Exchangeable Rasch Models

Steffen L. Lauritzen
University of Oxford

De Finetti 100, Rome 2006

Random Rasch matrices

Rasch model (1960):

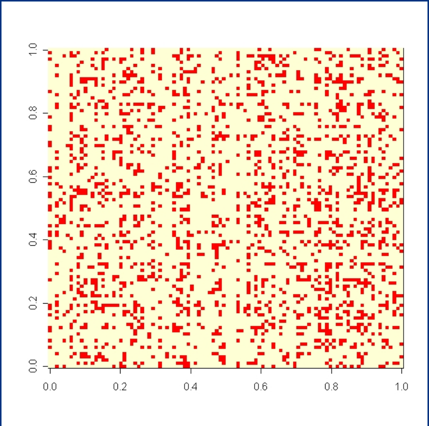
Problem i attempted by person j. There are 'easinesses' $\alpha = (\alpha_i)_{i=1,\ldots}$ and 'abilities' $\beta = (\beta_j)_{j=1,\ldots}$ so that binary responses X_{ij} are conditionally independent given (α, β) and

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

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



Exchangeable sequences

$X_1, \ldots, X_n, \ldots$ is *exchangeable* if for all n

$$X_1, \ldots, X_n \overset{\mathcal{D}}{=} X_{\pi(1)}, \ldots, X_{\pi(n)} \text{ for all } \pi \in S(n).$$

For example:

$$p(1,1,0,0,0,1,1,0) = p(1,0,1,0,1,0,0,1).$$

If $X_1, \ldots, X_n, \ldots$ are independent and identically distributed, they are exchangeable, but not conversely.

de Finetti's Theorem

de Finetti (1931) shows that all exchangeable sequences are mixtures of Bernoulli sequences:

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 $p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$ and $t_n = \sum_{i=1}^n x_i$.

More about de Finetti's Theorem

It further holds that *F is the distribution function of the limiting frequency:*

$$Y = \lim_{n\to\infty} \sum_i X_i/n, \quad P(Y \leq y) = F(y)$$

and *the Bernoulli distribution is obtained by conditioning with $Y = \theta$:*

$$P(X_1 = x_1, \ldots, X_n = x_n \,|\, Y = \theta) = \theta^{t_n}(1-\theta)^{n-t_n}.$$

Exchangeability and sufficiency

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

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

Because *$S(n)$ acts transitively on binary n-vectors with fixed sum,* i.e. if x and y are two such vectors, there is a permutation which sends x into y.

So *exchangeability is equivalent to $t_n = \sum_i x_i$ being sufficient and*

$$p(x_1, \ldots, x_n \,|\, t_n) = \binom{n}{t_n}^{-1}.$$

Summarizing statistics

We say that $t(x)$ is *summarizing* for p if $p(x) = \phi(t(x))$ for some ϕ.

Note that if $t(x)$ *is summarizing, it is sufficient* and

$$p(x \,|\, t) \text{ is uniform on } \{x : t(x) = t\}$$

So *exchangeability is equivalent to* $t_n = \sum_i x_i$ *summarizing* the probability.

Often $t(x)$ takes values in an Abelian semigroup, generally leading to mixture representation of all distributions summarized by t in terms of the *characters* of the semigroup, i.e. functions satisfying $\rho(s+t) = \rho(s)\rho(t)$.

Row- and column-exchangeable matrices

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 $\pi \in S(n)$

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

Summarized matrices

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.*

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

RCE versus RCS

Group G_{RC} of row and column permutations does *not* act transitively on matrices with fixed row- and column sums:

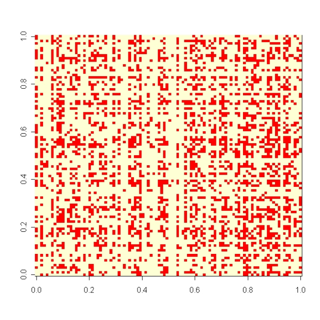
$$M_1 = \left\{ \begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{array} \right\}, \quad M_2 = \left\{ \begin{array}{ccc} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{array} \right\}$$

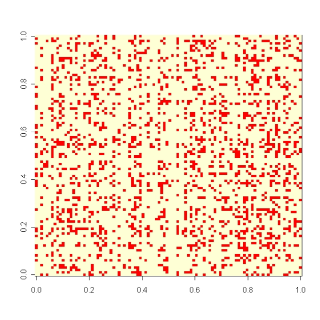
$$M_3 = \left\{ \begin{array}{ccc} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{array} \right\}, \quad M_4 = \left\{ \begin{array}{ccc} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{array} \right\}$$

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

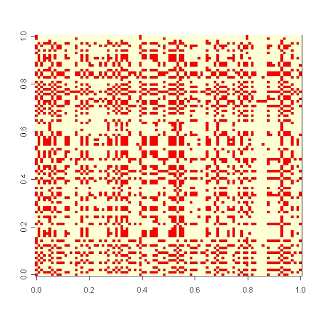
$$|\det M_1| = |\det M_2| = |\det M_3| = |\det M_4| = 1, |\det M_5| = 0.$$

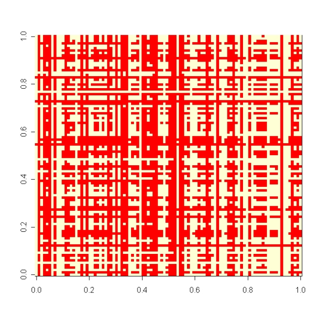
RCE versus RCE and RCS (RCES)





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

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

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.

(S)WE versus (S)WS

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

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

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.

Convexity formulation

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 (1978,1981): *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 $A_1 \cap A_2 = B_1 \cap B_2 = \emptyset$

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

Random bipartite graphs

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*.

Exchangeable random graphs

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.*

de Finetti for RCE matrices

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

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

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 indeed a semigroup statistic, so RCES matrices can be represented as mixtures of characters on the image semigroup.

Rasch type ϕ-matrices

If a ϕ-matrix is RCS it must satisfy

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

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

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

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

Any RCES matrix is a mixture of Rasch type ϕ-matrices.

A random binary matrix is *regular* if

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

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.

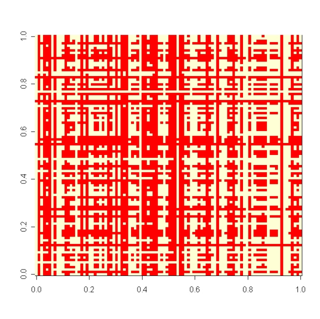
$$\phi(u,v) = \chi_{\{u \le v\}} = \begin{cases} 1 & \text{if } u \le v \\ 0 & \text{otherwise.} \end{cases}$$

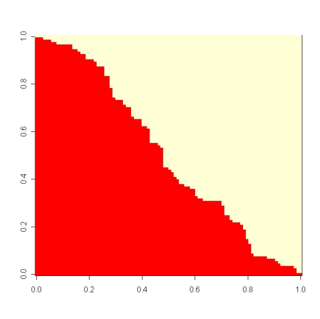
or

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

corresponding to incomparable groups.

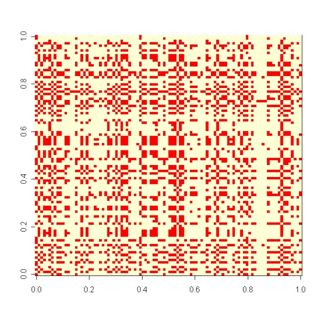
Non-regular Rasch with sorted rows and columns

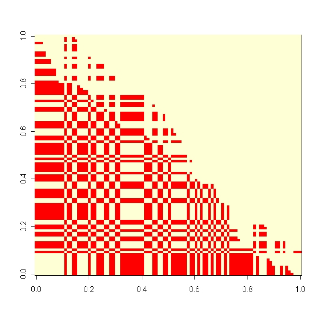




$$\phi(u,v) = \chi_{\{u \leq v\}}$$

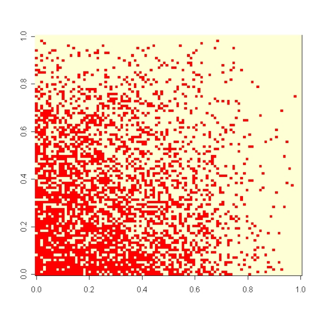
Non-regular RCE with sorted rows and columns

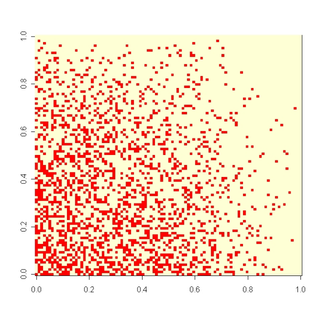




$$\phi(u,v) = \chi_{\{|u-v|\leq 1/2\}}$$

RCE vs Rasch with sorted rows and columns





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

Cantor–Rasch matrices

Rasch model prevails between comparable groups, determinism between incomparable ones: Keep cutting out middle thirds of the unit interval to get

$$\phi(u,v) = \begin{cases} \dfrac{a(u)b(v)}{1+a(u)b(v)} & \text{if } 1/9 < u,v < 2/9 \\ \dfrac{a(u)b(v)}{1+a(u)b(v)} & \text{if } 1/3 < u,v < 2/3 \\ \dfrac{a(u)b(v)}{1+a(u)b(v)} & \text{if } 7/9 < u,v < 8/9 \\ \chi_{\{u \le v\}} & \text{otherwise} \end{cases},$$

and so on.

General results of Ressel imply that the limit will correspond to a ϕ-matrix.

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

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

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?

Social network analysis

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

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

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.

Exchangeable random graphs

Problem: *characterize exchangeable random graphs* of the form

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

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.

Limiting behaviour of RCES matrices

Second half of de Finetti's Theorem relates parameter to limiting frequency behaviour. For RCES-matrices we have a clear analogue:

Let (F_m, G_n) denote the pair of empirical distributions of the row- and column- averages $\bar{X}_{i+} = R_i/n, \bar{X}_{+j} = C_j/m$.

Consistency demands (F_m, G_n) to be a *subconjugate pair*, i.e. that $F_m \preceq G_n^*$ where

$$F \preceq G^* \iff \int_0^s F(x)\,dx \leq \int_0^s G^*(x)\,dx \text{ for all } s \in [0,1],$$

where

$$G^*(x) = 1 - G^{-1}(1-x).$$

Note that in fact $F \preceq G^* \implies G \preceq F^*$

Since (F_m, G_n) are both distributions on the unit interval, it easily follows that *any RCES-matrix has an a.s. limit:*

$$\lim_{m,n\to\infty}(F_m, G_n) = (F, G).$$

For ϕ-matrices this limit is a degenerate random variable and plays the role of θ in the standard deFinetti's theorem.

Clearly, *also (F, G) must then be subconjugate pairs.*

Note that *the pair (F, G) plays the role of θ* in the standard deFinetti theorem, being the limiting value of the sufficient statistic.

Marginal problem

If we consider models for batters vs. pitchers, F may determine the distribution of the batting average of a random batter U and G the average result of a random pitcher V.

With the probability of a hit when this batter meets this pitcher defined by $Y = \gamma(U, V)$, consistency implies that

$$\mathbf{E}(Y \mid U) = U, \quad \mathbf{E}(Y \mid V) = V.$$

Using results of Guttmann et al. (1991), it can be shown (Lauritzen 2003) that such *a function γ exists if and only if (F, G) is a subconjugate pair.*

A conjecture concerning random Rasch matrices

Conjecture: *Let (F, G) be pair of cdfs on $[0, 1]$. Then there is a Rasch ϕ-matrix with limits of marginal averages having distributions F and G if and only if $F \preceq G^*$.*

Distributions of ϕ-matrices are injectively parametrized by (F, G).

The ϕ-matrix is regular if and only if $F \prec G^*$.

True if (F, G) are discrete with support on rational points!

A different formulation of the conjecture says that *the extreme points $\mathcal{E}_{RCES}$ can be identified with the set of subconjugate pairs (F, G).*

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 produce possibly interesting random graphs.
- Rasch type ϕ-matrices are (probably) parametrized by subconjugate pairs (F, G) of distributions of limiting marginal averages. Regular by strictly subconjugate.