On the basic estimation problem for symmetric random graphs and networks

Olav Kallenberg

—

Aarhus, 8/14–16, 2017

1. Graphs and Graphons

A **directed graph** G consists of a countable set S of nodes, and an array of pairwise interactions $x_{ij}: i \to j$, where we allow $x_{ij} \neq x_{ji}$. In a general **network** we may have two sets of nodes, S and T, along with some pairwise interactions x_{ij}, $(i,j) \in S \times T$. To reduce to the previous case, we may replace S by $S \cup T$.

We often assume $x_{ij} \in \{0,1\}$ for all (i,j), where 1 means existence of a (directed) link from i to j. The array $x = (x_{ij})$ is then called the **adjacency matrix**. A **colored graph** is one where the x_{ij} take more general values (=colors).

Random graphs and graphons

In a **random graph**, the interactions x_{ij} are random variables, now denoted by ξ_{ij}. The simplest case is when the ξ_{ij} are **independent with distributions** μ_{ij}. Or, we may take the pairs (ξ_{ij}, ξ_{ji}) to be independent with distributions μ_{ij}, which includes the case of symmetric interactions $\xi_{ij} = \xi_{ji}$.

For on/off interactions, it is enough to specify the probabilities $p_{ij} = P\{\xi_{ij} = 1\}$, but in general we need to specify the entire distributions μ_{ij}. The array $M = (\mu_{ij})$, called the **graphon** of X, clearly determines the distribution of the whole array $X = (\xi_{ij})$.

Conditional graphon models

Now randomize with respect to the graphon $M = (\mu_{ij})$. Thus, even M is considered as random, and we choose the ξ_{ij} to be **conditionally independent**, given M, with **random distributions** μ_{ij}, so that

$$\mathcal{L}(\xi_{ij}|M) = \mu_{ij}, \quad i,j \in S.$$

In other words, we first choose the graphon M at random, and then, given M, we choose an array X with graphon M. Though the distribution $\mathcal{L}(M)$ clearly determines $\mathcal{L}(X)$, **the converse is false** in general. (Thus, estimation of $\mathcal{L}(M)$ makes no sense!)

2. Exchangeable Representations

An random graph or array $X = (\xi_{ij})$ is said to be **jointly exchangeable** if its distribution is invariant under permutations $\pi = (\pi_i)$ of the index set S, in the sense that

$$X \circ \pi^{\otimes 2} = (\xi_{\pi_i,\pi_j};\ i,j \in S) \stackrel{d}{=} X.$$

We may also consider **separately exchangeable** arrays X, where invariance is assumed under possibly different permutations π' and π'' in the two indices:

$$X \circ (\pi' \otimes \pi'') = (\xi_{\pi'_i,\pi''_j};\ i,j \in S) \stackrel{d}{=} X.$$

- An infinite array $X = (\xi_{ij})$ is **jointly exchangeable** iff

$$\xi_{ij} = f(\alpha, \beta_i, \beta_j, \gamma_{ij}), \quad i, j \in S,$$

for some measurable function f on $[0,1]^4$ and some i.i.d. $U(0,1)$ variables α, β_i, and $\gamma_{ij} = \gamma_{ji}$.

- An infinite array X is **separately exchangeable** iff

$$\xi_{ij} = f(\alpha, \beta'_i, \beta''_j, \gamma_{ij}), \quad i, j \in S,$$

for some f as before and some i.i.d. $U(0,1)$ variables α, β'_i, β''_j, γ_{ij}.

Reduction to ergodic case

If we consider only a single graph or network (sample of size 1), we may reduce to the **ergodic** case where α is a constant:

- $X = (\xi_{ij})$ is **ergodic jointly exchangeable** iff

$$\xi_{ij} = f(\beta_i, \beta_j, \gamma_{ij}), \quad i, j \in S,$$

- X is **ergodic separately exchangeable** iff

$$\xi_{ij} = f(\beta'_i, \beta''_j, \gamma_{ij}), \quad i, j \in S,$$

for some function f on $[0,1]^3$ and some i.i.d. $U(0,1)$ random variables β_i, γ_{ij} or β'_i, β''_j, γ_{ij}.

The latter arrays are conditional graphon models with **simple symmetric** arrays of conditional distributions

$$\mu_{ij} = \Phi(\beta_i, \beta_j), \quad i, j \in S,$$

$$\mu_{ij} = \Phi(\beta'_i, \beta''_j), \quad i, j \in S,$$

respectively, where

$$\Phi(x, y) = \mathcal{L}\{f(x, y, \vartheta)\}, \quad x, y \in [0, 1],$$

for a $U(0,1)$ random variable ϑ. In this case, the distributions of $X = (\xi_{ij})$ and $M = (\mu_{ij})$ do **determine each other uniquely**, and it makes sense to estimate $\mathcal{L}(M)$.

Non-uniqueness of representations

The representations of symmetric arrays are not unique: For a **simple, jointly exchangeable** array, we may replace f by

$$g(x, y) = f(T(x), T(y)), \quad x, y \in [0, 1],$$

for any measure-preserving function T. For a **simple, separately exchangeable** array, we may replace f by

$$g(x, y) = f(T_1(x), T_2(y)), \quad x, y \in [0, 1],$$

for some measure-preserving functions T_1 and T_2.

The general equivalence criteria involve additional **randomization variables**. (This is because measure-preserving functions, unlike permutations, are not invertible in general.)

3. Basic Estimation Problem

Estimation problem for exchangeable graphs

Estimate a version of the representation function Φ of the graphon $M = (\mu_{ij})$ from a single observation of the exchangeable array $X = (\xi_{ij})$.

Theorem 1: *For any separately or jointly exchangeable array X, there exist some* **simple** *exchangeable arrays $X_1, X_2, \ldots,$ such that as $n \to \infty$*

$$\mathcal{L}_m(X_n) \to \mathcal{L}_m(X), \quad m \in \mathrm{N},$$

where $\mathcal{L}_m$ denotes the distribution of the $m \times m$ subarray.

Conclusion: Based on observations of finite subarrays, we can't see the difference between simple and more general arrays, and we can just as well assume that X is simple to begin with, which seems to simplify the problem. But then $M = X$, and there is nothing to estimate!

Estimation by grid processes

Consider a real, infinite array $X = (\xi_{ij})$ with $n \times n$ sub-arrays X_n. For each $n \in \mathbb{N}$, divide $[0, 1]$ into sub-intervals I_{nj} of length n^{-1}, and introduce the **grid process**

$$\varphi_n(x, y) = \xi_{ij}, \quad (x, y) \in I_{ni} \times I_{nj},\ i, j \leq n.$$

Theorem 2: *If $X = (\xi_{ij})$ is simple, ergodic, jointly or separately exchangeable with representation*

$$\xi_{ij} = f(\beta_i, \beta_j) \quad \text{or} \quad \xi_{ij} = f(\beta'_i, \beta''_j),$$

then

$$\inf_{f' \sim f} \|\varphi_n - f'\| \xrightarrow{P} 0.$$

Invariant diagonalization

To achieve uniqueness, we may sometimes diagonalize.

Theorem 3: *Let $X = (\xi_{ij})$ be simple, ergodic, and L^2-valued.*

- *When X is symmetric, jointly exchangeable,*

$$f = \sum_k \alpha_k (\varphi_k \otimes \varphi_k).$$

- *When X is separately exchangeable,*

$$f = \sum_k \alpha_k (\varphi_k \otimes \psi_k).$$

Here the eigenvalues α_k are **unique**, *and so are the eigen-functions φ_k and ψ_k, up to suitable rotations.*

No invariant diagonalization seems to exist in the general jointly exchangeable case.

Consider a real-valued, ergodic, jointly exchangeable array $X = (\xi_{ij})$. For any $r \in \mathrm{R}$, form the 0-1 array

$$X_{ij}(r) = 1\{X_{ij} \geq r\}, \quad i, j \in S.$$

Given an observation of the $m \times m$ subarray X^m,

- form the grid processes based on $X(r)$,
- diagonalize and keep only the leading terms,
- estimate by functions that are monotone in r.

Graphon estimation as a filtering problem

The suggested procedure leads to the following **statistical problems**:

- Assuming suitable smoothness of the underlying graphon, choose an **optimal truncation level**, depending on the size m of the subarray.
- Prove that the resulting estimators are **consistent** and converge to the representation function of the graphon.

Note that this is essentially a problem of **optimal filtering**.

References

- Aldous, D.J. (1981). Representations for partially exchangeable arrays of random variables. JMA **11**, 581–598.
- Hoover, D.N. (1979). Relations on probability spaces and arrays of random variables. Preprint, Princeton University.

— — —

- Kallenberg, O. (1989). On the representation theorem for exchangeable arrays. JMA **30**, 137–154.
- — (1999). Multivariate sampling and the estimation problem for exchangeable arrays. JTP **12**, 859–883.
- — (2005). *Probabilistic Symmetries and Invariance Principles*. Springer.