A de Finetti theorem for infinite weighted exchangeable sequences

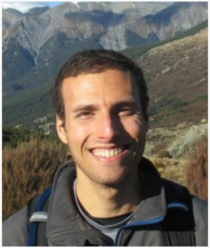
Rina Foygel Barber (with Emmanuel Candès, Aaditya Ramdas, Ryan Tibshirani)

http://rinafb.github.io/

Collaborators

Emmanuel Candès

Aaditya Ramdas



Ryan Tibshirani

- Thanks to American Institute of Math (AIM) for hosting & supporting our collaboration as an AIM SQuaRE
- Paper: *De Finetti's theorem and related results for infinite weighted exchangeable sequences*, Bernoulli 2024

Outline

Background

Exchangeable sequences

An infinite exchangeable sequence:

0 0 1 0 1 1 0 0 0 0 1 0 1...

De Finetti's theorem[1]

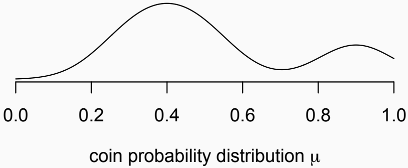
Any infinite exchangeable sequence $X_1, X_2, \ldots \in \{0, 1\}$ can be expressed as a mixture of i.i.d. sequences

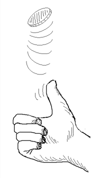
[1] de Finetti 1929, *Funzione caratteristica di un fenomeno aleatorio*

Exchangeable sequences

An infinite exchangeable sequence:

0 0 1 0 1 1 0 0 0 0 1 0 1...





De Finetti's theorem[1]

Any infinite exchangeable sequence $X_1, X_2, \ldots \in \{0, 1\}$ can be expressed as a mixture of i.i.d. sequences

[1] de Finetti 1929, *Funzione caratteristica di un fenomeno aleatorio*

Exchangeable sequences

Defining exchangeability

A sequence of random variables $X_1, \ldots, X_n \in \mathcal{X}$ is exchangeable if

$$(X_1, \ldots, X_n) \stackrel{\mathrm{d}}{=} (X_{\sigma(1)}, \ldots, X_{\sigma(n)})$$

for all permutations $\sigma \in \mathcal{S}_n$.

An infinite sequence $X_1, X_2, \ldots \in \mathcal{X}$ is exchangeable if every finite subsequence is exchangeable.

Exchangeable sequences

De Finetti–Hewitt–Savage theorem[2]

Any infinite exchangeable sequence $X_1, X_2, \ldots \in \mathcal{X}$ can be expressed as a mixture of i.i.d. sequences, under regularity conditions (e.g., $\mathcal{X}$ is a standard Borel space)

---

[2] Hewitt & Savage 1955, *Symmetric measures on Cartesian products*
Many extensions, e.g., Farrell 1962, Maitra 1977, Alam 2020, Fritz et al 2021

Exchangeability & distribution-free inference

Statistical problem—prediction in the supervised learning setting:

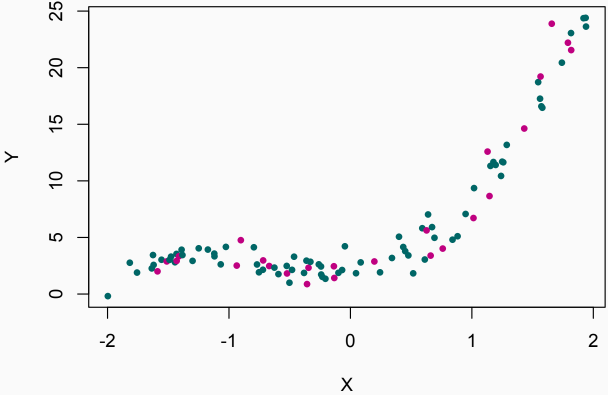
- Training data $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n) \in \mathcal{X} \times \mathcal{Y}$
- Prediction: predict value of Y given X
  for test points $(X_{n+1}, Y_{n+1}), (X_{n+2}, Y_{n+2}), \ldots$
- Predictive inference: quantify uncertainty for these predictions—construct a prediction set $\mathcal{C}$ such that

$$\mathbb{P}\left\{Y_{n+i} \in \mathcal{C}(X_{n+i})\right\} \geq 90\%$$

Exchangeability & distribution-free inference

Incorrect model assumptions $\rightsquigarrow$ inaccurate predictions

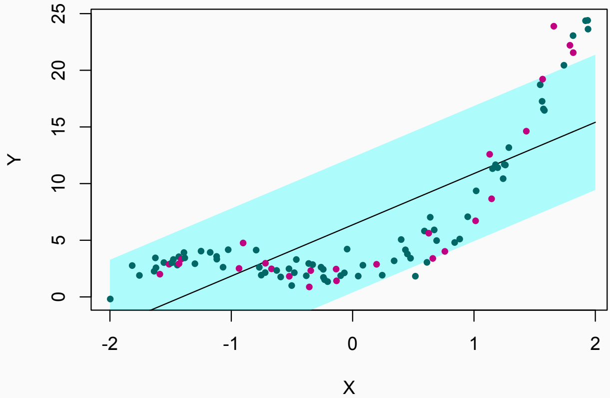
But under exchangeability of the data $\rightsquigarrow$ still get valid predictive intervals



Exchangeability & distribution-free inference

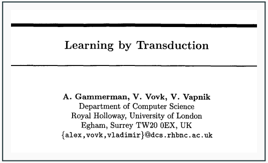
Incorrect model assumptions $\rightsquigarrow$ inaccurate predictions

But under exchangeability of the data $\rightsquigarrow$ still get valid predictive intervals



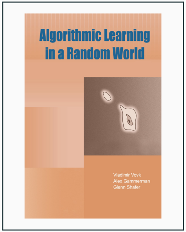
Exchangeability & distribution-free inference

The conformal prediction (CP) framework:
distribution-free predictive inference via exchangeability of the data



Learning by Transduction

A. Gammerman, V. Vovk, V. Vapnik
Department of Computer Science
Royal Holloway, University of London
Egham, Surrey TW20 0EX, UK
{alex,vovk,vladimir}@dcs.rhbnc.ac.uk



JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2018, VOL. 113, NO. 523, 1094–1111, Theory and Methods
https://doi.org/10.1080/01621459.2017.1307116

Taylor & Francis
Taylor & Francis Group

Check for updates

Distribution-Free Predictive Inference for Regression

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman

Department of Statistics, Carnegie Mellon University, Pittsburgh, PA

ABSTRACT
We develop a general framework for distribution-free predictive inference in regression, using conformal inference. The proposed methodology allows for the construction of a prediction band for the response variable using any estimator of the regression function. The resulting prediction band preserves the consistency properties of the original estimator under standard assumptions, while guaranteeing finite-sample marginal coverage even when these assumptions do not hold. We analyze and compare, both empirically and theoretically, the two major variants of our conformal framework: full conformal inference and split conformal inference, along with a related jackknife method. These methods offer different tradeoffs between statistical accuracy (length of resulting prediction intervals) and computational efficiency. As extensions, we develop a method for constructing valid in-sample prediction intervals called *rank-one-out* conformal inference, which has essentially the same computational efficiency as split conformal inference. We also describe an extension of our procedures for producing prediction bands with locally varying length, to adapt to heteroscedasticity in the data. Finally, we propose a model-free notion of variable importance, called *leave-one-covariate-out* or LOCO inference. Accompanying this article is an R package `conformalInference` that implements all of the proposals we have introduced. In the spirit of reproducibility, all of our empirical results can also be easily (re)generated using this package.

ARTICLE HISTORY
Received April 2016
Revised February 2017

KEYWORDS
Distribution-free; Model misspecification; Prediction band; Regression; Variable importance

Gammerman, Vovk, Vapnik
UAI 1998

Vovk, Gammerman, Shafer
2005 — see `alrw.net`

Lei, G'Sell, Rinaldo,
Tibshirani, Wasserman
JASA 2018

Some notation

The exchangeable sigma algebra

$\mathcal{E}_\infty$ is the set of all events $A \subseteq \mathcal{X}^\infty$ for which

$$(x_1, x_2, \dots) \in A \Longleftrightarrow (x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}, x_{n+1}, \dots) \in A$$

holds for all x, all $n \geq 1$, all $\sigma \in \mathcal{S}_n$.

Examples:

- The event that $\sup_{i \geq 1} x_i \leq C$
- The event that the sequence $(x_1, x_2, \dots)$ contains exactly k zeros

Additional properties of i.i.d. sequences

The Hewitt–Savage zero-one law[3]

If $X_1, X_2, \ldots \overset{\text{iid}}{\sim} P$ for any P, then

$$\mathbb{P}\left\{(X_1, X_2, \ldots) \in A\right\} \in \{0, 1\}$$

for all $A \in \mathcal{E}_\infty$.

[3] Hewitt & Savage 1955, *Symmetric measures on Cartesian products*

Additional properties of i.i.d. sequences

The Hewitt–Savage zero-one law[3]

If $X_1, X_2, \ldots \overset{\text{iid}}{\sim} P$ for any P, then

$$\mathbb{P}\{(X_1, X_2, \ldots) \in A\} \in \{0, 1\}$$

for all $A \in \mathcal{E}_\infty$.

In particular, the zero-one law implies the law of large numbers (LLN):

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \in A\}}_{=\widehat{P}_n(A)} \overset{\text{a.s.}}{\to} P(A)$$

for all events $A \subseteq \mathcal{X}$.

[3] Hewitt & Savage 1955, *Symmetric measures on Cartesian products*

Additional properties of i.i.d. sequences

Relating back to de Finetti's theorem:

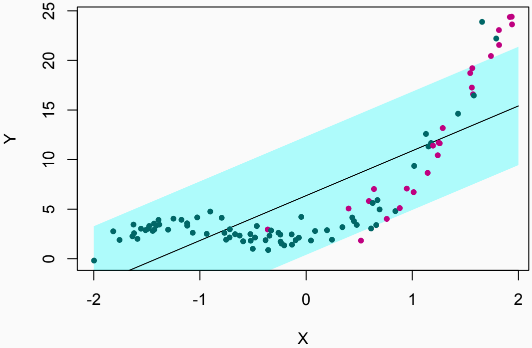
- Suppose $X_1, X_2, \ldots$ is exchangeable

  $\rightsquigarrow X_i \overset{\text{iid}}{\sim} P$ where P is random (drawn from a distribution μ)
- Conditional on P, for all A it holds that $\widehat{P}_n(A) \overset{\text{a.s.}}{\to} P(A)$ (by LLN)
- Therefore, the random distribution P can be recovered as the limit of the empirical distributions

Weighted exchangeability

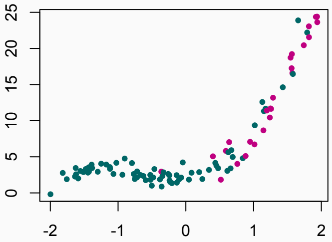
Motivation: prediction under distribution drift

- Training data $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$
- Predict for test data $(X_{n+1}, Y_{n+1}), (X_{n+2}, Y_{n+2}), \ldots$
- But the training & test data is not i.i.d. due to *distribution shift*

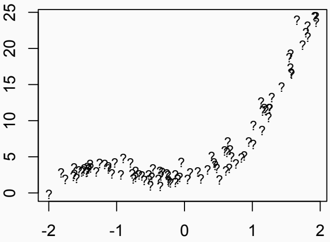


Motivation: prediction under distribution drift

If we know (or can estimate) the distribution shift,
can correct for it & obtain (approximately) valid predictive coverage
via *weighted conformal prediction*[4]



$\rightsquigarrow$



[4] Tibshirani, B., Candès, & Ramdas 2019, *Conformal Prediction Under Covariate Shift*
Podkopaev & Ramdas 2021, *Distribution-free uncertainty quantification for classification under label shift*
Prinster et al 2024, *Conformal Validity Guarantees Exist for Any Data Distribution*

Defining weighted exchangeability

Let $\Lambda = \{\text{measurable functions } \mathcal{X} \to (0, \infty)\}$.
Fix a sequence $\lambda = (\lambda_1, \ldots, \lambda_n) \in \Lambda^n$ (the "weight functions")

Finite weighted exchangeability

A distribution Q on $(X_1, \ldots, X_n) \in \mathcal{X}^n$ is λ-weighted exchangeable if

$$\frac{\mathrm{d}Q(x_1, \ldots, x_n)}{\lambda_1(x_1) \cdot \ldots \cdot \lambda_n(x_n)}$$

is an exchangeable measure.

Defining weighted exchangeability

Let $\Lambda = \{\text{measurable functions } \mathcal{X} \to (0, \infty)\}$.
Fix a sequence $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda^n$ (the "weight functions")

Finite weighted exchangeability

A distribution Q on $(X_1, \dots, X_n) \in \mathcal{X}^n$ is λ-weighted exchangeable if

$$\frac{\mathrm{d}Q(x_1, \dots, x_n)}{\lambda_1(x_1) \cdot \ldots \cdot \lambda_n(x_n)}$$

is an exchangeable measure.

For the distribution shift setting (weighted conformal prediction):

- Training data $\overset{\text{iid}}{\sim} P_0$, test data $\overset{\text{iid}}{\sim} P_1$, with $\frac{\mathrm{d}P_1}{\mathrm{d}P_0}$ known (or estimated)
- Then define $\lambda_i \equiv 1$ for training points, and $\lambda_i = \frac{\mathrm{d}P_1}{\mathrm{d}P_0}$ for test points

Defining weighted exchangeability

Weight functions $\lambda = (\lambda_1, \lambda_2, \dots) \in \Lambda^\infty$

Infinite weighted exchangeability

A distribution Q on $(X_1, X_2, \ldots) \in \mathcal{X}^\infty$ is λ-weighted exchangeable if, for every $n \geq 1$, the marginal Q_n satisfies that

$$\frac{\mathrm{d}Q_n(x_1, \dots, x_n)}{\lambda_1(x_1) \cdot \ldots \cdot \lambda_n(x_n)}$$

is an exchangeable measure.

Weighted exchangeable & weighted i.i.d.

Weighted i.i.d. distribution

Given a measure P on $\mathcal{X}$ & weight functions $\lambda = (\lambda_1, \lambda_2, \dots) \in \Lambda^\infty$,

$$P \circ \lambda = (P \circ \lambda_1) \times (P \circ \lambda_2) \times \dots$$

where each component is defined as

$$\frac{\mathrm{d}(P \circ \lambda_i)}{\mathrm{d}P}(x) \propto \lambda_i(x)$$

Weighted exchangeable & weighted i.i.d.

Weighted i.i.d. distribution

Given a measure P on $\mathcal{X}$ & weight functions $\lambda = (\lambda_1, \lambda_2, \dots) \in \Lambda^\infty$,

$$P \circ \lambda = (P \circ \lambda_1) \times (P \circ \lambda_2) \times \dots$$

where each component is defined as

$$\frac{\mathrm{d}(P \circ \lambda_i)}{\mathrm{d}P}(x) \propto \lambda_i(x)$$

well-defined if $0 < \int_{\mathcal{X}} \lambda_i(x)\, \mathrm{d}P(x) < \infty$ for all i
(denoted by $P \in \mathcal{M}_{\mathcal{X}}(\lambda)$)

Weighted exchangeable & weighted i.i.d.

What is $P \circ \lambda$ in the binary case, $\mathcal{X} = \{0, 1\}$?

- The distribution P is Bernoulli(p) for some $p \in [0, 1]$
- Each weight function is $\lambda_i = (\lambda_i(0), \lambda_i(1))$

$$\rightsquigarrow \quad P \circ \lambda = (P \circ \lambda_1) \times \cdots \times \underbrace{(P \circ \lambda_i)}_{=\text{Bernoulli}\left(\frac{p\lambda_i(1)}{p\lambda_i(1) + (1-p)\lambda_i(0)}\right)} \times \ldots$$

Weighted exchangeable & weighted i.i.d.

Any mixture of λ-weighted i.i.d. distributions, is λ-weighted exchangeable.

Notation

$Q = (P \circ \lambda)_\mu$ denotes that a draw from Q is generated as follows:

- Draw $P \sim \mu$ (where μ is a distribution on $\mathcal{M}_\mathcal{X}(\lambda)$)
- Draw $(X_1, X_2, \ldots) \sim P \circ \lambda$

Weighted de Finetti?

A weighted version of de Finetti's theorem?

Is it true that any λ-weighted exchangeable distribution Q on $\mathcal{X}^\infty$ can be expressed as a mixture of λ-weighted i.i.d. distributions:

$$Q \text{ is } \lambda\text{-weighted exch.} \implies Q = (P \circ \lambda)_\mu \text{ for some } \mu \text{ ?}$$

Weighted de Finetti?

A weighted version of de Finetti's theorem?

Is it true that any λ-weighted exchangeable distribution Q on $\mathcal{X}^\infty$ can be expressed as a mixture of λ-weighted i.i.d. distributions:

$$Q \text{ is } \lambda\text{-weighted exch.} \implies Q = (P \circ \lambda)_\mu \text{ for some } \mu \text{ ?}$$

FALSE

Consider the binary case, $\mathcal{X} = \{0, 1\}$, and

$$\lambda_i = (\lambda_i(0), \lambda_i(1)) = (1, 2^{-i}).$$

Define a λ-weighted exchangeable Q as:

$$\underbrace{(1, 0, 0, 0, 0, \ldots)}_{\text{prob. } \frac{1}{2}} \quad \underbrace{(0, 1, 0, 0, 0, \ldots)}_{\text{prob. } \frac{1}{4}} \quad \underbrace{(0, 0, 1, 0, 0, \ldots)}_{\text{prob. } \frac{1}{8}} \quad \ldots$$

Weighted de Finetti?

Theorem: binary case (Lauritzen)[5]

Let $\lambda_i = (\lambda_i(0), \lambda_i(1)) \in \mathbb{R}_+^2$ be a sequence of weight functions.

Then the statement

> *Any λ-weighted exchangeable distribution Q on $\{0,1\}^\infty$*
> *can be expressed as a mixture of λ-weighted i.i.d. distributions*

holds if and only if

$$\sum_{i \geq 1} \frac{\min\{\lambda_i(0), \lambda_i(1)\}}{\max\{\lambda_i(0), \lambda_i(1)\}} = \infty.$$

[5] Lauritzen 1988, *Extremal families and systems of sufficient statistics*

Weighted de Finetti?

Theorem: binary case (Lauritzen)[5]

Let $\lambda_i = (\lambda_i(0), \lambda_i(1)) \in \mathbb{R}_+^2$ be a sequence of weight functions.

Then the statement

> *Any λ-weighted exchangeable distribution Q on $\{0,1\}^\infty$ can be expressed as a mixture of λ-weighted i.i.d. distributions*

holds if and only if

$$\sum_{i\geq 1} \frac{\min\{\lambda_i(0), \lambda_i(1)\}}{\max\{\lambda_i(0), \lambda_i(1)\}} = \infty.$$

Question posed by Vladimir Vovk in $\sim$2020:
Does this result generalize beyond the binary case?

[5] Lauritzen 1988, *Extremal families and systems of sufficient statistics*

Main results: three properties

Three sets

For a general space $\mathcal{X}$ (we will assume standard Borel space)....

Question 1: which $\lambda \in \Lambda^\infty$ lead to a de Finetti theorem?

$$\Lambda_{\mathsf{deFin}} = \left\{ \begin{array}{c} \text{All sequences } \lambda \in \Lambda^\infty \text{ for which it holds that} \\ \text{any } \lambda\text{-weighted exchangeable distribution } Q \\ \text{can be expressed as } Q = (P \circ \lambda)_\mu \text{ for some } \mu \end{array} \right\}$$

Three sets

Question 2: which $\lambda \in \Lambda^\infty$ lead to a zero-one law?

$$\Lambda_{01} = \left\{ \begin{array}{c} \text{All sequences } \lambda \in \Lambda^\infty \text{ for which it holds that} \\ (P \circ \lambda)(A) \in \{0,1\} \text{ for all } P \in \mathcal{M}_{\mathcal{X}}(\lambda) \text{ and all } A \in \mathcal{E}_\infty \end{array} \right\}$$

Three sets

Question 3: which $\lambda \in \Lambda^\infty$ lead to a Law of Large Numbers?

$$\Lambda_{\mathsf{LLN}} = \left\{ \begin{array}{c} \text{All sequences } \lambda \in \Lambda^\infty \text{ for which it holds that} \\ \text{if } (X_1, X_2, \ldots) \sim P \circ \lambda \text{ for some } P \in \mathcal{M}_{\mathcal{X}}(\lambda), \\ \text{then } \underbrace{\widetilde{P}_{n,i}(A)}_{\text{a weighted version of the empirical distrib.}} \overset{\text{a.s.}}{\to} (P \circ \lambda_i)(A) \text{ for all } i \geq 1 \text{ and all events } A \end{array} \right\}$$

Main results

Theorem: nested sets

For any standard Borel space $\mathcal{X}$ it holds that

$$\Lambda_{\mathsf{deFin}} \subseteq \Lambda_{01} \subseteq \Lambda_{\mathsf{LLN}}.$$

Proof

Proof idea for $\lambda \in \Lambda_{01} \implies \lambda \in \Lambda_{\mathsf{LLN}}$

For the weighted i.i.d. sequence $(X_1, X_2, \ldots) \sim P \circ \lambda$, want to show

$$\widetilde{P}_{n,i}(A) \overset{\text{a.s.}}{\to} (P \circ \lambda_i)(A)$$

where

$$\underbrace{\widetilde{P}_{n,i} = \sum_{j=1}^{n} \frac{\sum_{\sigma \in \mathcal{S}_n : \sigma(i)=j} \prod_{k=1}^{n} \lambda_k(x_{\sigma(k)})}{\sum_{\sigma \in \mathcal{S}_n} \prod_{k=1}^{n} \lambda_k(x_{\sigma(k)})} \cdot \delta_{X_j}}_{= \text{ distribution of } X_i \text{ conditional on } \widehat{P}_n}$$

Proof

Proof idea for $\lambda \in \Lambda_{01} \implies \lambda \in \Lambda_{\mathsf{LLN}}$

For the weighted i.i.d. sequence $(X_1, X_2, \ldots) \sim P \circ \lambda$, want to show

$$\widetilde{P}_{n,i}(A) \overset{\text{a.s.}}{\to} (P \circ \lambda_i)(A)$$

where

$$\underbrace{\widetilde{P}_{n,i} = \sum_{j=1}^{n} \frac{\sum_{\sigma \in \mathcal{S}_n : \sigma(i)=j} \prod_{k=1}^{n} \lambda_k(x_{\sigma(k)})}{\sum_{\sigma \in \mathcal{S}_n} \prod_{k=1}^{n} \lambda_k(x_{\sigma(k)})} \cdot \delta_{X_j}}_{= \text{ distribution of } X_i \text{ conditional on } \widehat{P}_n}$$

- Martingale theorems $\implies \widetilde{P}_{n,i}(A) \overset{\text{a.s.}}{\to} \mathbb{P}\{X_i \in A \mid \mathcal{E}_\infty\}$
- $\lambda \in \Lambda_{01} \implies \mathbb{P}\{X_i \in A \mid \mathcal{E}_\infty\} \overset{\text{a.s.}}{=} \mathbb{E}[\mathbb{P}\{X_i \in A \mid \mathcal{E}_\infty\}] = (P \circ \lambda_i)(A)$

Proof

Proof idea for $\lambda \in \Lambda_{\mathsf{deFin}} \implies \lambda \in \Lambda_{01}$

Let $A \in \mathcal{E}_\infty$, $P_* \in \mathcal{M}_{\mathcal{X}}(\lambda)$, $p = (P_* \circ \lambda)(A)$. Then

$$P_* \circ \lambda = p \cdot \underset{\substack{\uparrow \\ \text{condition on } A}}{Q_0} + (1-p) \cdot \underset{\substack{\uparrow \\ \text{condition on } A^c}}{Q_1}$$

Proof

Proof idea for $\lambda \in \Lambda_{\mathsf{deFin}} \implies \lambda \in \Lambda_{01}$

Let $A \in \mathcal{E}_\infty$, $P_* \in \mathcal{M}_{\mathcal{X}}(\lambda)$, $p = (P_* \circ \lambda)(A)$. Then

$$P_* \circ \lambda = p \cdot \underset{\substack{\uparrow \\ \text{condition on } A}}{Q_0} + (1-p) \cdot \underset{\substack{\uparrow \\ \text{condition on } A^c}}{Q_1}$$

- If $0 < p < 1$, then Q_0, Q_1 are λ-weighted exchangeable
- $\lambda \in \Lambda_{\mathsf{deFin}} \implies Q_0 = (P \circ \lambda)_{\mu_0}$, $Q_1 = (P \circ \lambda)_{\mu_1}$
- Therefore, $P_* \circ \lambda = (P \circ \lambda)_\mu$ where $\mu = p \cdot \mu_0 + (1-p) \cdot \mu_1$
- But also, $P_* \circ \lambda = (P \circ \lambda)_\mu$ where $\mu = \delta_{P_*} \rightsquigarrow$ contradiction

Main results: the binary case

Theorem: the binary case (extending Lauritzen's result)

For $\mathcal{X} = \{0,1\}$ it holds that

$$\Lambda_{\mathsf{deFin}} = \Lambda_{\mathsf{01}} = \Lambda_{\mathsf{LLN}} = \left\{ \lambda : \sum_{i \geq 1} \frac{\min\{\lambda_i(0), \lambda_i(1)\}}{\max\{\lambda_i(0), \lambda_i(1)\}} = \infty \right\}.$$

Main results: the binary case

Intuition for why Lauritzen's condition is necessary

Counterexample:

$$\lambda_1 = (1, \tfrac{1}{2}),\ \lambda_2 = (1, \tfrac{1}{4}),\ \lambda_3 = (1, \tfrac{1}{8}),\ \lambda_4 = (1, \tfrac{1}{16}),\ \dots$$

- $\lambda \notin \Lambda_{\mathsf{deFin}}$: define λ-weighted exchangeable Q as:

$$\underbrace{(1,0,0,0,0,\dots)}_{\text{prob. } \frac{1}{2}} \quad \underbrace{(0,1,0,0,0,\dots)}_{\text{prob. } \frac{1}{4}} \quad \underbrace{(0,0,1,0,0,\dots)}_{\text{prob. } \frac{1}{8}} \quad \dots$$

Main results: the binary case

Intuition for why Lauritzen's condition is necessary

Counterexample:

$$\lambda_1 = (1, \tfrac{1}{2}),\ \lambda_2 = (1, \tfrac{1}{4}),\ \lambda_3 = (1, \tfrac{1}{8}),\ \lambda_4 = (1, \tfrac{1}{16}),\ \ldots$$

- $\lambda \notin \Lambda_{\mathrm{deFin}}$: define λ-weighted exchangeable Q as:

$$\underbrace{(1, 0, 0, 0, 0, \ldots)}_{\text{prob. } \frac{1}{2}} \quad \underbrace{(0, 1, 0, 0, 0, \ldots)}_{\text{prob. } \frac{1}{4}} \quad \underbrace{(0, 0, 1, 0, 0, \ldots)}_{\text{prob. } \frac{1}{8}} \quad \ldots$$

- $\lambda \notin \Lambda_{01}$: let $P = \text{Bernoulli}(0.5)$ and $A = \{\text{ all 0's }\}$

$$P \circ \lambda = \text{Bernoulli}(\tfrac{1}{1+2}) \times \text{Bernoulli}(\tfrac{1}{1+4}) \times \text{Bernoulli}(\tfrac{1}{1+8}) \times \ldots$$

Main results: the binary case

Intuition for why Lauritzen's condition is necessary

Counterexample:

$$\lambda_1 = (1, \tfrac{1}{2}),\ \lambda_2 = (1, \tfrac{1}{4}),\ \lambda_3 = (1, \tfrac{1}{8}),\ \lambda_4 = (1, \tfrac{1}{16}),\ \ldots$$

- $\lambda \notin \Lambda_{\mathsf{deFin}}$: define λ-weighted exchangeable Q as:

$$\underbrace{(1,0,0,0,0,\ldots)}_{\text{prob. } \frac{1}{2}} \quad \underbrace{(0,1,0,0,0,\ldots)}_{\text{prob. } \frac{1}{4}} \quad \underbrace{(0,0,1,0,0,\ldots)}_{\text{prob. } \frac{1}{8}} \quad \ldots$$

- $\lambda \notin \Lambda_{01}$: let $P = \text{Bernoulli}(0.5)$ and $A = \{$ all 0's $\}$

$$P \circ \lambda = \text{Bernoulli}(\tfrac{1}{1+2}) \times \text{Bernoulli}(\tfrac{1}{1+4}) \times \text{Bernoulli}(\tfrac{1}{1+8}) \times \ldots$$

- $\lambda \notin \Lambda_{\mathsf{LLN}}$: let $P = \text{Bernoulli}(0.5)$.
  Then $(P \circ \lambda_1)(\{1\}) = \frac{1}{3}$, but $\widetilde{P}_{n,1}(\{1\}) = 0$ on the event A

Main results: the binary case

Intuition for why Lauritzen's condition is necessary

Another counterexample:

$$\lambda_1 = (1, \tfrac{1}{2}),\ \lambda_2 = (\tfrac{1}{4}, 1),\ \lambda_3 = (1, \tfrac{1}{8}),\ \lambda_4 = (\tfrac{1}{16}, 1),\ \ldots$$

- $\lambda \notin \Lambda_{01}$: let $P = \text{Bernoulli}(0.5)$, and

$$A = \left\{ \sum_{i=1}^{n} X_i = n/2, \text{ for all sufficiently large even integers } n \right\}$$

  Define also

$$x_0 = ({\color{magenta}1}, 0, 1, 0, 1, 0, 1, 0, \ldots), \quad x_1 = ({\color{magenta}0}, 0, 1, 0, 1, 0, 1, 0, \ldots)$$

  Then $X = x_0 \implies A$ occurs, and $X = x_1 \implies A^c$ occurs

Main results: the finite case

Theorem: the finite case

For $|\mathcal{X}| < \infty$ it holds that

$$\Lambda_{\mathsf{deFin}} = \Lambda_{01} = \Lambda_{\mathsf{LLN}}.$$

Finite vs. infinite sequences

For exchangeability:

- $(X_1, X_2, \ldots) \rightsquigarrow$ mixture of i.i.d. distributions
- $(X_1, \ldots, X_N) \rightsquigarrow$ if $n \ll N$, $(X_1, \ldots, X_n) \approx$ mixture of i.i.d.'s[6]

[6] Diaconis & Freedman 1980, *Finite exchangeable sequences*

Finite vs. infinite sequences

For exchangeability:

- $(X_1, X_2, \ldots) \rightsquigarrow$ mixture of i.i.d. distributions
- $(X_1, \ldots, X_N) \rightsquigarrow$ if $n \ll N$, $(X_1, \ldots, X_n) \approx$ mixture of i.i.d.'s[6]

For λ-weighted exchangeability: under conditions on λ,

- $(X_1, X_2, \ldots) \rightsquigarrow$ mixture of λ-weighted i.i.d.
- Recent result by Wenpin Tang:
  $(X_1, \ldots, X_N) \rightsquigarrow$ if $n \ll N$, $(X_1, \ldots, X_n) \approx$ mixture of λ-wtd. i.i.d.'s[7]

---

[6] Diaconis & Freedman 1980, *Finite exchangeable sequences*

[7] Tang 2023, *Finite and infinite weighted exchangeable sequences*

Necessary & sufficient conditions

Necessary & sufficient conditions

Theorem: a necessary condition

If $\lambda \in \Lambda_{\mathsf{LLN}}$ then

$$\sum_{i \geq 1} \min\{(P \circ \lambda_i)(A), (P \circ \lambda_i)(A^c)\} = \infty$$

for all $P \in \mathcal{M}_{\mathcal{X}}(\lambda)$, and all events A with $\min\{P(A), P(A^c)\} > 0$.

Necessary & sufficient conditions

Theorem: a necessary condition

If $\lambda \in \Lambda_{\mathsf{LLN}}$ then

$$\sum_{i\geq 1} \min\{(P \circ \lambda_i)(A), (P \circ \lambda_i)(A^c)\} = \infty$$

for all $P \in \mathcal{M}_{\mathcal{X}}(\lambda)$, and all events A with $\min\{P(A), P(A^c)\} > 0$.

Theorem: a sufficient condition

If $\lambda \in \Lambda^\infty$ satisfies

$$\sum_{i\geq 1} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)}{\sup_{x\in\mathcal{X}} \lambda_i(x)} = \infty$$

then $\lambda \in \Lambda_{\mathsf{deFin}}$.

Necessary & sufficient conditions

Theorem: a necessary condition

If $\lambda \in \Lambda_{\mathsf{LLN}}$ then

$$\sum_{i \geq 1} \min\{(P \circ \lambda_i)(A), (P \circ \lambda_i)(A^c)\} = \infty$$

for all $P \in \mathcal{M}_{\mathcal{X}}(\lambda)$, and all events A with $\min\{P(A), P(A^c)\} > 0$.

Theorem: a sufficient condition

If $\lambda \in \Lambda^\infty$ satisfies

$$\sum_{i \geq 1} \frac{\inf_{x \in \mathcal{X}} \lambda_i(x) / \lambda_*(x)}{\sup_{x \in \mathcal{X}} \lambda_i(x) / \lambda_*(x)} = \infty \qquad \text{for some } \lambda_* \in \Lambda$$

then $\lambda \in \Lambda_{\mathsf{deFin}}$.

Overview of all results

Theorem: all results

For any standard Borel space $\mathcal{X}$ it holds that

$$\Lambda_{\mathsf{suff}} \subseteq \Lambda_{\mathsf{deFin}} \subseteq \Lambda_{01} \subseteq \Lambda_{\mathsf{LLN}} \subseteq \Lambda_{\mathsf{nec}}.$$

Moreover,

- In the binary case $\mathcal{X} = \{0, 1\}$, all five sets are equal (necessary condition = sufficient condition = Lauritzen's condition)
- In the finite case $3 \leq |\mathcal{X}| < \infty$, $\Lambda_{\mathsf{suff}} \subsetneq \Lambda_{\mathsf{deFin}} = \Lambda_{01} = \Lambda_{\mathsf{LLN}} = \Lambda_{\mathsf{nec}}$
- In the infinite case $|\mathcal{X}| = \infty$, $\Lambda_{\mathsf{suff}} \subsetneq \Lambda_{\mathsf{deFin}}$

The sufficient condition: a closer look

If $|\mathcal{X}| \geq 3$, then $\Lambda_{\mathsf{suff}} \subsetneq \Lambda_{\mathsf{deFin}}$. Why is the sufficient condition too strong?

- For $|\mathcal{X}| = 3$, here is a sequence $\lambda \notin \Lambda_{\mathsf{suff}}$:

$$\lambda_1 = (1, 1, \tfrac{1}{2}),\ \lambda_2 = (1, \tfrac{1}{4}, 1),\ \lambda_3 = (1, 1, \tfrac{1}{8}),\ \lambda_4 = (1, \tfrac{1}{16}, 1),\ \ldots$$

- Compare to our counterexample for the case $|\mathcal{X}| = 2$:

$$\lambda_1 = (1, \tfrac{1}{2}),\ \lambda_2 = (\tfrac{1}{4}, 1),\ \lambda_3 = (1, \tfrac{1}{8}),\ \lambda_4 = (\tfrac{1}{16}, 1),\ \ldots$$

(violates the necessary condition)

Proof: the sufficient condition

Sufficient condition (simplified version)

$$\sum_{i \geq 1} \frac{\inf_{x \in \mathcal{X}} \lambda_i(x)}{\sup_{x \in \mathcal{X}} \lambda_i(x)} = \infty$$

- Goal: any λ-weighted exch. Q can be written as $Q = (P \circ \lambda)_\mu$
- Moreover, after observing the sequence $X = (X_1, X_2, \ldots) \sim Q$, the random P is equal to a limit of weighted empirical distrib.'s

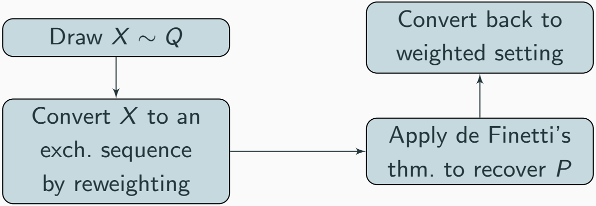
Proof: the sufficient condition

Sufficient condition (simplified version)

$$\sum_{i\geq 1} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)}{\sup_{x\in\mathcal{X}} \lambda_i(x)} = \infty$$

- Goal: any λ-weighted exch. Q can be written as $Q = (P \circ \lambda)_\mu$
- Moreover, after observing the sequence $X = (X_1, X_2, \ldots) \sim Q$, the random P is equal to a limit of weighted empirical distrib.'s

Proof strategy:



Proof: the sufficient condition

Given $X = (X_1, X_2, \ldots) \sim Q$, how do we convert to an exch. sequence?

Proof: the sufficient condition

Given $X = (X_1, X_2, \ldots) \sim Q$, how do we convert to an exch. sequence?

First attempt: reweight the distribution

- For each n, reweight the marginal Q_n:

$$\mathrm{d}P_n(x_1, \ldots, x_n) \propto \frac{\mathrm{d}Q_n(x_1, \ldots, x_n)}{\lambda_1(x_1) \cdot \ldots \cdot \lambda_n(x_n)}$$

  Q is λ-weighted exch. $\Longrightarrow$ P_n is exch. for each n

Proof: the sufficient condition

Given $X = (X_1, X_2, \ldots) \sim Q$, how do we convert to an exch. sequence?

First attempt: reweight the distribution

- For each n, reweight the marginal Q_n:
$$\mathrm{d}P_n(x_1, \ldots, x_n) \propto \frac{\mathrm{d}Q_n(x_1, \ldots, x_n)}{\lambda_1(x_1) \cdot \ldots \cdot \lambda_n(x_n)}$$
  Q is λ-weighted exch. $\implies P_n$ is exch. for each n

- Can we take $n \to \infty$ and apply de Finetti's theorem?
  No — the P_n's are not consistent with each other!

Proof: the sufficient condition

Given $X = (X_1, X_2, \ldots) \sim Q$, how do we convert to an exch. sequence?

Proof: the sufficient condition

Given $X = (X_1, X_2, \ldots) \sim Q$, how do we convert to an exch. sequence?

Second attempt: subsample the sequence

- Sample $X = (X_1, X_2, \ldots) \sim Q$
- Independently for each i,

$$\begin{cases} \text{Accept sample } X_i \text{ with probability } p_i = \frac{\inf_{x \in \mathcal{X}} \lambda_i(x)}{\lambda_i(X_i)} \\ \text{Otherwise, reject sample } X_i \end{cases}$$

Lemma

The subsequence of accepted samples has an exchangeable distribution.

Proof: the sufficient condition

Given $X = (X_1, X_2, \ldots) \sim Q$, how do we convert to an exch. sequence?

Second attempt: subsample the sequence

- Sample $X = (X_1, X_2, \ldots) \sim Q$
- Independently for each i,

$$\begin{cases} \text{Accept sample } X_i \text{ with probability } p_i = \frac{\inf_{x \in \mathcal{X}} \lambda_i(x)}{\lambda_i(X_i)} \\ \text{Otherwise, reject sample } X_i \end{cases}$$

$\sum_{i \geq 1} p_i = \infty$

Lemma

The subsequence of accepted samples has an exchangeable distribution. Moreover, it has infinite length a.s. under the sufficient condition.

Proof: the sufficient condition

The rest of the proof....

- Apply de Finetti's theorem to the infinite exchangeable subsequence $\rightsquigarrow$ the subsequence is $\overset{\text{iid}}{\sim} P$ for a *random* distribution P
- Q is equivalent to sampling from $P \circ \lambda$ for this *random* P

Proof: the sufficient condition

What is the random distribution P?

- By de Finetti's theorem + LLN,

$$P(A) = \lim_{n\to\infty} \underbrace{\frac{\sum_{i=1}^{n} \mathbb{1}\{X_i \in A \text{ and } X_i \text{ accepted}\}}{\sum_{i=1}^{n} \mathbb{1}\{X_i \text{ accepted}\}}}_{\text{empirical distrib. of the exchangeable subsequence}}$$

- Since X_i's are accepted or rejected independently for each i, an equivalent limit:

$$P(A) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathbb{1}\{X_i \in A\} \cdot \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)}{\lambda_i(X_i)}}{\sum_{i=1}^{n} \frac{\inf_{x\in\mathcal{X}} \lambda_i(x)}{\lambda_i(X_i)}}$$

Summary

Summary & open questions

Our results prove a weighted version of de Finetti's theorem

(and related properties like the 0-1 law),

generalizing Lauritzen's results for $\mathcal{X} = \{0, 1\}$ to a general space $\mathcal{X}$

Open questions:

- A gap in the theory for the infinite case $|\mathcal{X}| = \infty$:

$$\Lambda_{\mathsf{suff}} \subsetneq \Lambda_{\mathsf{deFin}} \underset{\uparrow}{\subseteq} \Lambda_{\mathsf{01}} \underset{\uparrow}{\subseteq} \Lambda_{\mathsf{LLN}} \underset{\uparrow}{\subseteq} \Lambda_{\mathsf{nec}}$$

are these $=$ or $\subsetneq$?

- Implications for statistical inference?
  e.g., asymptotics for predictive coverage under distribution drift

---

Thank you!