Exchangeable Array and Random Graph

Dongwoo Kim
ANU

April 13, 2016

1 Exchangeable array for graph

Definition 1 (Jointly exchangeable array)**.** A random 2-array $(X_{ij})_{i,j\in\mathbb{N}}$ is jointly exchangeable if

$$(X_{ij}) \stackrel{d}{=} (X_{\pi(i)\pi(j)}) \quad \text{for } i,j \in \mathbb{N}^2 \tag{1}$$

for any permutation π of $\mathbb{N}$.

Theorem 2 (Aldous, Hoover Theorem (Aldous, 1981; Hoover, 1979))**.** A random 2-array $(X_{ij})_{i,j\in\mathbb{N}}$ is jointly exchangeable if and only if there exists a random measurable function $f : [0,1]^3 \to \mathcal{X}$ such that

$$(X_{ij}) \stackrel{d}{=} (f(U_i, U_j, U_{ij})), \tag{2}$$

where $(U_i)_{i\in\mathbb{N}}$ and $(U_{ij})_{ij\in\mathbb{N}^2}$ with $U_{ij} = U_{ij}$ are a sequence and matrix of i.i.d. Uniform[0,1] random variables.

If the function f is a symmetric in its first two argument, then the matrix X is symmetric. In general, however, it does not need to be a symmetric (e.g. directed graph).

For the undirected graph case $\mathcal{X} = \{0,1\}$, the theorem can be simplified further through a random function called *graphon* $W : [0,1]^2 \to [0,1]$, symmetric in its arguments, where

$$f(U_i, U_j, U_{ij}) = \begin{cases} 1 & U_{ij} < W(U_i, U_j) \\ 0 & \text{otherwise} \end{cases}$$

Example 2.1 (Random function prior on function f (Lloyd et al., 2013))**.** Lloyd et al. use a Gaussian process prior to define function f for an undirected graph. They define $W(U_i, U_j) = \phi(\Theta(U_i, U_j))$ where ϕ is a logistic function, and $\Theta(\cdot, \cdot)$ is a continuous function with a Gaussian process prior. Thus, the probability of edge between node i and j is equal to Bernoulli($\phi(\Theta(U_i, U_j))$).

Example 2.2 (Stochastic block model)**.** The stochastic block model is another generative model for random graphs. The model assumes that each node belongs to a community and there is a edge probability between communities. Then the probability of edge between two different nodes are determined by the edge probability between communities where the two nodes belong. In stochastic block model with Aldous, Hoover Theorem, U_i and U_j determine the communities of node i and j, respectively, and U_{ij} determines the existence of edge between node i and j as done in graphon. The stick breaking process is one way to partition unit length interval into an infinite number of communities (Sethuraman, 1994).

Definition 3 (Sparse Graph)**.** Let the number of nodes in a graph be n. The graph is sparse if the number of edges are $o(n^2)$ or dense if the number of edges are $\Theta(n^2)$.

Remark 1 (Graphon is trivially dense)**.** Every graph represented by graphon W are either empty or dense. The asymptotic proportion of edges is $p = \frac{1}{2} \int W(x,y)dxdy$ and the graph is hence either empty ($p = 0$) or dense (since $O(pn^2) = O(n^2)$).

In (Lloyd et al., 2013), the authors place a Gaussian process prior over graphon $W : [0,1]^2 \to \mathbb{R}$ and transform the output through the logistic function to model the edge probability between nodes.

For the undirected graph case where $X_{ij} = X_{ji}$, one can sample upper triangle of the adjacency matrix, and use the same result for the lower triangle. For the directed graph case, however, the adjacency matrix is no longer symmetric, also, by the theorem, both X_{ij} and X_{ji} rely on single parameter U_{ij}. Therefore, one should jointly sample (X_{ij}, X_{ji}) together from three parameters U_i, U_j, U_{ij}, which means X_{ij} and X_{ji} are not conditionally independent. In addition, asymmetric graphon W could be employed for directed random graph. However, (Cai et al., 2015) show that assymetric graphon is inappropriate to impose certain structures on a graph such as the partial ordering and propose a class of priors for directed graphs.

2 Exchangeable array for matrix factorisation

Jointly exchangeable array assume that both rows and columns index the same entities. However, the assumption does not hold for the general matrix where rows and columns represent different entities (e.g. users and items in recommendation system).

Definition 4 (Separately exchangeable array)**.** A random 2-array $(X_{ij})_{i,j \in \mathbb{N}}$ is separately exchangeable if

$$(X_{ij}) \stackrel{d}{=} (X_{\pi(i)\sigma(j)}) \quad \text{for } i,j \in \mathbb{N}^2 \tag{3}$$

for any permutation π and σ of $\mathbb{N}$.

Corollary 4.1 (Separately exchangeable array)**.** A random 2-array (X_{ij}) is separately exchangeable if and only if there exists a random measurable function $f : [0,1]^3 \to \mathcal{X}$ such that

$$(X_{ij}) \stackrel{d}{=} (f(U_i, U_j, U_{ij})), \tag{4}$$

where $(U_i)_{i \in \mathbb{N}}$, $(U_j)_{j \in \mathbb{N}}$, and $(U_{ij})_{ij \in \mathbb{N}^2}$ are i.i.d. Uniform[0,1] random variables.

In the separately exchangeable array case, function f is not a symmetric in its first two arguments.

Example 4.1 (Probabilistic matrix factorisation (PMF) (Salakhutdinov and Mnih, 2008))**.** PMF is one instantiation of Corollary 4.1. Let the generative process of PMF be

$$U_i \sim \mathcal{MN}_d(0, \Sigma_U) \tag{5}$$
$$V_j \sim \mathcal{MN}_d(0, \Sigma_V) \tag{6}$$
$$X_{ij} \sim \mathcal{N}(U_i^\top V_j, \sigma_x) \tag{7}$$

where $\mathcal{MN}_d$ is a d-dimensional zero-mean multivariate normal distribution with covariance Σ. This corresponds to random measurable function $f(U_i, U_j, U_{ij}) = \Phi_1(U_{ij}; \Phi_d(U_i; 0, \Sigma_U)^\top \Phi_d(U_j; 0, \Sigma_U), \sigma_x^2)$, where $\Phi_d(\cdot; \mu, \sigma^2)$ is an inverse-CDF of d-dimensional multivariate function with mean μ and variance σ^2. Note that σ here differs from the permutation notation.

3 Exchangeable array for knowledge graph

A knowledge graph can be represented as a 3-array tensor X_{ijk}, where the first two dimensions represent entities, and the third dimension represent relations. Therefore, X_{ijk} is jointly exchangeable in its first two dimensions, and the third dimension is separately exchangeable to the first two dimensions.

Definition 5 (π-exchangeable array for knowledge graph (Orbanz and Roy, 2015)). A random 3-array $(X_{ijk})_{i,j,k\in\mathbb{N}}$ is jointly exchangeable in its first two dimensions and separately exchangeable with the third dimension if

$$(X_{ijk}) \stackrel{d}{=} (X_{\pi(i)\pi(j)\sigma(k)}) \quad \text{for } i,j,k \in \mathbb{N}^3 \tag{8}$$

for any permutation π and σ of $\mathbb{N}$.

Corollary 5.1 (π-exchangeable array for knowledge graph). A random 3-array X_{ijk} is jointly exchangeable in its first two dimension and separately exchangeable for the third dimension if and only if there exists a random measurable function $f : [0,1]^7 \to \mathcal{X}$ such that

$$(X_{ijk}) \stackrel{d}{=} (f(U_i, U_j, U_k, U_{ij}, U_{ik}, U_{jk}, U_{ijk})), \tag{9}$$

where $(U_i)_{i\in\mathbb{N}}$, $(U_k)_{k\in\mathbb{N}}$, $(U_{ij})_{ij\in\mathbb{N}^2}$, $(U_{ik})_{ik\in\mathbb{N}^2}$, and $(U_{ijk})_{ijk\in\mathbb{N}^3}$ are i.i.d. Uniform[0,1] random variables.

Remark 2 (Compositional model (Gu et al., 2015)). Compositional vector space models cannot be modelled through the exchangeable array theorem because the dependency between triples breaks the exchangeability.

Example 5.1 (Bayesian RESCAL (BRESCAL)). The generative process of BRESCAL with normally distributed output variable X_{ijk} is as follows:

$$e_i, e_j \sim \mathcal{MN}_d(0, \sigma_e^2 I_d) \tag{10}$$

$$R_k \sim \mathcal{MN}_{d^2}(0, \sigma_r^2 I_{d^2}) \tag{11}$$

$$X_{ijk} \sim \mathcal{N}(R_k^\top (e_i \otimes e_j), \sigma_x^2). \tag{12}$$

Again, we can transform uniform random variables U_i, U_j, U_k to e_i, e_j, R_k via the inverse-CDF of zero-mean multivariate normal distribution. Then X_{ijk} is determined by U_{ijk} via the inverse-CDF of normal distribution with mean $R_k^\top (e_i \otimes e_j)$ and variance σ_x^2, i.e. $X_{ijk} = \Phi_1(U_{ijk}; R_k^\top (e_i \otimes e_j), \sigma_x^2)$. The model is invariant w.r.t the changes in the other uniform variables U_{ij}, U_{ik}, and U_{jk}.

4 Sparse exchangeable graph

Caron and Fox (2015) propose a symmetric simple point process on $\mathbb{R}_+^2$ as the edge set of a random graph, and Veitch and Roy (2015) formalise the probabilistic symmetry as joint exchangeability of the point process.

Definition 6 (Jointly exchangeable random measure on $\mathbb{R}^2$). A random measure ξ on $\mathbb{R}^2$ is jointly exchangeable if for every measure preserving transformation f on $\mathbb{R}_+$, we have

$$\xi \circ (f \otimes f)^{-1} \stackrel{d}{=} \xi \tag{13}$$

Theorem 7 (Kallenberg's jointly exchangeable theorem (Kallenberg, 1990, 2005)). A random measure ξ on $\mathbb{R}^2$ is jointly exchangeable iff it has an almost surely representation

$$\xi = \sum_{i,j} f(\alpha, \vartheta_i, \vartheta_j, \zeta_{i,j}) \delta_{\theta_i, \theta_j} \tag{14}$$

$$+ \sum_{j,k} \{g(\alpha, \vartheta_j, \chi_{jk}) \delta_{\theta_j, \sigma_{jk}} + g'(\alpha, \vartheta_j, \chi_{jk}) \delta_{\sigma_{jk}, \theta_j}\} \tag{15}$$

$$+ \sum_{k} \{l(\alpha, \eta_k) \delta_{\rho_k, \rho'_k} + l'(\alpha, \eta_k) \delta_{\rho'_k, \rho_k}\} \tag{16}$$

$$+ \sum_{j} \{h(\alpha, \vartheta_j)(\delta_{\theta_j} \otimes \Lambda) + h'(\alpha, \vartheta_j)(\Lambda \otimes \delta_{\theta_j})\} + \beta \Lambda_D + \gamma \Lambda^2, \tag{17}$$

for some measurable function $f : \mathbb{R}_+^4 \to \mathbb{R}_+$, $g, g' : \mathbb{R}_+^3 \to \mathbb{R}_+$, and $l, l', h, h' : \mathbb{R}_+^2 \to \mathbb{R}_+$, some collection of independent uniformly distributed random variables $(\zeta_{i,j})$ on $[0, 1]$, some independent unit rate Poisson processes $\{(\theta_j, \vartheta_j)\}$ and $\{(\sigma_{ij}, \chi_{ij})\}$ on $\mathbb{R}_+^2$ and $\{(\rho_j, \rho'_j, \eta_j)\}$ on $\mathbb{R}^3$, and some independent set of random variables $\alpha, \beta, \gamma \geq 0$. Λ is Lebesgue measure on $\mathbb{R}_+$, and Λ_D is Lebesgue measure on the diagonal of $\mathbb{R}_+^2$.

If we only focus on an atomic measure, all terms with the Lebesgue measure have measure zero. The atomic measure will be symmetric if f is symmetric in its the second and third arguments, $g = g'$, and $l = l'$. The atomic measure will be simple if f, g, g', l, and l' are $\{0, 1\}$-valued. This random measure is locally finite under certain conditions, and a truncated measure is defined as a Kallenberg exchangeable graph by Veitch and Roy (2015). Like *graphon* in exchangeable random array, they also characterise the space of exchangeable atomic measure in terms of *graphex*.

Theorem 8 (Graphex (Veitch and Roy, 2015)). Let ξ be a random adjacency atomic measure. Then ξ is jointly exchangeable iff it has a.s. representation

$$\xi = \sum_{i,j} \mathbf{1}[W(\alpha, \vartheta_i, \vartheta_j) \leq \zeta_{i,j}] \delta_{\theta_i, \theta_j} \tag{18}$$

$$+ \sum_{j,k} \mathbf{1}[\chi_{jk} \leq S(\alpha, \vartheta_j)](\delta_{\theta_j, \sigma_{jk}} + \delta_{\sigma_{jk}, \theta_j}) \tag{19}$$

$$+ \sum_{k} \mathbf{1}[\eta_k \leq I(\alpha)](\delta_{\rho_k, \rho'_k} + \delta_{\rho'_k, \rho_k}), \tag{20}$$

where $W : \mathbb{R}^3 \to [0, 1]$, $S : \mathbb{R}^2 \to \mathbb{R}$, and $I : \mathbb{R} \to \mathbb{R}$, $(\zeta_{i,j})$ are collection of uniform random variable in $[0,1]$, $\{(\theta_i, \vartheta_j)\}$, $\{(\sigma_{jk}, \chi_{jk})\}$, and $\{(\rho_k, \rho'_k, \eta_k)\}$ are independent unit rate Poisson processes on $\mathbb{R}_+^2$ and $\mathbb{R}_+^3$, and α is an independent random variable on $\mathbb{R}_+$.

Example 8.1 (Caron and Fox (2015)). Caron and Fox (2015) propose a simple point process on $\mathbb{R}^2$ as a product measure of a complete random measure.

$$G_0 = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} \qquad\qquad G_0 \sim \text{CRM}(\rho, \lambda) \tag{21}$$

$$G = \sum_{i,j} n_{ij} \delta_{\theta_i, \theta_j} \qquad\qquad G|G_0 \sim \text{PP}(G_0 \times G_0) \tag{22}$$

$$Z = \sum_{i,j} \min(n_{ij} + n_{ji}, 1) \delta_{\theta_i, \theta_j}, \tag{23}$$

for Lévy intensity measure ρ and λ where ρ is a jump part of the measure. Z can be directly specified as

$$p(z_{ij} = 1|w) = \begin{cases} 1 - \exp(-2w_i w_j) & i \neq j \\ 1 - \exp(-w_i^2) & i = j \end{cases}$$

This model can be reformulated in the Kallenberg's exchangeable theorem. The CRM G_0 can be constructed from a two-dimensional unit-rate Poisson process on $\mathbb{R}_+^2$ by the inverse Lévy method. Let (θ_i, ϑ_i) be a unit rate Poisson process on $\mathbb{R}_+^2$, and ρ^{-1} be the inverse Lévy intensity of ρ. Then $w_i = \rho^{-1}(\vartheta_i)$. The model can be formulated under the representation of Equation 14 as

$$f(\alpha, \vartheta_i, \vartheta_j, \zeta_{i,j}) = \begin{cases} 1 - \exp(-2\rho^{-1}(\vartheta_i)\rho^{-1}(\vartheta_j)) & i \neq j \\ 1 - \exp(\rho^{-1}(\vartheta_i)^2) & i = j \end{cases}$$

and $g, g', l, l' = 0$.

5   Sparse exchangeable bipartite graph

Theorem 9 (Kallenberg's separately exchangeable theorem (Kallenberg, 1990))**.** A random measure ξ on $\mathbb{R}^2$ is separately exchangeable iff it has an almost surely representation

$$\xi = \sum_{i,j} f(\alpha, \vartheta_i, \vartheta'_j, \zeta_{i,j})\delta_{\theta_i, \theta'_j} \tag{24}$$

$$+ \sum_{j,k}\{g(\alpha, \vartheta_j, \chi_{jk})\delta_{\theta_j, \sigma_{jk}} + g'(\alpha, \vartheta'_j, \chi'_{jk})\delta_{\sigma'_{jk}, \theta'_j}\} \tag{25}$$

$$+ \sum_{k} l(\alpha, \eta_k)\delta_{\rho_k, \rho'_k} \tag{26}$$

$$+ \sum_{j}\{h(\alpha, \vartheta_j)(\delta_{\theta_j} \otimes \Lambda) + h'(\alpha, \vartheta'_j)(\Lambda \otimes \delta_{\theta'_j})\} + \gamma\Lambda^2 \tag{27}$$

for some measurable function $f : \mathbb{R}_+^4 \rightarrow \mathbb{R}_+$, $g, g' : \mathbb{R}_+^3 \rightarrow \mathbb{R}_+$, and $l, l', h, h' : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$, some collection of independent uniformly distributed random variables $(\zeta_{i,j})$ on $[0, 1]$, some independent unit rate Poisson processes $\{(\theta_i, \vartheta_i)\}$, $\{(\theta'_j, \vartheta'_j)\}$, $\{(\sigma_{ij}, \chi_{ij})\}$ and $\{(\sigma'_{ij}, \chi'_{ij})\}$ on $\mathbb{R}_+^2$ and $\{(\rho_j, \rho'_j, \eta_j)\}$ on $\mathbb{R}^3$, and some independent set of random variables $\alpha, \gamma \geq 0$. Λ is Lebesgue measure on $\mathbb{R}_+$.

Matrix factorisation for the sparse bipartite graph has been partially discussed in (Caron, 2012), although the formal proof of the sparsity is lacking.

References

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598.

Cai, D., Ackerman, N., and Freer, C. (2015). Priors on exchangeable directed graphs. pages 1–17.

Caron, F. (2012). Bayesian nonparametric models for bipartite graphs. *Advances in Neural Information Processing Systems 25 (Proceedings of NIPS)*, pages 1–9.

Caron, F. and Fox, E. B. (2015). Sparse graphs using exchangeable random measures. pages 1–64.

Gu, K., Miller, J., and Liang, P. (2015). Traversing knowledge graphs in vector space. In *EMNLP*.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ*, 2.

Kallenberg, O. (1990). Exchangeable random measures in the plane. *Journal of Theoretical Probability*.

Kallenberg, O. (2005). *Probabilistic symmetries and invariance principles*. Springer Science & Business Media.

Lloyd, J., Orbanz, P., Ghahramani, Z., and Roy, D. (2013). Random function priors for exchangeable arrays with applications to graphs and relational data. In *Advances in Neural Information Processing Systems 26*, pages 1–9.

Orbanz, P. and Roy, D. M. (2015). Bayesian Models of Graphs, Arrays and Other Exchangeable Random Structures. *TPAMI*, pages 1–25.

Salakhutdinov, R. and Mnih, A. (2008). Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *Proceedings of the 25th International Conference on Machine Learning (2008)*, volume 25, pages 880–887.

Sethuraman, J. (1994). A constructive definition of dirichlet priors. *Statistica sinica*, pages 639–650.

Veitch, V. and Roy, D. M. (2015). The Class of Random Graphs Arising from Exchangeable Random Measures. *arXiv.org*, math.ST.