Exchangeable graphs, conditional independence, and computably-measurable samplers

DANIEL M. ROY
UNIVERSITY OF CAMBRIDGE

Joint work with

Nate Ackerman (Harvard)

Jeremy Avigad (CMU)

Cameron Freer (MIT)

Jason Rute (U of Hawaii–Manoa)

Computability and Complexity in Analysis
Nancy, France, July 2013

Three vignettes

(1) Exchangeable sequences of random variables

(2) Exchangeable sequences of random sets with exchangeable increments

(3) Exchangeable arrays of random variables

In each case, statisticians have come up against computational difficulties and in each case computably analysis sheds some light on what's going on.

Recurring themes

(a) How can we represent such processes?

Representation

Computability

(b) Implications for *probabilistic programming*

Computable a.e. versus computably measurable

Conditional independence

(c) Inference in stochastic process models

"Exact approximate" inference

Exchangeable sequences of random variables

Let H be a probability measure on $\mathbb{R}$ and consider the sequence $Y_1, Y_2, \dots$ of random variables such that

$$\mathbb{P}(Y_1 \in \cdot\,) = H \tag{1}$$

and for every $n \in \mathbb{N}$,

$$\mathbb{P}(Y_{n+1} \in \cdot \mid Y_1, \dots, Y_n) = \frac{1}{n+1} H + \frac{n}{n+1} \hat{P}_n, \tag{2}$$

where $\hat{P}_n \equiv \sum_{i=1}^n \delta_{Y_i}$ is the empirical distribution.

$Y_1, Y_2, \dots$ is a (labeled) **Chinese restaurant process** and this process has been hugely influential in nonparametric Bayesian statistics in the last 15 years in **clustering**.

Despite the dependence of Y_{n+1} on earlier values

$$(Y_1, Y_2, \dots) \overset{d}{=} (Y_{\pi_1}, Y_{\pi_2}, \dots) \tag{3}$$

for every permutation π of $\mathbb{N}$, i.e., **the sequence is exchangeable**.

Thm (de Finetti)**.** *An infinite sequence of random variables $Y = (Y_1, Y_2, \dots)$ is exchangeable if and only if it is conditionally i.i.d. (independent and identically distributed). In particular, there is a random probability measure ν s.t.*

$$\mathbb{P}(Y \in \cdot \mid \nu) = \nu^\infty \ \ a.s. \tag{4}$$

If you know ν, you can sample Y_i's in parallel.

In the case of a Chinese restaurant process, we can described ν quite explicitly. In particular,

$$\nu = \sum_{i=1}^{\infty} V_i \delta_{\tilde{Y}_i} \text{ a.s.} \tag{5}$$

$$\tilde{Y}_1, \tilde{Y}_2, \ldots \sim H^{\infty} \tag{6}$$

$$U_1, U_2, \ldots \sim \mathrm{U}(0,1)^{\infty} \tag{7}$$

$$V_j \equiv U_j \prod_{i<j} (1 - U_i), \quad i \in \mathbb{N}. \tag{8}$$

ν is a so-called Dirichlet process, an infinite dimensional object. This was a major algorithmic road block for statisticians until Papaspiliopoulos and Roberts (2008) suggested to only generate random variables as you need them. This is (naïve) computable analysis in practice!

Can we expose the conditional independence in general?

Thm (Freer and **R.**, 2012)**.** *The distribution of an exchangeable sequence $Y_1, Y_2, \ldots$ is computable if and only if the distribution of its directing random measure ν is computable.*

In theory, you can always parallelize an algorithm for generating an exchangeable sequence.
In practice, conditional independence (i.e., the opportunity to parallelize) is absolutely critical for efficient inference.

Exchangeable sequences of random sets

In some cases, there is additional conditional independence structure. Recall that a Poisson (point) process with (finite) mean γH is a random set

$$\{S_1, \dots, S_\kappa\} \tag{9}$$

$$S_1, S_2, \dots \sim H^\infty \tag{10}$$

$$\kappa \sim \text{Poisson}(\gamma) \tag{11}$$

Consider the following exchangeable sequence of *sets*: Y_1 is a Poisson (point) process with mean H, and for each $n \in \mathbb{N}$,

$$Y_{n+1} \setminus (Y_1 \cup \cdots \cup Y_n) \tag{12}$$

is a Poisson (point) process with mean $\frac{1}{n+1} H$ and

$$\mathbb{P}(s \in Y_{n+1} \mid Y_1, \dots, Y_n) = \frac{\#\{j \le n : s \in Y_j\}}{n+1}.$$

$Y_1, Y_2, \dots$ is a (labeled) **Indian buffet process** and it has also been hugely influential in nonparametric Bayesian statistics in the past 6 years in **clustering** with overlapping groups.

Now again, the sequence $Y = (Y_1, Y_2, \dots)$ is exchangeable and so there is a random probability measure ν (on the space of finite sets) such that

$$\mathbb{P}(Y \in \cdot \mid \nu) = \nu^\infty.$$

But there's a lot more structure!

In particular,

(1) If $A_1, \ldots, A_k$ are disjoint sets, then the sets

$$Y_1 \cap A_1, \ldots, Y_1 \cap A_k$$

are independent conditional on ν, i.e., the Y_j have exchangeable increments; and

(2) if ϕ is a H-measure preserving transformation, then the sequence $Y'_n = \phi(Y_n)$, $n \in \mathbb{N}$, has the same distribution as Y_n, $n \in \mathbb{N}$.

This implies that there is a random countable sequence P in $[0,1]$ such that

$$P_1 \geq P_2 \geq \cdots > 0 \quad \text{and} \quad \sum_i P_i < \infty \text{ a.s.}$$

and an i.i.d.-H collection $\tilde{S} = \{\tilde{S}_1, \tilde{S}_2, \ldots\}$ such that

$$Y_j \subset \tilde{S} \text{ a.s.} \tag{13}$$

$$\mathbb{P}(\tilde{S}_j \in Y_j | \tilde{S}, P) = P_i. \tag{14}$$

In particular, one can show that

$$P_n = \prod_{j \leq n} U_j \tag{15}$$

$$U_1, U_2, \ldots \sim \mathrm{U}(0,1)^\infty. \tag{16}$$

Again, ν (equivalently, P and $\tilde{S}$) are infinite dimensional, but **the same tricks for computation don't work.** In practice, the sequence is truncated so that $P_m = 0$ for all sufficiently large m.

Lem (**R.**). *The probability* $\mathbb{P}(Y_1 = \emptyset \mid P = \cdot)$ *is a discontinuous everywhere function on every measure one set.*

Statisticians were worried about truncation. So they developed an *auxiliary variable method* called **slice sampling** to remove the approximation induced by truncation, but maintain the conditional independence.

Thm (slice sampling). *Define*

$$T = \min\{P_j \,:\, \tilde{S}_j \in Y_1 \cup \cdots \cup Y_n\},$$

and let ξ *be uniformly distributed on* $[0, T]$. *Then* $\mathbb{P}(Y_1 \in \cdot \mid \tilde{S}, P, \xi)$ *and* $\mathbb{P}(\xi | Y_1, \dots, Y_n, \tilde{S}, P)$ *are computable a.e.*

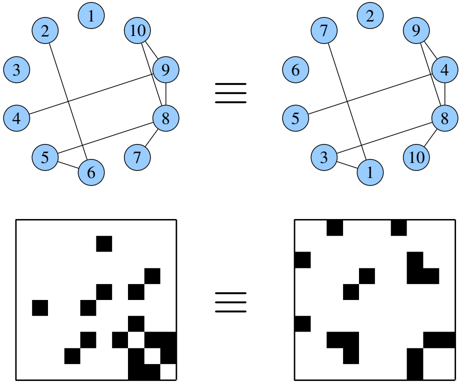
What's going on here?

Thm (**R.**). $\mathbb{P}(Y_1 \in \cdot \mid \tilde{S}, P)$ *is computable on a set of measure* $1 - 2^{-k}$, *uniformly in* k.

Say such a function is **computably measurable**. This representation dates back to Kriesel-Lacombe (1957) and Šanin (1968), who proposed notions of effectively measurable sets. Later, Ko (1986) built on this work, studying computably measurable functions.This is also related to layerwise-computable functions and L^p-computable functions.

Exchangeable arrays of random variables

Let $X = (X_{i,j})_{i,j\in\mathbb{N}}$ be an array of random variables in some space S. (E.g., $X_{i,j} \in \{0,1\}$, representing the adjacency matrix of a graph.)

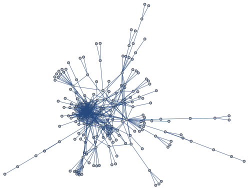


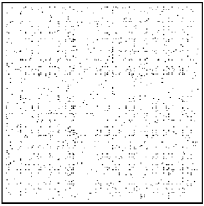
Defn. Call X **(jointly) exchangeable** when

$$(X_{i,j})_{i,j\in\mathbb{N}} \overset{d}{=} (X_{\pi(i),\pi(j)})_{i,j\in\mathbb{N}} \tag{17}$$

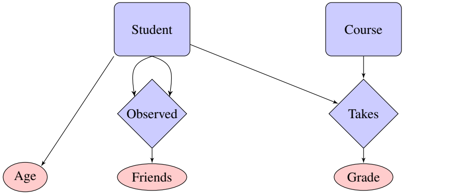
holds for every permutation π of $\mathbb{N}$.

Most figures by James Lloyd (Cambridge) and Peter Orbanz (Columbia)





- Links between websites
- Products that customers have purchased
- Proteins that interact
- Relational databases



15
15
15
14
14
16

	✓	✓	✓	✓	✓
✓		×			
✓	×				
✓					✓
✓					
✓			✓		

A									
							E	A	
	C			D			B		F
				D					
		B			C				

Let λ be Lebesgue measure on $[0,1]$.
Let $\tilde{\mathbb{N}}^d \equiv \{s \subset \mathbb{N} : |s| \le d\}$.
Let U_s, $s \in \tilde{\mathbb{N}}^2$, be i.i.d.-$\lambda$.
Write $U_i \equiv U_{\{i\}}$.

$$\begin{array}{llllll} U_\emptyset & U_1 & U_2 & U_3 & U_4 & \cdots \\ & & U_{\{1,2\}} & U_{\{1,3\}} & U_{\{1,4\}} & \cdots \\ & & & U_{\{2,3\}} & U_{\{2,4\}} & \cdots \\ & & & & U_{\{3,4\}} & \cdots \\ & & & & & \ddots \end{array}$$

Defn (standard exchangeable array)**.** Let $f : [0,1]^4 \to S$ be a measurable function, and put

$$X_{i,j} = f(U_\emptyset, U_i, U_j, U_{\{i,j\}}), \quad i,j \in \mathbb{N}. \tag{18}$$

By a **standard (exchangeable) array** we mean an array with the same distribution as X for some f.

Thm (Aldous, Hoover)**.** *An infinite array X is exchangeable if and only it is standard, i.e.,*

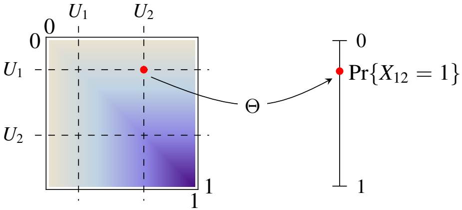
$$(X_{i,j})_{i,j\in\mathbb{N}} \overset{d}{=} (f(U_\emptyset, U_i, U_j, U_{\{i,j\}}))_{i,j\in\mathbb{N}} \tag{19}$$

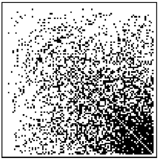
for some measurable function $f : [0,1]^4 \to S$.

Example (exchangeable graph).
Assume $X_{i,j} \in \{0,1\}$ and $X_{i,j} = X_{j,i}$ a.s.
X is the adjacency matrix of a random graph on $\mathbb{N}$.

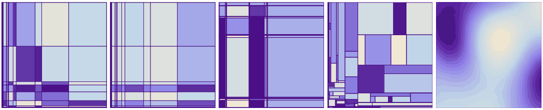
Let $\mathcal{W}$ be the space of symmetric measurable functions from $[0,1]^2$ to $[0,1]$.
Such functions are called "graphons".

If X is exchangeable, it's standard w.r.t some f.
Let $\Theta(x,y) \equiv \lambda\{u \in [0,1] : f(U_\emptyset, x, y, u) = 1\}$
then Θ is a random element in $\mathcal{W}$.





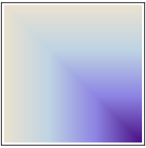
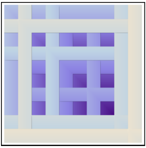
.



Question: Let X be an exchangeable array, standard w.r.t. a function f. If X has a computable distribution, is f computable?

Note that the element Θ is not uniquely determined by the distribution of X. Let $T : [0,1] \to [0,1]$ be a measure preserving transformation, and define

$$\Theta^T(x,y) \equiv \Theta(T(x),T(y)). \tag{20}$$

Then Θ' and Θ induce the same distribution on graphs. Let $\sim$ be equivalence up to a measure preserving transformation.

Thm (Hoover). *The measurable function f underlying an exchangeable array is unique up to a measure preserving transformation.*

de Finetti's theorem is a special case of Aldous-Hoover.

Cor. *An infinite sequence $Y = (Y_i)_{i \in \mathbb{N}}$ is exchangeable if and only if*

$$(Y_i)_{i \in \mathbb{N}} \overset{d}{=} (g(U_\emptyset, U_i))_{i \in \mathbb{N}} \tag{21}$$

for some measurable function $g : [0,1]^2 \to S$.

The random measure

$$\nu = \mathbb{P}(Y_1 \in \cdot \mid U_\emptyset) = \mathbb{P}(g(U_\emptyset, U_1) \in \cdot \mid U_\emptyset) \tag{22}$$

is the a.s. unique random measure satisfying

$$\mathbb{P}(Y \in \cdot \mid \nu) = \nu^\infty \text{ a.s.} \tag{23}$$

Thm (Freer and **R.**, 2012). *The distribution of the sequence $Y_1, Y_2, \dots$ is computable if and only if the distribution of ν is computable.*

Cor. *Let $Y : [0,1] \to S^\infty$ be a measurable function such that $Y(U_\emptyset)$ is a exchangeable sequence. If Y is λ-a.e. computable then there exists a function $g : [0,1]^2 \to S$ that is λ^2-a.e. computable that satisfies*

$$Y(U_\emptyset) \overset{d}{=} (g(U_\emptyset, U_1), g(U_\emptyset, U_2), \dots). \tag{24}$$

Question: Is the analogous result for exchangeable arrays true?

Thm (AFRR)**.** *No.*

Proof sketch. Let μ be the distribution of an exchangeable graph with a nonrandom graphon Θ. Such an exchangeable graph is **ergodic**. Then Lovász and Szegedy (2006) proved that the map

$$\mu \mapsto \int_0^1 \int_0^1 [\Theta(x,y)]^2 \mathrm{d}x\mathrm{d}y \tag{25}$$

is discontinuous w.r.t the weak topology. This already rules out computability. $\square$

But note that if f only takes values in $[0,1]$, then this function is continuous.

Question: If we restrict attention to graphons taking values in $\{0,1\}$, can we compute a graphon from the distribution of a graph it induces?

Thm (AFRR)**.** *No.*

AFRR = Avigad,Freer,**R.**,Rute

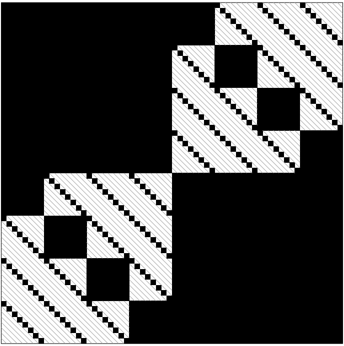
Construction

Write $x_1x_2\ldots$ for the a.s. unique binary expansion of a uniform random variable x in $[0,1]$.

Consider the symmetric function $\Psi : [0,1]^2 \to \{0,1\}$ given by

$$\Psi(x_1x_2\ldots,\, y_1y_2\ldots) = \begin{cases} 1 & (\exists n \in \mathbb{Z}_+)\,(\forall j \in \{2^n, 2^{n+1}-1\})\ (x_j = y_j), \\ 0 & \text{otherwise.} \end{cases}$$

Here is a picture of this function (1=black, 0=white):



Thm (AFRR). *Let $U_1, U_2, \dots$ be i.i.d. uniform, and consider the exchangeable graph with edges*

$$X_{i,j} = \Psi(U_i, U_j). \tag{26}$$

Then the distribution of X is computable, but there is no a.e. computable version of Ψ.

Proof sketch. For Ψ to be a.e. computable it must be continuous on a measure one set. However, $\Psi^{-1}\{0\}$ is a nowhere dense set of positive measure

$$\frac{1}{2} \cdot \frac{3}{4} \cdot \frac{7}{8} \cdots \approx 0.289, \tag{27}$$

and so Ψ is not continuous on a measure one set. The (slightly harder) part is showing that this property is true also for every *weakly isomorphic* function g, i.e., functions g that generate graphs X' with the same distribution as X. $\square$

Now what?

Silver lining?

Let μ be a computable distribution on a computable metric space T, let S be a computable metric space, and let $f : T \to S$ be a measurable function.

Defn. Recall f is **computably-measurable** when it is computable on a set of μ-measure $1 - 2^{-k}$, uniformly in k.

Thm (AFRR). *Let X be an ergodic exchangeable array that is computable and such that there is an underlying nonrandom graphon Θ that takes values in $\{0, 1\}$. Then there is a computably-measurable version of Θ, uniformly in the distribution of X.*

Let $f : [0, 1]^3 \to [0, 1]$ and define the exchangeable **multigraph**

$$X^k_{i,j} = f(U_\emptyset, U_i, U_j, U^k_{\{i,j\}}). \tag{28}$$

Each X^k is an ergodic exchangeable array with graphon $\Theta(x, y) = \lambda\{u : f(x, y, u) = 1\}$.

Thm (AFRR). *Let X be an exchangeable multigraph that is computable and such that there is an underlying nonrandom graphon Θ. Then there is a computably-measurable version of Θ, uniformly in the distribution of X.*

Probabilistic programming

Probabilistic programming is an approach to statistical modeling where the statistician

(1) uses a program to define a probabilistic model (X, Y, Θ) of some quantities (x, y, θ), and

(2) performs statistical analysis using generic algorithms that take these *programs* as input and compute various conditional distributions, e.g., $\mathbb{P}(\Theta = \theta \mid X = x, Y = y)$.

Probabilistic programs have been identified with a.e. computable functions from $\{0,1\}^{\mathbb{N}} \to S$ for some computable metric space S.

This work suggests that we should possibly consider re-founding probabilistic programming on computably-measurable representations of distributions as a.e. computable representations rule out exposing important conditional independencies in some cases.

Conclusions

(1) All computable exchangeable sequences can be sampled in a parallel way.

(2) This is no longer true for exchangeable arrays.

(3) If we are happy with the sampler failing with some probability that we control, we can produce parallel samplers again.

(4) Given how important conditional independence is to efficient inference, the main representational result suggests that we might rethink the current foundation of probabilistic programming on a.e. computability.

(5) We can potentially eliminate the error introduced through "truncation" by using more general versions of slice sampling.