Exchangeable Matrices and Random Networks

Steffen Lauritzen
University of Oxford

May 31, 2007

Overview of Lectures

1. Exchangeability and de Finetti's Theorem
2. Sufficiency, Partial Exchangeability, and Exponential Families
3. Exchangeable Matrices and Random Networks

Basic references for the series of lectures include Aldous (1985) and Lauritzen (1988). Other references will be given as we go along. For this lecture, Aldous (1981); Diaconis and Freedman (1981); Lauritzen (2003) are particularly relevant.

$X_1, \ldots, X_n, \ldots$ is *exchangeable* if for all $n = 2, 3, \ldots,, \pi \in S(n)$

$$X_1, \ldots, X_n \stackrel{\mathcal{D}}{=} X_{\pi(1)}, \ldots, X_{\pi(n)}.$$

de Finetti (1931):
A binary sequence $X_1, \ldots, X_n, \ldots$ is exchangeable if and only if there exists a distribution function F on $[0, 1]$ such that for all n

$$p(x_1, \ldots, x_n) = \int_0^1 \theta^{t_n}(1-\theta)^{n-t_n}\, dF(\theta),$$

where $t_n = \sum_{i=1}^n x_i$. Further, *F is distribution function of $Y = \bar{X}_\infty$ and, conditionally on $Y = \theta$, $X_1, \ldots, X_n, \ldots$ are i.i.d. with expectation θ.*

$t(x)$ is *summarizing* p if for some ϕ

$$p(x) = \phi(t(x)).$$

For binary variables, $X_1, \ldots, X_n, \ldots$ is exchangeable if and only if for all n

$$P(X_1 = x_1, \ldots, X_n = x_n) = \phi_n(\sum_i x_i).$$

Thus *exchangeability in binary case is equivalent to $t_n = \sum_i x_i$ summarizing* the distribution of $X_1, \ldots, X_n$.

Rephrasing de Finetti–Hewitt–Savage

If a family of distributions for a sequence $X_1, \ldots, X_n, \ldots$ is summarized by the empirical measure, then every distribution in the family is conditionally i.i.d. given the infinitely remote future $\mathcal{T}$ or, equivalently, given the limiting empirical measure M_∞.

Let $t : \mathcal{X} \to \mathcal{S}$ be a semigroup valued statistic i.e. $\mathcal{S}$ has a composition $\oplus$ which satisfies

$$a \oplus b = b \oplus a, \quad (a \oplus b) \oplus c = a \oplus (b \oplus c).$$

The distribution of $X_1, \ldots, X_n$ of is summarized by $t_n(x_1, \ldots, x_n) = t(x_1) \oplus \cdots \oplus t(x_n)$ for all n if and only if $X_1, \ldots, X_n, \ldots$ are conditionally i.i.d. given the tail $\mathcal{T}$ and

$$P(X_i = x \mid \mathcal{T}) = p(x) = p(x \mid \theta) = c(\theta)^{-1} \rho_\theta\{t(x)\}$$

where ρ_θ is a *character* on the semigroup generated by $t(\mathcal{X})$, i.e. an 'exponential function', satisfying

$$\rho_\theta(u)\rho_\theta(v) = \rho(u \oplus v), \quad \rho_\theta(u) \geq 0.$$

Diaconis and Freedman (1980) show for countable $\mathcal{X}$ that *if the distribution of $X_1, \ldots, X_n$ is for all n summarized by*

$$t_n(x_1, \ldots, x_n) = (x_1, \{n_{xy}\}_{x,y \in \mathcal{X}})$$

where n_{xy} are the transition counts:

$$n_{xy} = \#\{i : (x_i, x_{i+1}) = (x, y)\}$$

and the process is recurrent, then it is a mixture of stationary Markov chains.

Similar results true for

$$t_n(x_1, \ldots, x_n) = \{x_1, \oplus_i t(x_i, x_{i+1})\}$$

where $t : \mathcal{X} \times \mathcal{X} \to \mathcal{S}$ is semigroup valued:

The extreme recurrent processes are Markov chains with

$$P(X_{n+1} = y \mid X_n = x) = \rho_\theta\{t(x, y)\} \frac{c_\theta(y)}{c_\theta(x)},$$

where c_θ are eigenvectors with eigenvalue 1 for the matrix $m_{xy} = \rho_\theta\{t(x, y)\}$; see Ressel (1988) for full details.

Clearly, then

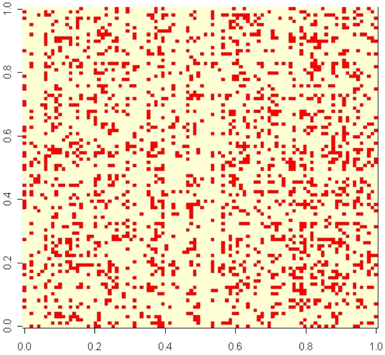
$$p(x_1, \ldots, x_n) = p(x_1)\rho_\theta\{\oplus_i t(x_i, x_{i+1})\} \frac{c_\theta(x_n)}{c_\theta(x_1)}.$$

Rasch model (Rasch, 1960):
Problem i attempted by person j. There are 'easinesses' $\alpha = (\alpha_i)_{i=1,\dots}$ and 'abilities' $\beta = (\beta_j)_{j=1,\dots}$ so that binary responses X_{ij} are conditionally independent given (α, β) and

$$P(X_{ij} = 1 \mid \alpha, \beta) = 1 - P(X_{ij} = 0 \mid \alpha, \beta) = \frac{\alpha_i \beta_j}{1 + \alpha_i \beta_j}.$$

A *random Rasch matrix* has (α_i) i.i.d. with distribution A and (β_j) i.i.d. B.
Also potential model for *hit of batter i against pitcher j, occurrence of species i on island j,* etc.

Example of random Rasch matrix



A doubly infinite matrix $X = \{X_{ij}\}_{1,1}^{\infty,\infty}$ is said to be

- *row–column exchangeable* (RCE-matrix) if for all m, n, $\pi \in S(m), \rho \in S(n)$

$$\{X_{ij}\}_{1,1}^{m,n} \stackrel{\mathcal{D}}{=} \{X_{\pi(i)\rho(j)}\}_{1,1}^{m,n}.$$

- *weakly exchangeable* (WE-matrix) if for all n and $\pi \in S(n)$

$$\{X_{ij}\}_{1,1}^{n,n} \stackrel{\mathcal{D}}{=} \{X_{\pi(i)\pi(j)}\}_{1,1}^{n,n}.$$

A doubly infinite (binary) matrix $X = \{X_{ij}\}_{1,1}^{\infty,\infty}$ is said to be *row-column summarized* (RCS-matrix) if for all m, n

$$p(\{x_{ij}\}_{1,1}^{m,n}) = \phi_{m,n}\{R_1, \ldots, R_m; C_1, \ldots, C_n\},$$

where $R_i = \sum_j x_{ij}$ and $C_j = \sum_j x_{ij}$ are the row- and column sums. Note that, in contrast to the case of binary sequences, *RCE-matrices are generally not RCS-matrices and vice versa* because group G_{RC} of row and column permutations does *not* act transitively on matrices with fixed row- and column sums:

If a matrix is both RCE and RCS, it is an *RCES-matrix.*

RCE versus RCS

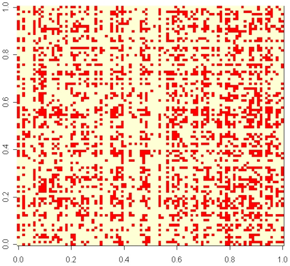
$$M_1 = \left\{\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right\}, \quad M_2 = \left\{\begin{array}{ccc} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{array}\right\}$$

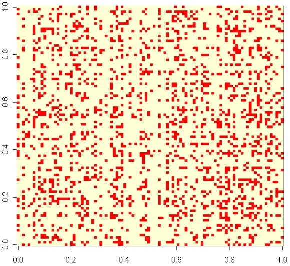
$$M_3 = \left\{\begin{array}{ccc} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{array}\right\}, \quad M_4 = \left\{\begin{array}{ccc} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{array}\right\}$$

$$M_5 = \left\{\begin{array}{ccc} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{array}\right\}$$

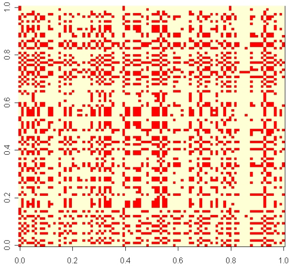
$$|\det M_1| = |\det M_2| = |\det M_3| = |\det M_4| = 1, |\det M_5| = 0.$$

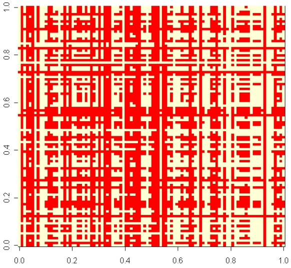
RCE versus RCE and RCS (RCES)





RCE versus RCES





Weakly summarized matrices

A doubly infinite (binary) matrix $X = \{X_{ij}\}_{1,1}^{\infty,\infty}$ is *weakly summarized* (WS-matrix) if for all n

$$p(\{x_{ij}\}_{1,1}^{n,n}) = \phi_n\{R_1 + C_1, \ldots, R_n + C_n\},$$

where $R_i = \sum_j x_{ij}$ and $C_j = \sum_j x_{ij}$ are the row- and column sums as before.

Also here *WE-matrices are generally not WS-matrices and vice versa.*

If a matrix is both WE and WS, it is an *WES-matrix.*

If in addition, $\{X_{ij} = X_{ji}\}$, i.e. the matrix is *symmetric*, we may consider SWE, SWS, SWES matrices, etc.

$$M_6 = \begin{Bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{Bmatrix}$$

$$M_7 = \begin{Bmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{Bmatrix}$$

No joint permutation of rows and columns take M_6 into M_7: M_6 is adjacency matrix of two triangles and M_7 adjacency matrix of 6-cycle.

The set of distributions $\mathcal{P}_{RCE}$ is a convex simplex.
In particular, *every $P \in \mathcal{P}_{RCE}$* has a unique representation as a mixture of extreme points $\mathcal{E}_{RCE}$, i.e.

$$P(A) = \int_{\mathcal{E}} Q(A)\mu_P(Q).$$

The same holds if RCE is replaced by RCS, RCES, WE, SWE, SWES, etc. In addition, it can be shown that

$$\mathcal{E}_{RCES} = \mathcal{E}_{RCE} \cap \mathcal{P}_{RCS}, \quad \mathcal{E}_{WES} = \mathcal{E}_{WE} \cap \mathcal{P}_{WS},$$

etc.

Features of extreme measures

Aldous (1981): *for any $P \in \mathcal{P}_{RCE}$ the following are equivalent:*

- *$P \in \mathcal{E}_{RCE}$*
- *The tail σ-field $\mathcal{T}$ is trivial*
- *The corresponding RCE-matrix X is dissociated.*

Here the *tail* $\mathcal{T}$ is $\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma\{X_{ij}, \min(i,j) \geq n\}$ and a matrix is *dissociated* if for all A_1, A_2, B_1, B_2 with $A_1 \cap A_2 = B_1 \cap B_2 = \emptyset$

$$\{X_{ij}\}_{i \in A_1, j \in B_1} \perp\!\!\!\perp \{X_{ij}\}_{i \in A_2, j \in B_2}.$$

A binary doubly infinite random matrix X is a *ϕ-matrix* if X_{ij} are independent given $U = (U_i)_{i=1,\dots}$ and $V = (V_j)_{j=1,\dots}$ where U_i and V_j are independent and uniform on $(0, 1)$ and

$$P(X_{ij} = 1 \mid U = u, V = v) = \phi(u_i, v_j),$$

Aldous (1981); Diaconis and Freedman (1981) show that *distributions of ϕ-matrices are the extreme points of $\mathcal{P}_{RCS}$*, i.e. binary RCE matrices are mixtures of ϕ-matrices.
Many ϕ give same distribution of ϕ-matrix.

RCE versus RCS

Consider ϕ-matrix defined by $\phi(u_i, v_j) = u_i v_j$. Then

$$P(M_1) = P(M_2) = P(M_3) = P(M_4) = \frac{665}{2985984}$$

whereas $P(M_5) = 1/4096$. $(665 \times 4096 = 2723840)$
RCE matrices have no simple summarizing statistics whereas *RCES-matrices are summarized by the empirical distributions of row- and column sums:*

$$t_{mn} = \left(\sum_{i=1}^{m} \delta_{r_i}, \sum_{j=1}^{n} \delta_{s_j} \right).$$

This is a semigroup statistic, and RCES matrices can be represented via mixtures of characters on the image semigroup (Ressel 2002, personal communication).

Rasch type ϕ-matrices

If a ϕ-matrix is RCS it must satisfy

$$P\left(\left\{\begin{array}{cc}1 & 0\\ 0 & 1\end{array}\right\} \mid U, V\right) = P\left(\left\{\begin{array}{cc}0 & 1\\ 1 & 0\end{array}\right\} \mid U, V\right).$$

This holds if ϕ is of *Rasch type,* i.e. if for all u, v, u^*, v^*:

$$\begin{aligned}\phi(u,v)\bar{\phi}(u,v^*)\bar{\phi}(u^*,v)\phi(u^*,v^*) =\\ \bar{\phi}(u,v)\phi(u,v^*)\phi(u^*,v)\bar{\phi}(u^*,v^*),\end{aligned}$$

where we have let $\bar{\phi} = 1 - \phi$. Above is *Rasch functional equation*. General solutions of this equation represent characters of the image semigroup of the empirical row- and column sum measures.

de Finetti for RCES

Lauritzen (2003): *Any RCES matrix is a mixture of Rasch type ϕ-matrices.*

A random binary matrix is *regular* if

$$0 < P(X_{ij} = 1 \mid \mathcal{S}) < 1 \text{ for all } i, j,$$

where the *shell σ-algebra* $\mathcal{S}$ is

$$\mathcal{S} = \bigcap_{n=1}^{\infty} \sigma\{X_{ij}, \max(i,j) \geq n\}.$$

Any regular RCES matrix is a mixture of random Rasch matrices.

Solutions to Rasch functional equation

Regular solutions ($0 < \phi < 1$) all of form

$$\phi(u, v) = \frac{a(u)b(v)}{1 + a(u)b(v)}$$

leading to random Rasch models.
Regular random Rasch matrices are parametrized by distributions (A, B) of $a(U)$ and $b(V)$, up to multiplication of a and division of b with constant.

$$(A, B\ ) \sim (A', B') \iff A'(x) = A(cx), B'(y) = B(y/c)$$

for some $c > 0$.

Non-regular solutions to Rasch equation

There are other interesting solutions, e.g.

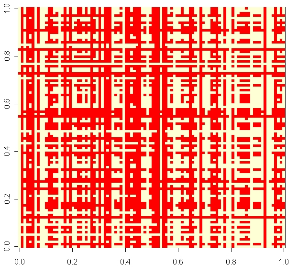
$$\phi(u, v) = \chi_{\{u \leq v\}} = \begin{cases} 1 & \text{if } u \leq v \\ 0 & \text{otherwise.} \end{cases}$$

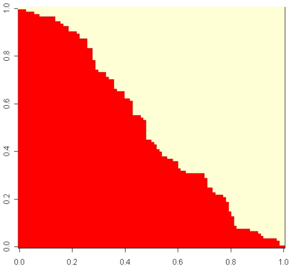
or

$$\phi(u, v) = \begin{cases} \dfrac{a(u)b(v)}{1 + a(u)b(v)} & \text{if } 1/3 < u, v < 2/3 \\ \chi_{\{u \leq v\}} & \text{otherwise} \end{cases}$$

corresponding to incomparable groups.

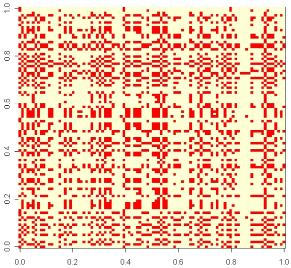
Non-regular Rasch with sorted rows and columns

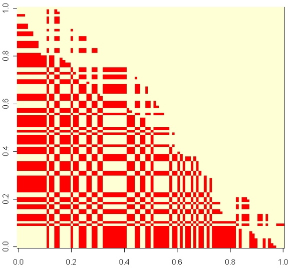




$$\phi(u, v) = \chi_{\{u \leq v\}}$$

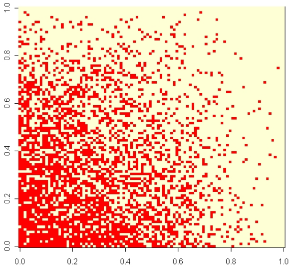
Non-regular RCE with sorted rows and columns





$$\phi(u, v) = \chi_{\{|u-v|\leq 1/2\}}$$

RCE vs Rasch with sorted rows and columns



$$\phi(u,v) = uv, \qquad \phi(u,v) = uv/(1+uv).$$

de Finetti for WE matrices

A binary doubly infinite random matrix X is a *ψ-matrix* if $X_{\{i,j\}}$ are all independent given $U = (U_i)_{i=1,\dots}$ where U_i are mutually independent and uniform on $(0,1)$ and

$$P(X_{\{i,j\}}) = (y,z) \mid U = u, V = v) = \psi_{yz}(u_i, u_j).$$

Here we have let $X_{\{i,j\}} = (X_{ij}, X_{ji})$ for $i < j$.
Reformulating results in Aldous (1981) yield that *binary WE matrices are mixtures of ψ-matrices.*
Note that we may further impose *full symmetry* by restricting to $\psi_{yz} = 0$ unless $y = z$ and *distributional symmetry* by assuming $\psi_{yz} = \psi_{zy}$ or, equivalently, $\psi_{yz}(u,v) = \psi_{yz}(v,u)$.

Regular SWES matrices

Exactly as before, it is easy to show that $\mathcal{E}_{SWES} = \mathcal{E}_{SWE} \cap \mathcal{P}_{SWS}$, implying that SWES matrices are mixtures of ψ-matrices where ψ satisfies the Rasch functional equation.

Hence *regular SWES ψ-matrices are generated as*

$$\psi(u, v) = \frac{a(u)a(v)}{1 + a(u)a(v)}.$$

Probably no interesting non-regular solutions?

A binary matrix X defines a random graph in several ways.

If we consider the rows and colums as labels of two different sets of vertices, a *random bipartite* graph can be defined from X by letting $X_{ij} = 1$ if and only if there is a directed edge from i to j.

An RCE-matrix then corresponds to a *random graph with exhangeable labels within each partition* of the graph vertices.

An RCS-matrix is similarly one where *any two graphs having the same in-degree and out-degree for every vertex are equally likely*.

If we consider the row-and column numbers to label the *same vertex set*, the matrix X represents in a similar way a random graph.
The graph is in general directed, but if we further restrict the matrix X to be symmetric, X can represent a random *undirected graph*.
A WE-matrix now represents *a random graph with exchangeable labels,* and an SWE-matrix similarly an undirected random exchangeable graph.
An SWS-matrix represents a random graph with the probability of any graph *only depending on its vertex degrees.*

Random graphs with exchangeability properties form natural models for *social networks*.

Frank and Strauss (1986) consider *Markov graphs* which are random graphs with

$$X_{\{i,j\}} \perp\!\!\!\perp X_{\{k,l\}} \mid X_{E \setminus \{\{i,j\},\{k,l\}\}} \tag{1}$$

whenever all indices i, j, k, l are different. Here E denotes the edges in the complete graph on $\{1, \ldots, n\}$.

They show that *weakly exchangeable Markov graphs* all have the form

$$p(\{x_{ij}\}_{1,1}^{n,n}) \propto \exp\{\tau_n t(x) + \sum_{k=1}^{n-1} \delta_{nk}\nu_k(x)\}$$

where $x = \{x_{ij}\}_{1,1}^{n,n}$, $t(x)$ is the number of triangles in x, and $\nu_k(x)$ is the number of vertices in x of degree k.
Such *Markov graphs are SWE*, but *generally not extendable as such.*
They are SWES if $\tau = 0$, and not otherwise if $n > 5$.

Note that ψ-matrices typically differ from Markov graphs in that they are *dissociated*, hence marginally rather than conditionally independent:

$$X_{\{i,j\}} \perp\!\!\!\perp X_{\{k,l\}} \tag{2}$$

whenever all indices i, j, k, l are different.

In fact *infinite weakly exchangeable Markov graphs are Bernoulli graphs,* essentially because the conjunction of (1) and (2) implies complete independence.

Problem: *characterize exchangeable random graphs* which for every n also are summarized by the number of triangles and the empirical distribution of vertex degrees:

$$p(\{x_{ij}\}_{1,1}^{n,n}) = f_n(t(x), \sum_{k=1}^{n} \delta_{r_k(x)})$$

or similar graphs with sufficient statistics being counts of specific types of subgraph.
Rasch-type graphs, i.e. regular SWES-matrices, are as above, but without triangles.

Summary

- RCES matrices are mixtures of ϕ-matrices of Rasch type
- Regular RCES matrices are mixtures of random Rasch matrices
- Non-regular RCES matrices can be natural and interesting
- RCE, RCES, WE and SWE, matrices may produce possibly interesting random graphs, in particular in combination with other types of summarizing statistics.

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

Aldous, D. (1985). Exchangeability and related topics. In Hennequin, P., editor, *École d'Été de Probabilités de Saint–Flour XIII — 1983*, pages 1–198. Springer-Verlag, Heidelberg. Lecture Notes in Mathematics 1117.

de Finetti, B. (1931). Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Academia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Mathematice e Naturale*, 4:251–299.

Diaconis, P. and Freedman, D. (1980). de Finetti's theorem for Markov chains. *Annals of Probability*, 8:115–130.

Diaconis, P. and Freedman, D. (1981). On the statistics of vision: the Julesz conjecture. *Journal of Mathematical Psychology*, 24:112–138.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81:832–842.

Lauritzen, S. L. (1988). *Extremal Families and Systems of Sufficient Statistics*. Springer-Verlag, Heidelberg. Lecture Notes in Statistics 49.

Lauritzen, S. L. (2003). Rasch models with exchangeable rows and columns. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 7*, pages 215–232, Oxford. Clarendon Press.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*, volume 1 of *Studies in Mathematical Psychology*. Danmarks Pædagogiske Institut, Copenhagen.

Ressel, P. (1988). Integral representations for distributions of symmetric stochastic processes. *Probability Theory and Related Fields*, 79:451–467.