STA4513: Statistical Models of Networks Fall 2014

Lecture 3 — 24 September, 2014

Prof. Daniel M. Roy *Scribe: Victor Veitch*

1 Overview

The topic of this lecture is exchangeability of random variables and the associated representation theorems. The flow of the lecture is roughly

1. Introduce the concept of exchangeability for both sets of random variables and arrays of random variables.
2. Explain de Finetti's representation theorem for exchangeable sets of random variables, and
3. Explain the Aldous-Hoover representation theorem for exchangeable arrays.

2 Review of Exchangeability and Conditional Independence

We begin with a few simple examples:

Example 1. Suppose we wish to model the age, weight and height of individuals in a population as a random variable; writing the random variable for the ith individual as $X_i = (A_i, H_i, W_i)$. An (unqualified) independence assumption is probably not appropriate here: if, for instance, we measure the heights of the first 100 individuals this information will impact our prediction of the height of the 101st individual. However, the sequence of random variables $X_1, X_2, \dots$ can be assumed to be exchangeable in the sense that our inferences do not depend on the order in which we observed the individuals.

Example 2. If we wish to make inferences about the weather we might measure the peak temperature and humidity each day, denoting these values for the ith day as X_i. In this case an exchangeability assumption would not be appropriate because, for instance, if we want to predict the temperature on the 21st day the temperature on the 20th day is relevant in a different way than the temperature on the 15th day. However, it may be appropriate to assume a translation symmetry here: $\Pr(X_1, X_2, \dots) = \Pr(X_{k+1}, X_{k+2}, \dots)$. The physical content of this equation is that the absolute values of the labeling indices are not significant, only the distance between them matters; i.e. it does not matter if the calendar starts at day 0 or at day k.

Example 3. If we model symmetric connections between people i and j as a random variable X_{ij} then it is reasonable to model the joint distribution of the observations as invariant under joint permutations σ of the indices, $X_{ij} \to X_{\sigma(i),\sigma(j)}$. For instance, if we want to model the relationships between people A, B, C then this assumption would imply in particular:

$$\Pr_X \begin{pmatrix} x_{AA} & x_{AB} & x_{AC} \\ x_{AB} & x_{BB} & x_{BC} \\ x_{AC} & x_{BC} & x_{CC} \end{pmatrix} = \Pr_X \begin{pmatrix} x_{AA} & x_{AC} & x_{AB} \\ x_{AB} & x_{CC} & x_{BC} \\ x_{AB} & x_{BC} & x_{BB} \end{pmatrix},$$

i.e. that the distribution is invariant under swapping the labels B and C.

In this lecture we are interested in the types of probabilistic symmetry present in the 1st and 3rd examples. To that end we now give a more formal definition of exchangeability,

Definition. Let $X = (X_1, X_2, \dots)$ be a set of random variables on a space S,[1] then we say that X is *exchangeable* when any of the following equivalent conditions hold:

1. For all permutations $\pi : \mathbb{N} \to \mathbb{N}$ it holds that $X_1, X_2, \dots \overset{d}{=} X_{\pi(1)}, X_{\pi(2)}, \dots$
2. $\forall n \in \mathbb{N}$ and $\forall \pi \in S_n$ the group of permutations of $[n] = \{1, \dots n\}$ it holds that $X_1, \dots, X_n \overset{d}{=} X_{\pi(1)}, \dots, X_{\pi(n)}$
3. $\forall n \in \mathbb{N}$ and distinct $k_1, \dots k_n \in \mathbb{N}$ it holds that $X_1, \dots, X_n \overset{d}{=} X_{k_1}, \dots, X_{k_n}$
4. $\forall A_1, \dots A_n \subseteq S$ st A_i are measurable and $\forall \pi \in S_n$ it holds that
$$\Pr\left(X_1 \in A_1, \dots, X_n \in A_n\right) = \Pr\left(X_{\pi(1)} \in A_1, \dots, X_{\pi(n)} \in A_n\right)$$
or equivalently
$$\Pr\left(X_1 \in A_1, \dots, X_n \in A_n\right) = \Pr\left(X_1 \in A_{\pi(1)}, \dots, X_n \in A_{\pi(n)}\right).$$

That 2 implies 1 is a consequence of the Kolmogorov extension theorem.

Our short term goal is an explanation of de Finetti's representation theorem. This is going to require a careful statement of what it means for a set of random variables to be conditionally i.i.d. We first give the unconditional definition as a warm up,

Definition 4. A set $X_1, X_2, \dots$ of random variables defined on S is independent if for all measurable $A_1, \dots A_n \in S$ it holds that$\Pr\left(X_1 \in A_1, \dots, X_n \in A_n\right) = \prod_{i=1}^n \Pr\left(X_i \in A_i\right)$

Definition 5. A set $X_1, X_2, \dots$ of random variables defined on S is independently identically distributed (iid) if for all measurable $A_1, \dots A_n \in S$ it holds that$\Pr\left(X_1 \in A_1, \dots, X_n \in A_n\right) = \prod_{i=1}^n \Pr\left(X_1 \in A_i\right)$

Conditional independence is nearly identical notationally but conceptually trickier. The key point is that if $\theta : S \to T$ is some random variable than $\Pr\left(X_1 \in A | \theta\right)$ is itself a random variable (mapping from S to $[0, 1]$). Indeed, we can define a random measure ν_1^θ on S by taking $\nu_1^\theta(A) = \Pr\left(X_1 \in A | \theta\right)$ for all $A \in \sigma(S)$.[2] Any particular realization of this random variable is a measure on S and ν_1^θ is a measurable function from the domain of theta to the space of measures on S.

Definition 6. A set $X_1, X_2, \dots$ of random variables defined on S is independent conditional on random variable θ if for all measurable $A_1, \dots A_n \in S$ it holds that $\Pr\left(X_1 \in A_1, \dots, X_n \in A_n | \theta\right) = \prod_{i=1}^n \Pr\left(X_i \in A_i | \theta\right)$

Definition 7. A set $X_1, X_2, \dots$ of random variables defined on S is independently identically distributed (iid) conditional on random variable θ if for all measurable $A_1, \dots A_n \in S$ it holds that$\Pr\left(X_1 \in A_1, \dots, X_n \in A_n | \theta\right) = \prod_{i=1}^n \Pr\left(X_1 \in A_i | \theta\right) = \prod_{i=1}^n \nu_1^\theta(A_i)$

[1] Assume S is a Polish space if you're fussed about such things.

[2] $\sigma(S)$ denotes the sigma algebra of S.

Notice that, in contrast to the unconditional case, these definitions deal with the equality of *random variables.*

Also notice that many different random variables θ can result in the same random measure ν^{θ}. For example, suppose we were interested in conditioning on a normally distributed random variable with fixed variance Λ and random mean $K+L$. In this case we might take $\theta=(K,L)$. However, we might instead have chosen distinct random variables W and Z such that $K+L=W+Z$ and conditioned on random variable $\tilde{\theta}=(W,Z)$. For this model

$$\Pr\left(X_1\in A|\theta\right)=\Pr\left(X_1\in A|\tilde{\theta}\right)\ \forall\text{ measureable }A\subseteq S$$

and the corresponding random measures ν^{θ} and $\nu^{\tilde{\theta}}$ are equal. This suggests an obvious canonical choice of random variable to condition on: the random measure ν itself. That is, we have

$$\begin{gathered}\nu^{\theta}=\nu^{\tilde{\theta}}=\nu\text{ and}\\ \Pr\left(X_1\in A|\theta\right)=\Pr\left(X_1\in A|\tilde{\theta}\right)=\Pr\left(X_1\in A|\nu\right)\forall\text{ measureable }A\subseteq S\end{gathered}$$

3 de Finetti's Representation Theorem

Notice that if the set $X_1,X_2,\ldots$ is iid conditional on the random measure ν then

$$\begin{aligned}\Pr\left(X_1\in A_1,\ldots X_n\in A_n\right)&=\mathbb{E}_{\nu}\left[\Pr\left(X_1\in A_1,\ldots X_n\in A_n|\nu\right)\right]\\&=\mathbb{E}_{\nu}\left[\prod_{i=1}^{n}\nu\left(A_i\right)\right],\end{aligned}$$

from which it is immediately manifest that $X_1,X_2,\ldots$ is exchangeable. That is, conditional iid implies exchangeability. The remarkable content of de Finetti's eponymous theorem is that these two things are in fact equivalent:

Theorem 8. *(de Finetti, Hewitt-Savage) Let $X_1,X_2,\ldots$ be an infinite sequence of random variables on S. The following are equivalent:*

1. *X is exchangeable, i.e. $\left(X_1,X_2,\ldots\right)\overset{d}{=}\left(X_{\pi(1)},X_{\pi(2)},\ldots\right)\ \forall\pi\in S_{\infty}$*
2. *X is conditionally iid; i.e. there exists a random measure ν such that $\forall n\in\mathbb{N}\ \Pr\left(X_1\in A_1,\ldots,X_n\in A_n|\nu\right)=\prod_{i=1}^{n}\nu\left(A_i\right)$*

This theorem is often alternatively phrased as saying that exchangeability implies that $\forall n\in\mathbb{N}$

$$\Pr\left(X_1\in A_1,\ldots,X_n\in A_n\right)=\int_{v}\prod_{i=1}^{n}v\left(A_i\right)\mu\left(dv\right),$$

where μ is a measure on the space of random measures on S.

To build intuition for what this means lets consider a concrete example. We can model flipping a (not necessarily fair) coin by assigning a random variable

$$X_i=\begin{cases}0&\text{tails}\\1&\text{heads}\end{cases}$$

to the ith flip. An unconditional independence assumption is inappropriate for this situation: if the coin is flipped once and comes up heads then, in the absence of any relevant prior knowledge, we should bet that the second toss will also be heads. However, an exchangeability assumption is reasonable: we will base our inferences on the number of observed heads and tails, but not on the order in which they occurred. In the language of de Finetti's theorem we then have that

$$\Pr\left(X_1 = x_1, \ldots, X_n = x_n | \nu\right) = \prod_{i=1}^{n} \nu\left(x_n\right)$$

where $x_j \in \{0,1\}$ and ν is a random measure on $\{0,1\}$. In this particular instance the collection of random measures has a simple parametric form: a realization of ν is a Bernoulli distribution. Since Bernoulli distributions are indexed by a single parameter $p \in [0,1]$ we have a simple correspondence

$$\nu \in_R \left\{\mathrm{Bern}\left(p\right)\right\}_{p\in[0,1]} \leftrightarrow \nu = \mathrm{Bern}\left(p\right),\ p \in_R [0,1],\ \text{whence}$$

$$\begin{aligned}\Pr\left(X_1 = x_1, \ldots, X_n = x_n\right) &= \int_v \prod_{i=1}^{n} v\left(x_n\right)\mu\left(dv\right)\\ &= \int_{[0,1]} \prod_{i=1}^{n} p^{x_i}\left(1-p\right)^{1-x_i} F\left(dp\right)\\ &= \int_{[0,1]} p^{S}\left(1-p\right)^{n-S} F\left(dp\right),\ S = \sum_{i=1}^{n} X_i.\end{aligned}$$

This recovers the familiar treatment of repeated flips of a biased coin given in introductory probability courses: conditional on the "propensity" p of the coin the number of heads in n flips is $\mathrm{Bin}\left(n,p\right)$. If we do not know the parameter p then we must integrate it out.

de Finetti's theorem can enable very powerful models even in cases where the random measure ν lacks a simple form, or even a finite dimensional representation. Dirichlet process models (e.g. DP mixture models) are a well developed example of this.

4 Exchangeable Graphs and Arrays

We now turn to the treatment of exchangeable arrays. Here we'll be dealing with random variables X_{ij} that we think of as representing some (symmetric) relation between units i and j. These can generally be thought of as edges of some graph. We are interested in infinite exchangeable arrays of these things, with the exposition of the corresponding representation theorem as our goal for this section.

Definition 9. A random array $X = \left\{X_{ij}\right\}_{i,j\in\mathbb{N}}$ is *jointly exchangeable* if for every permutation π of $\mathbb{N}$,

$$\left\{X_{ij}\right\}_{i,j\in\mathbb{N}} \overset{d}{=} \left\{X_{\pi(i)\pi(j)}\right\}_{i,j\in\mathbb{N}}.$$

A random graph is called jointly exchangeable if its adjacency matrix is jointly exchangeable.

To build some intuition we consider a series of examples:

Example 10. $X_{ij} = 0$ for $i < j$ is trivially jointly exchangeable, since every distinct pair of units is disconnected.

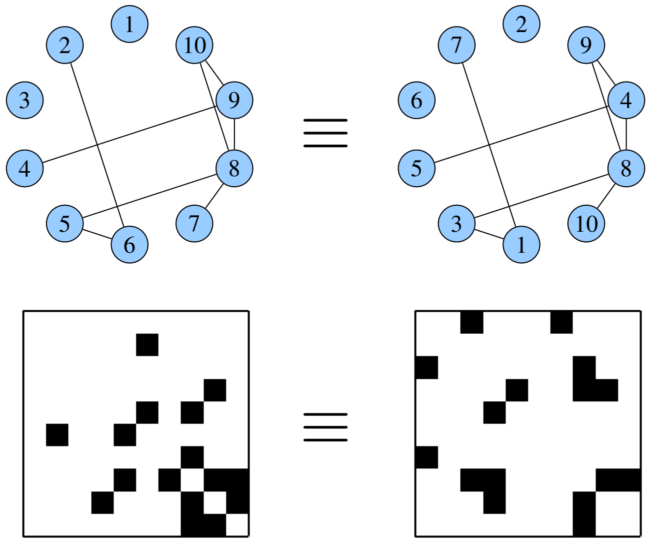

Figure 1: This represents a particular permutation of the labels of the vertices of a random graph. If the distribution of the random graph is invariant under all such permutations then it is jointly exchangeable.

Example 11. $X_{ij} = (j - i) \mod 2$ is not jointly exchangeable.

Example 12. $X_{ij} \overset{iid}{\sim} \text{Bern}\,(p)$ is jointly exchangeable.

Example 13. $X_{ij}|p \overset{iid}{\sim} \text{Bern}\,(p)$ is jointly exchangeable.

Example 14. $X_{ij}|\boldsymbol{N} \overset{iid}{\sim} \text{Bern}\,(\sigma\,(\langle N_i, N_j\rangle))$, where $N_i \overset{iid}{\sim} N\,(\mu, \Lambda)$, $\langle , \rangle$ denotes inner product and $\sigma : \mathbb{R} \to [0, 1]$. This model is also exchangeable. (nb. this is closely related to the eigenmodel).

To state the representation theorem corresponding to this flavour of exchangeability we will need one final definition:

Definition 15. Let $U_1, U_2, \ldots$ be iid uniform random variables in $[0, 1]$. Let $\Theta : [0, 1]^2 \to [0, 1]$ be a symmetric measurable function and let

$$X_{ij} = 1 \text{ with probability } \Theta\,(U_i, U_j)$$

independently for all $i < j \in \mathbb{N}$. A Θ-*random graph* is an array with the same distribution as X. A *graphon* is a symmetric measurable function from $[0, 1]^2$ to $[0, 1]$. See figure 2.

Example 14 is an example of a Θ-random graph, with

$$\Theta\,(\cdot, \cdot) = \sigma\left(\langle \Phi^{-1}\,(\cdot)\,, \Phi^{-1}\,(\cdot)\rangle\right),$$

where $\Phi^{-1}\,(\cdot)$ is the pseudo-inverse of the $N\,(\mu, \Lambda)$ cumulative distribution function. That is, $\Phi^{-1}\,(\cdot)$ is the function such that $\Phi^{-1}\,(U_i) \overset{d}{=} N_i$ where $U_i \sim U\,[0, 1]$ and $N_i \sim N\,(\mu, \Lambda)$. This

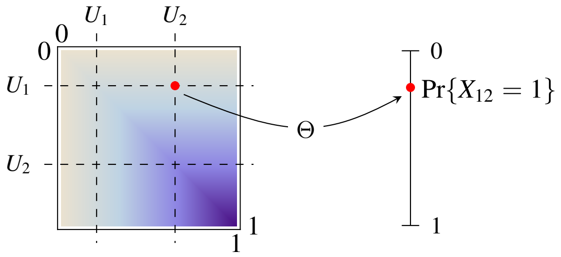

Figure 2: A pictorial representation of the construction of a Θ-random graph. Θ is defined on $[0,1]^2$ with value indicated by greyscale value. To determine the value of X_{ij} we select coordinates U_1 and U_2 then flip a weighted coin with weight $\Theta\left(U_1, U_2\right)$.

example makes it clear that the choice of uniformly random variables in the definition of Θ-random graph is arbitrary: any atom-free distribution[3] would work equally well.

It is easy to see that any Θ-random graph is exchangeable. The content of the representation theorem is that exchangeability also implies that the graph is conditionally Θ-random:

Theorem 16. *(Aldous, Hoover) Let $X = \left(X_{ij}\right)_{i,j\in\mathbb{N}}$ be the adjacency matrix of an undirected graph on $\mathbb{N}$. The following are equivalent:*

1. *X is jointly exchangeable.*
2. *X is conditionally Θ-random, given a random graphon Θ.*

Example 17. Erdős-Renyi graphs are Θ-random graphs with $\Theta\left(U_i, U_j\right) = p$, a constant function.

Example 18. Let $Y_1, Y_2, \ldots$ be a Pólya urn. Let $\phi : \mathbb{N}^2 \to \mathbb{N}$ be a bijection and let $X_{ij} = Y_{\phi(i,j)}$. Then since the Pólya urn is exchangeable we can conclude that the random graph X_{ij} is jointly exchangeable, and thus by the representation theorem it is conditionally Θ-random. In this case $\Theta\left(U_i, U_j\right) = p$ where $p \sim U\left[0,1\right]$.

We end by noting that in contrast to de Finetti's theorem the random graphon of the Aldous-Hoover representation theorem is not unique. To see this let $T : [0,1] \to [0,1]$ be a transformation such that $T\left(U\right) \overset{d}{=} U$ and let

$$\Theta^T\left(U_1, U_2\right) = \Theta\left(T\left(U_1\right), T\left(U_2\right)\right).$$

It is easy to see that Θ^T-random graph and a Θ-random graph will have the same distribution, thereby ruling out uniqueness.

[3] g is atom free if $X_1 \sim g$, $X_2 \sim g \implies \Pr\left(X_1 = X_2\right) = 0$