) \quad \text{as } n \to \infty,$$

then there is μ s.t.

$$P_{X_1^k} = M_{k,\mu}$$

- Applies to finite-alphabet
- Condition might fail

Proof Sketch

$$\sum_{m=k}^{n} I(X_1^{i-1}; X_i | X_{k+1}^{m}) = \sum_{m=k}^{n} I(X_1^{i-1}; X_m | X_k^{m-1})$$
$$= I(X_1^{i-1}; X_k^{n}),$$

Proof Sketch

$$\sum_{m=k}^{n} I(X_1^{i-1}; X_i | X_{k+1}^{m}) = \sum_{m=k}^{n} I(X_1^{i-1}; X_m | X_k^{m-1})$$
$$= I(X_1^{i-1}; X_k^{n}),$$

Proof Sketch

$$\frac{1}{n-k+1}\sum_{i=1}^{k}\sum_{m=k}^{n} I(X_1^{i-1}; X_i|X_{k+1}^{m}) = \frac{1}{n-k+1}\sum_{i=1}^{k}\sum_{m=k}^{n} I(X_1^{i-1}; X_m|X_k^{m-1})$$

$$= \frac{1}{n-k+1}\sum_{i=1}^{k} I(X_1^{i-1}; X_k^{n})$$

Proof Sketch

$$\frac{1}{n-k+1}\sum_{m=k}^{n}\sum_{i=1}^{k} I(X_1^{i-1}; X_i | X_{k+1}^{m}) = \frac{1}{n-k+1}\sum_{i=1}^{k} I(X_1^{i-1}; X_k^{n})$$

Proof Sketch

$$\frac{1}{n-k+1}\sum_{m=k}^{n}\sum_{i=1}^{k} I(X_1^{i-1}; X_i | X_{k+1}^{m}) = \frac{1}{n-k+1}\sum_{i=1}^{k} I(X_1^{i-1}; X_k^{n})$$

$\Rightarrow \exists\ m^* \in \{k, k+1, \ldots, n\}$ such that

$$\sum_{i=1}^{k} I(X_1^{i-1}; X_i | X_{k+1}^{m^*}) \leq \frac{1}{n-k+1}\sum_{i=1}^{k} I(X_1^{i-1}; X_k^{n})$$

Proof sketch

But now with $\mu_* =$ law of (random) $P_{X_1^k|X_{k+1}^{m^*}}(\cdot|x_{k+1}^{m^*})$

$$
\begin{aligned}
D\big(P_{X_1^k}\big\|M_{k,\mu_*}\big) &= D\Big(\int P_{X_1^k|X_{k+1}^{m^*}}(\cdot|x_{k+1}^{m^*})d\mu_*\Big\|\int P^k_{X_1|X_{k+1}^{m^*}}(\cdot|x_{k+1}^{m^*})d\mu_*\Big)\\
&\leq \int D\Big(P_{X_1^k|X_{k+1}^{m^*}}(\cdot|x_{k+1}^{m^*})\Big\|P^k_{X_1|X_{k+1}^{m^*}}(\cdot|x_{k+1}^{m^*})\Big)\ d\mu_*\\
&= \int D\Big(P_{X_1^k|X_{k+1}^{m^*}}(\cdot|x_{k+1}^{m^*})\Big\|\prod_{i=1}^k P_{X_i|X_{k+1}^{m^*}}(\cdot|x_{k+1}^{m^*})\Big)\ d\mu_*\\
&= \sum_{i=1}^k I(X_1^{i-1};X_i|X_{k+1}^{m^*}) \leq \frac{1}{n-k+1}\sum_{i=1}^k I(X_1^{i-1};X_k^n)
\end{aligned}
$$

□

On the optimal rate

Sharp de Finetti upper bound in [Johnson, G., Kontoyiannis '24] via sampling bounds:

Theorem

If X_1^n exchangeable on a finite alphabet A, there is $\mu = \mu_n$ s.t.

$$D\big(P_{X_1^k} \| M_{k,\mu}\big) \leq \frac{1}{2}(|A|-1)\frac{k(k-1)}{(n-1)(n-k+1)}$$

Open questions

- de Finetti's theorem is often used in machine learning!
  - Applications?
- Quantum Information Theory
- Extensions
  - Weighted exchangeability, Aldous–Hoover,...

Thank you!