Some statistical applications of constrained Monte Carlo

JULIAN BESAG

Department of Mathematical Sciences, University of Bath, England

Department of Statistics, University of Washington, Seattle, USA

LMS, Durham, Saturday 5th July, 2008

Agenda

A statistician plays Sudoku

Simple Monte Carlo p–values
Examples

Markov chain Monte Carlo (MCMC) p–values
Examples

Bayesian inference for incomplete $2 \times J \times K$ tables

MCMC for p–values in higher–dimensional contingency tables
Examples

Attaining irreducibility in constrained sample spaces
Example

Social networks: Markov random graphs: MCMC p–values
Example

A statistician plays Sudoku

The Times : daily Sudoku puzzle (“fiendish” $|\mathcal{S}| \approx 2 \times 10^{25}$)

Initial configuration

0	9	0	7	0	0	8	6	0
0	3	1	0	0	5	0	2	0
8	0	6	0	0	0	0	0	0
0	0	7	0	5	0	0	0	6
0	0	0	3	0	7	0	0	0
5	0	0	0	1	0	7	0	0
0	0	0	0	0	0	1	0	9
0	2	0	6	0	0	3	5	0
0	5	4	0	0	8	0	7	0

Eventual solution

2	9	5	7	4	3	8	6	1
4	3	1	8	6	5	9	2	7
8	7	6	1	9	2	5	4	3
3	8	7	4	5	9	2	1	6
6	1	2	3	8	7	4	9	5
5	4	9	2	1	6	7	3	8
7	6	3	5	2	4	1	8	9
9	2	8	6	7	1	3	5	4
1	5	4	9	3	8	6	7	2

Metropolis algorithm for solving Sudoku

Sample space : $\mathcal{S}$ = set of all 9×9 tables $\mathbf{x}$ with feasible 3×3 subtables.

$v(\mathbf{x})$ = number of like–like pairs among the rows and among the columns of $\mathbf{x} \in \mathcal{S}$.

Target distribution : $\pi(\mathbf{x}) \propto \exp\{-\beta v(\mathbf{x})\}, \qquad \mathbf{x} \in \mathcal{S},$

where β is a positive constant ($\beta = 3$ recommended).

Hence, if **Sudoku solutions** exist, they are **modes** of $\pi(\mathbf{x})$, with $v(\mathbf{x}) = 0$.

Algorithm : Initialize s.t. each subtable contains $1, \ldots, 9$. Then continually

Choose one of the nine subtables at random.

Select two of its flexible elements at random.

Propose swapping the two elements.

Accept or reject the swap according to Metropolis ratio.

Terminate the algorithm when a solution $v(\mathbf{x}) = 0$ is reached.

Metropolis algorithm

Sample space : $\mathcal{S}$ = set of all 9×9 tables $\mathbf{x}$ with feasible 3×3 subtables.

Given partial configuration

0	9	0	7	0	0	8	6	0
0	3	1	0	0	5	0	2	0
8	0	6	0	0	0	0	0	0
0	0	7	0	5	0	0	0	6
0	0	0	3	0	7	0	0	0
5	0	0	0	1	0	7	0	0
0	0	0	0	0	0	1	0	9
0	2	0	6	0	0	3	5	0
0	5	4	0	0	8	0	7	0

Feasible initial subtables

2	9	4	7	1	2	8	6	1
5	3	1	3	4	5	3	2	4
8	7	6	6	8	9	5	7	9
1	2	7	2	5	4	1	2	6
3	4	6	3	6	7	3	4	5
5	8	9	8	1	9	7	8	9
1	3	6	1	2	3	1	2	9
7	2	8	6	4	5	3	5	4
9	5	4	7	9	8	6	7	8

On the way . . .

2	9	5	7	3	1	8	6	4
4	3	1	8	6	5	9	2	7
8	7	6	2	4	9	1	3	5
3	1	7	9	5	2	4	8	6
9	6	4	3	8	7	5	1	2
5	8	2	4	1	6	7	9	3
6	7	8	5	2	3	1	4	9
1	2	9	6	7	4	3	5	8
3	5	4	1	9	8	2	7	6

Eventual solution

2	9	5	7	4	3	8	6	1
4	3	1	8	6	5	9	2	7
8	7	6	1	9	2	5	4	3
3	8	7	4	5	9	2	1	6
6	1	2	3	8	7	4	9	5
5	4	9	2	1	6	7	3	8
7	6	3	5	2	4	1	8	9
9	2	8	6	7	1	3	5	4
1	5	4	9	3	8	6	7	2

Gibbs sampler for generating large $m \times m$ Latin squares

Sample space : $\mathcal{S} = \{\text{all } m \times m \text{ arrays } \mathbf{x} : x_{ij} \in 1, 2, \ldots, m\}$.

$v(\mathbf{x})$ = number of like–like pairs among the rows and among the columns of $\mathbf{x} \in \mathcal{S}$.

Target distribution : $\pi(\mathbf{x}) \propto \exp\{-\beta v(\mathbf{x})\}, \quad \mathbf{x} \in \mathcal{S}$,

where β is a positive constant.

Hence, **Latin squares** are **modes** of $\pi(\mathbf{x})$, with $v(\mathbf{x}) = 0$.

Algorithm : Initialize table by completely random $\mathbf{x} \in \mathcal{S}$. Then continually

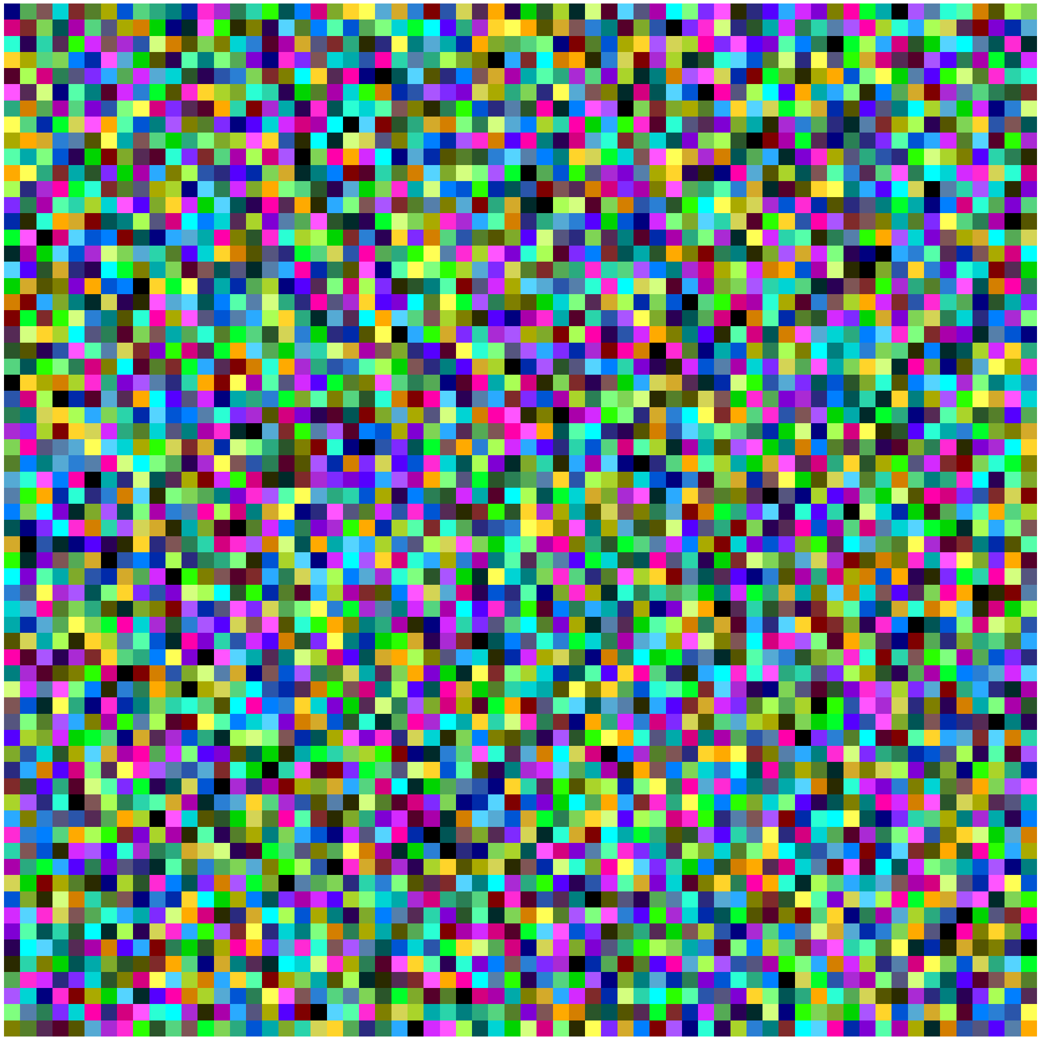
Choose a single cell (i, j) completely at random.

Update its value x_{ij} according to the full conditional at (i, j).

Terminate the algorithm at first table with $v(\mathbf{x}) = 0$.

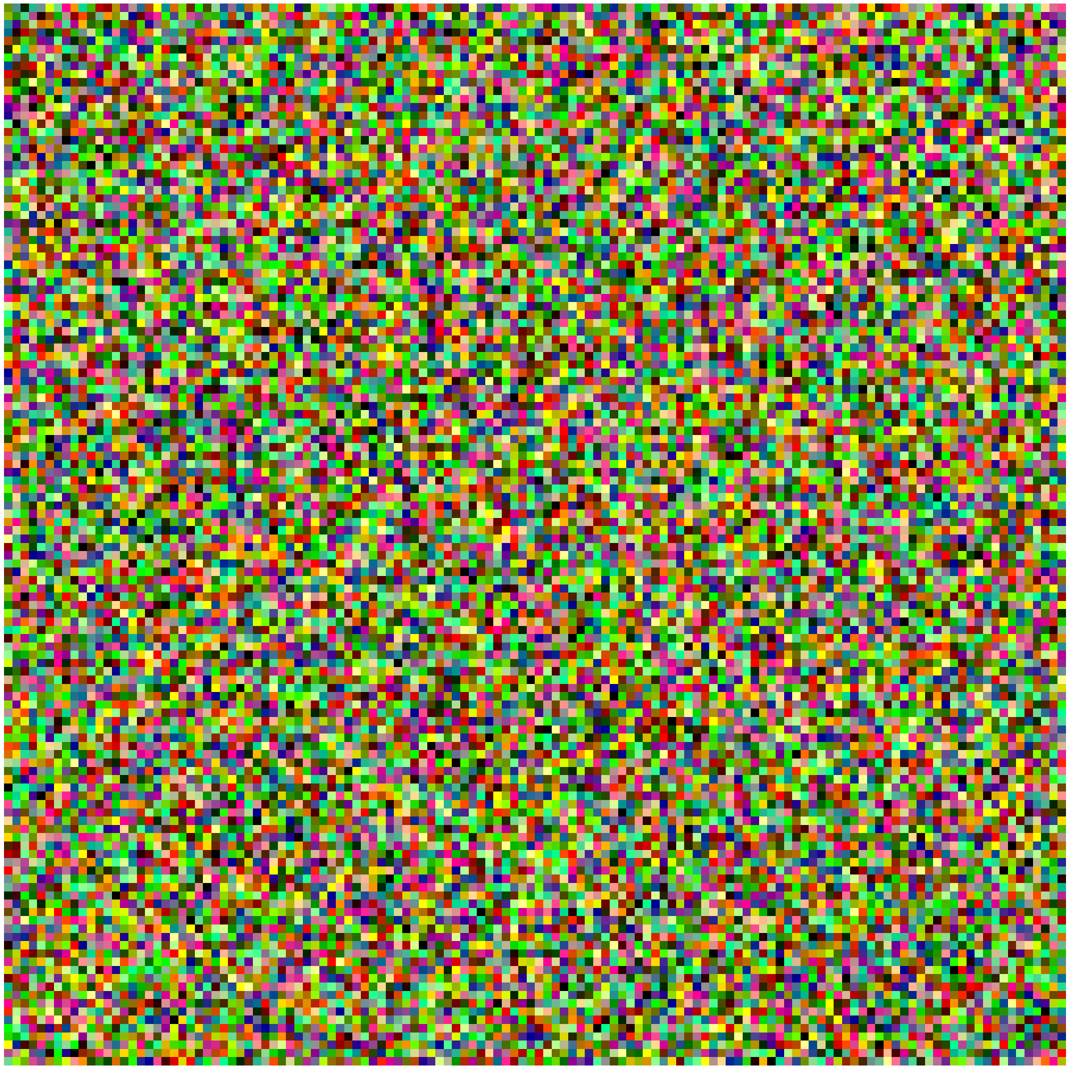
Eventual $\mathbf{x}$ sampled uniformly at random from $m \times m$ Latin squares?? **NO !!**

Constructing large Latin squares



64×64 **Latin square**

Constructing large Latin squares



128×128 **Latin square**

Ex. Deaths by horsekicks in the Prussian Army

Corps identifier

Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
1875	0	0	0	0	0	0	0	1	1	0	0	0	1	0	3
1876	2	0	0	0	1	0	0	0	0	0	0	0	1	1	5
1877	2	0	0	0	0	0	1	1	0	0	1	0	2	0	7
1878	1	2	2	1	1	0	0	0	0	0	1	0	1	0	9
1879	0	0	0	1	1	2	2	0	1	0	0	2	1	0	10
1880	0	3	2	1	1	1	0	0	0	2	1	4	3	0	18
1881	1	0	0	2	1	0	0	1	0	1	0	0	0	0	6
1882	1	2	0	0	0	0	1	0	1	1	2	1	4	1	14
1883	0	0	1	2	0	1	2	1	0	1	0	3	0	0	11
1884	3	0	1	0	0	0	0	1	0	0	2	0	1	1	9
1885	0	0	0	0	0	0	1	0	0	2	0	1	0	1	5
1886	2	1	0	0	1	1	1	0	0	1	0	1	3	0	11
1887	1	1	2	1	0	0	3	2	1	1	0	1	2	0	15
1888	0	1	1	0	0	1	1	0	0	0	0	1	1	0	6
1889	0	0	1	1	0	1	1	0	0	1	2	2	0	2	11
1890	1	2	0	2	0	1	1	2	0	2	1	1	2	2	17
1891	0	0	0	1	1	1	0	1	1	0	3	3	1	0	12
1892	1	3	2	0	1	1	3	0	1	1	0	1	1	0	15
1893	0	1	0	0	0	1	0	2	0	0	1	3	0	0	8
1894	1	0	0	0	0	0	0	0	1	0	1	1	0	0	4
Total	16	16	12	12	8	11	17	12	7	13	15	25	24	8	196

Row (year) and **column** (corps) categorizations **independent**?

Frequentist goodness–of–fit tests

Goodness–of–fit tests often required, particularly at **initial stage** of data analysis.

$\mathcal{H}_0$: observation $\mathbf{x}^{(1)}$ consistent with **fully specified** distribution $\{\pi(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$.

OK for **nonparametric tests** or via sufficient conditioning in **exponential** families.

E.g. logistic regression, multidimensional contingency tables, Markov random fields.

Observe value $u^{(1)}$ of **any** particular **test statistic** $u = u(\mathbf{x})$ of scientific interest.

Suppose large values of $u^{(1)}$ suggest a conflict between data and model.

Then ***p*–value** for $u^{(1)}$ is $\Pr\{u(\mathbf{X}) \geq u^{(1)}\}$ under π.

NB. Fallacious Bayesian dismissal of frequentist p–values, c.f. **Fisher** (1922):

"More or less elaborate forms will be suitable according to the volume of the data".

What if one **cannot evaluate** $\Pr\{u(\mathbf{X}) \geq u^{(1)}\}$ under π?

Exact Monte Carlo p–values   Dwass (1957), Barnard (1963), B&D (1977)

Suppose can draw **random sample** $\mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}$ from π.

Compare test statistic $u^{(1)}$ with corresponding $u^{(2)}, \ldots, u^{(m)}$.

If $u^{(1)}$ is kth largest among all m values, declare **exact** p–value k/m.

If ties between ranks occur, quote range or use randomized rule.

Typically choose $m = 99$ or 999 or 9999.

Larger $m \;\Rightarrow\;$ finer gradation, increased power, more consistency.

NB. **Any** choice of test statistic u is OK! Always exact.

Exact sequential Monte Carlo p–values   Besag & Clifford (1991)

Substantially reduces expected sample size when $\mathcal{H}_0$ holds.

Ex. Hardy–Weinberg equilibrium for 13 alleles at a single locus ?

A_j

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}	A_i
0	0	0	0	0	0	0	2	0	0	0	0	0	A_1
	0	1	1	1	0	0	2	4	2	0	0	0	A_2
		2	0	2	1	0	4	1	2	2	0	0	A_3
			1	0	0	0	2	0	0	0	0	0	A_4
				1	0	0	2	2	1	1	0	0	A_5
					0	0	4	1	1	0	0	0	A_6
						1	0	0	1	0	0	0	A_7
							5	3	3	0	0	0	A_8
								3	4	0	0	0	A_9
									0	0	1	0	A_{10}
										0	0	0	A_{11}
											0	1	A_{12}
												0	A_{13}

$$\textbf{HWE} \quad \Rightarrow \quad \Pr(A_i \times A_j) = \begin{cases} p_i^2 & i = j \\ 2p_i p_j & i < j \end{cases} \qquad\qquad p_i = \Pr(A_i) = \ ?$$

Independence in folded square contingency tables (HWE)

Simple Monte Carlo test : Besag & Seheult (1983), Guo & Thompson (1992).

$$\Pr(A_i) = p_i \quad \Rightarrow \quad \Pr(A_i \times A_j) = \begin{cases} p_i^2 & i = j \\ 2p_ip_j & i < j \end{cases} \qquad i, j = 1, \ldots, m$$

$$\Pr(\mathbf{x}) = \frac{x_{++}!}{\prod_{i \le j} x_{ij}!} \prod_i p_i^{2x_{ii}} \prod_{i<j} (2p_ip_j)^{x_{ij}} = \frac{x_{++}!}{\prod_{i \le j} x_{ij}!} \prod_{i<j} 2^{x_{ij}} \prod_i p_i^{t_i}$$

where $t_i = (x_{1i} + x_{2i} + \ldots + x_{ii}) + (x_{ii} + x_{ii+1} + \ldots + x_{im})$.

$\Rightarrow \quad \mathbf{t} = (t_1, t_2, \ldots, t_m)$ is sufficient statistic for $\mathbf{p} = (p_1, p_2, \ldots, p_m)$.

$$\Pr(\mathbf{t}) = \frac{t_+!}{\prod_i t_i!} \prod_i p_i^{t_i} \quad \Rightarrow \quad \Pr(\mathbf{x} \,|\, \mathbf{t}) = \frac{n! \prod_i t_i! \prod_{i<j} 2^{x_{ij}}}{(2n)! \prod_{i \le j} x_{ij}!}$$

where $n =$ sample size $= x_{++} \quad \Rightarrow \quad t_+ = 2n$.

Monte Carlo *p*–value $= 0.305$ from 999 samples, $= 0.303$ from 999999 samples.

MCMC approximate *p*–value (Lazzarone & Lange, 1997) $= 0.316$.

Independence in folded square contingency tables (HWE)

Data from Cavalli–Sforza & Bodmer (1971), Guo & Thompson (1992).

A_i

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9		
1236	120	18	982	32	2582	6	2	115	A_1	
	3	0	55	1	132	0	0	5	A_2	
		0	7	0	20	0	0	2	A_3	
			249	12	1162	4	0	53	A_4	
				0	29	0	0	1	A_5	A_j
					1312	4	0	149	A_6	
						0	0	0	A_7	
							0	0	A_8	
								4	A_9	

Monte Carlo *p*–value $= 0.7154$ from 9999 samples.

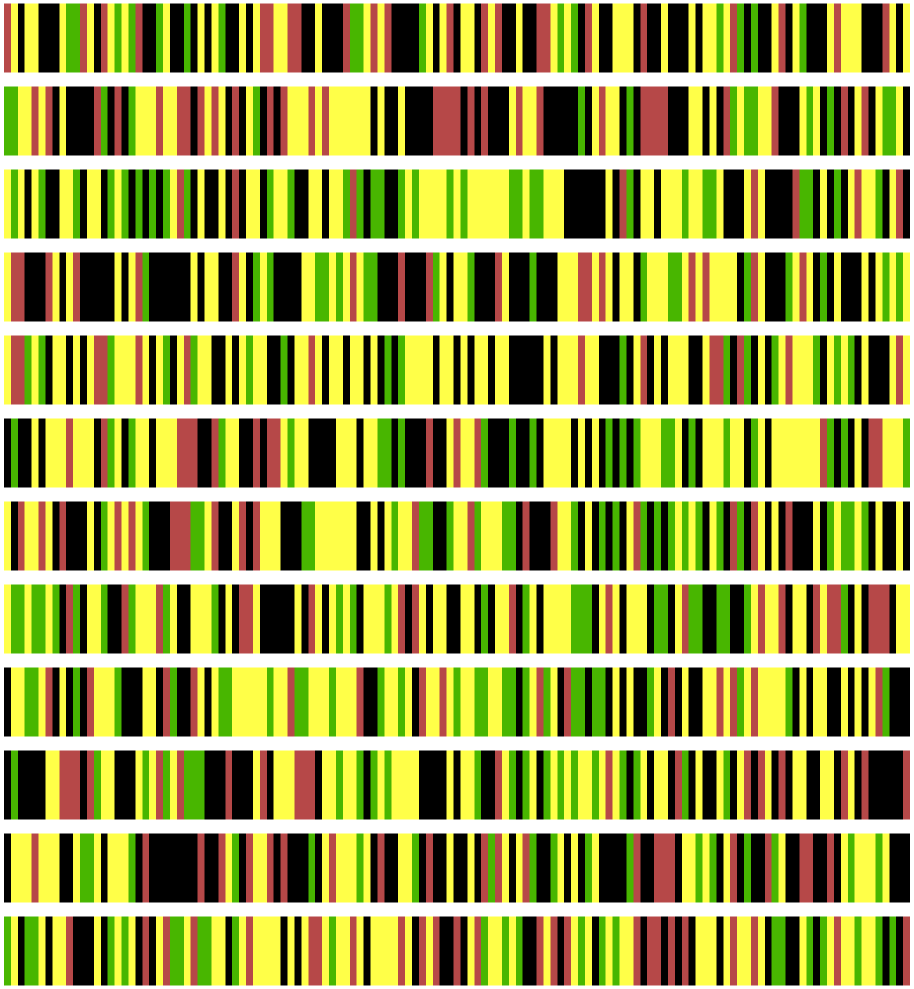
MCMC approximate *p*–value (Guo & Thompson, 1992) $= 0.6955$.

NB. Massive savings in both examples, using **sequential** Monte Carlo adaptation.

DNA sequence data

Avery & Henderson (1999)

1572 bases : black = A, green = C, purple = G, yellow = T



Goodness of fit for Markov chains

Bartlett (1951), Hoel (1954)

Observe sequence $\mathbf{x}^{(1)} = (x_0^{(1)}, \ldots, x_n^{(n)})$.

$\mathcal{H}_0$: sequence $\mathbf{x}^{(1)}$ is consistent with a Markov chain.

Corresponding **likelihood function**, given $x_0^{(1)}$, is

$$L(\mathbf{p}) = \prod_i \prod_j p_{ij}^{n_{ij}},$$

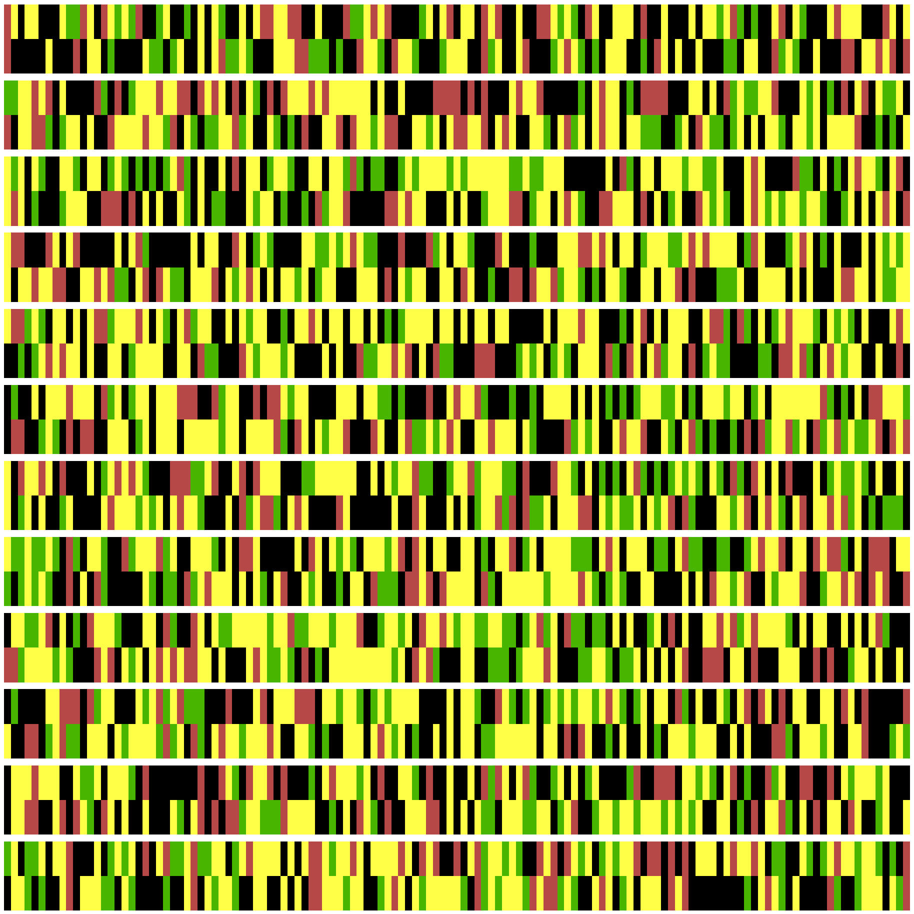
where p_{ij} is the (unknown) probability of one–step transition from state i to state j and n_{ij} is the **observed number of such transitions**.

The n_{ij}'s are **sufficient statistics** for the p_{ij}'s, and, **conditioning** on these values, distribution is **uniform** on the space $\mathcal{S}$ generated by $x_0^{(1)}$ and the observed n_{ij}'s.

Random samples, subject to required constraints, generated fast via **Euler tours**: Aldous (1990), Kandel, Matias, Unger & Winkler (1996), Besag & Mondal (?).

DNA sequence data

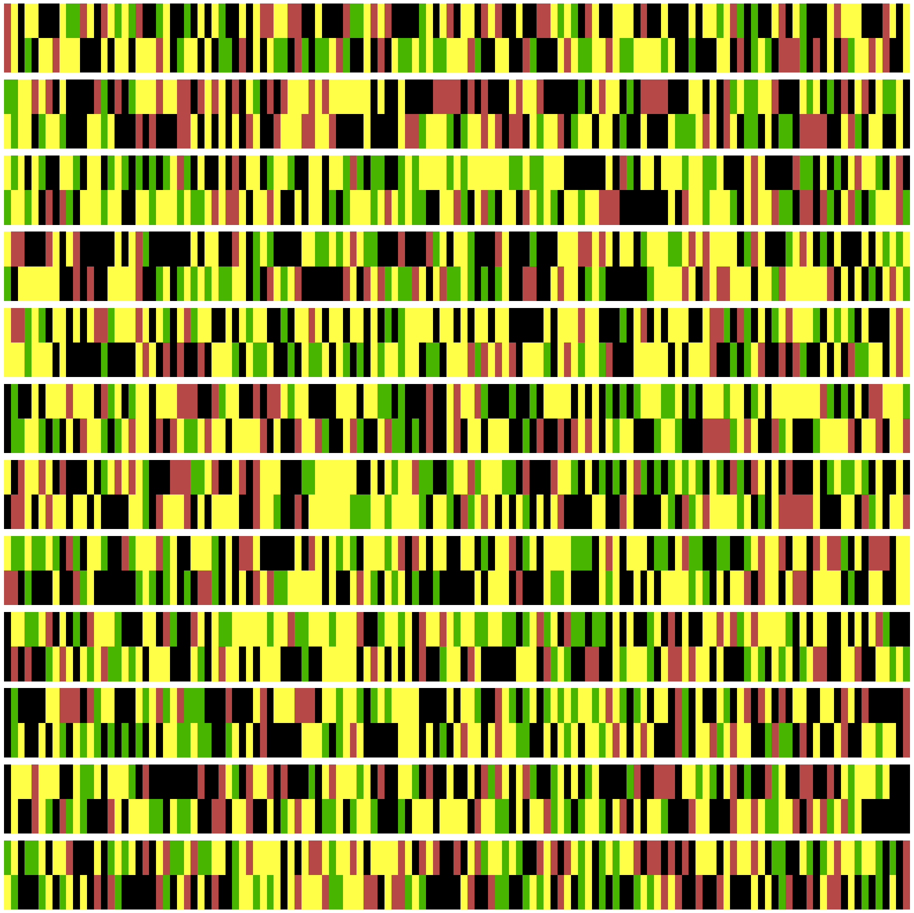
Data and sample with same frequencies of pairs.



Deviance = 55.66 on 36 d.fr. χ^2 : 0.019, MC : 0.030, MCMC : 0.027

DNA sequence data

Data and sample with same frequencies of triples.



Deviance = 150.31 on 144 d.fr. $\chi^2 : 0.342$, MC : 0.439, MCMC : 0.443

Binary data for 77 schizophrenics

Presence / absence of a particular response over 12 months

	Month											
Subject	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	1	1	1	1
4	1	1	0	0	1	0	0	0	0	1	1	1
5	1	0	0	0	0	0	0	0	0	0	0	0
6	1	1	0	0	0	0	0	0	0	0	0	0
7	1	1	0	0	0	0	0	0	0	0	0	0
8	1	1	0	0	0	0	0	0	0	0	0	0
9	1	1	0	0	0	0	0	0	0	0	0	0
10	1	1	1	0	0	0	0	0	0	0	0	0
11	1	1	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0
13	1	1	1	1	1	1	1	1	1	1	1	1
14	1	1	0	0	0	0	0	0	0	0	0	0
15	1	1	1	1	1	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
71	1	1	0	0	0	0	0	0	0	0	0	0
72	1	0	0	0	0	0	0	0	0	0	0	0
73	1	1	1	1	1	1	1	0	0	0	0	0
74	1	0	1	1	0	0	0	0	0	0	0	0
75	0	1	1	0	1	1	0	0	0	0	0	0
76	1	1	1	1	0	0	0	0	0	0	0	0
77	1	1	1	0	0	0	0	0	0	0	0	0

Binary data for 77 schizophrenics

Presence / absence of a particular response over 12 months.

Are data consistent with 77 **individual** Markov chains ?

Deviance v. 2nd–order chains = 47.32 nominally on 154 d.fr.

but ordinary Monte Carlo and MCMC p–values = 0.03.

NB. Above MC tests fix number of 00's, 01's, 10's & 11's for each subject.

To test for **single** Markov chain, must fix **overall** 00's, 01's, 10's & 11's.

1	1	1	1	0	1	0	0	0	0	0	1
1	1	1	1	0	0	1	0	0	0	0	1
1	1	1	1	0	0	0	1	0	0	0	1
1	1	1	1	0	0	0	0	1	0	0	1
1	1	1	1	0	0	0	0	0	1	0	1
1	1	1	0	1	1	0	0	0	0	0	1
1	1	1	0	1	0	0	0	0	0	1	1
1	1	1	0	0	1	1	0	0	0	0	1
1	1	1	0	0	1	0	0	0	0	1	1
1	1	1	0	0	0	1	1	0	0	0	1
1	1	1	0	0	0	1	0	0	0	1	1
1	1	1	0	0	0	0	1	1	0	0	1
1	1	1	0	0	0	0	1	0	0	1	1
1	1	1	0	0	0	0	0	1	1	0	1
1	1	1	0	0	0	0	0	1	0	1	1
1	1	0	1	1	1	0	0	0	0	0	1
1	1	0	1	1	0	0	0	0	0	1	1
1	1	0	1	0	0	0	0	0	1	1	1
1	1	0	0	1	1	1	0	0	0	0	1
1	1	0	0	1	1	0	0	0	0	1	1
1	1	0	0	1	0	0	0	0	1	1	1
1	1	0	0	0	1	1	1	0	0	0	1
1	1	0	0	0	1	1	0	0	0	1	1
1	1	0	0	0	1	0	0	0	1	1	1
1	1	0	0	0	0	1	1	1	0	0	1
1	1	0	0	0	0	1	1	0	0	1	1
1	1	0	0	0	0	1	0	0	1	1	1
1	1	0	0	0	0	0	1	1	1	0	1
1	1	0	0	0	0	0	1	1	0	1	1
1	1	0	0	0	0	0	1	0	1	1	1
1	0	1	1	1	1	0	0	0	0	0	1
1	0	1	1	1	0	0	0	0	0	1	1
1	0	1	1	0	0	0	0	0	1	1	1
1	0	1	0	0	0	0	0	1	1	1	1
1	0	0	1	1	1	1	0	0	0	0	1
1	0	0	1	1	1	0	0	0	0	1	1
1	0	0	1	1	0	0	0	0	1	1	1
1	0	0	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	1	1	0	0	0	1
1	0	0	0	1	1	1	0	0	0	1	1
1	0	0	0	1	1	0	0	0	1	1	1
1	0	0	0	1	0	0	0	1	1	1	1
1	0	0	0	0	1	1	1	1	0	0	1
1	0	0	0	0	1	1	1	0	0	1	1
1	0	0	0	0	1	1	0	0	1	1	1
1	0	0	0	0	1	0	0	1	1	1	1
1	0	0	0	0	0	1	1	1	1	0	1
1	0	0	0	0	0	1	1	1	0	1	1
1	0	0	0	0	0	1	1	0	1	1	1
1	0	0	0	0	0	1	0	1	1	1	1

Data for the 4th subject.

Permutations preserving 00's, 01's, 10's & 11's.

MC p–value from random sampling = 0.03.

Besag (1983), Besag & Clifford (1989)

Exact Markov chain Monte Carlo p–values

. . . **but what if random sampling from** $\{\pi(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$ **is not feasible ?**

e.g. Rasch model, hierarchical models for higher–dimensional contingency tables, Markov random fields in spatial statistics, random graphs for social networks, . . .

Can always construct **Metropolis algorithms** with π as **stationary** distribution.

Problems

1. Need to cope with burn–in for MCMC.
2. Need to cope with dependence in the MCMC.
3. MCMC may not be irreducible w.r.t. $\mathcal{S}$.

Metropolis algorithm

Target distribution $\pi = \{\pi(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$.

Consider *any* **symmetric transition matrix** R with elements $R(\mathbf{x}, \mathbf{x}')$

$$P(\mathbf{x}, \mathbf{x}') := R(\mathbf{x}, \mathbf{x}')\, A(\mathbf{x}, \mathbf{x}'), \qquad \mathbf{x}' \neq \mathbf{x} \in \mathcal{S},$$

where $A(\mathbf{x}, \mathbf{x}') := \min\{1, \pi(\mathbf{x}')/\pi(\mathbf{x})\}, \qquad \mathbf{x}' \neq \mathbf{x} \in \mathcal{S},$

with $P(\mathbf{x}, \mathbf{x})$ determined by subtraction. Then, for $\mathbf{x}' \neq \mathbf{x}$,

$$\begin{aligned} \pi(\mathbf{x})\, P(\mathbf{x}, \mathbf{x}') &= R(\mathbf{x}, \mathbf{x}') \min\{\pi(\mathbf{x}), \pi(\mathbf{x}')\} \\ \pi(\mathbf{x}')\, P(\mathbf{x}', \mathbf{x}) &= R(\mathbf{x}', \mathbf{x}) \min\{\pi(\mathbf{x}'), \pi(\mathbf{x})\} \end{aligned}$$

so that P satisfies **detailed balance** w.r.t. π!

R is the **proposal matrix**. If current state is $\mathbf{x}$, then $\mathbf{x}'$ is proposed as next state with probability $R(\mathbf{x}, \mathbf{x}')$ and is accepted with **acceptance probability** $A(\mathbf{x}, \mathbf{x}')$, else $\mathbf{x}$ is retained as next state.

Exact Markov chain Monte Carlo p–values

$\mathcal{H}_0$: $\mathbf{X} = (X_1, \ldots, X_n)$ has **known** but **very complex** distribution $\pi(\mathbf{x})$.

Dataset $\mathbf{x}^{(1)}$ $\Rightarrow$ **test statistic** $u^{(1)} = u(\mathbf{x}^{(1)})$.

Reject $\mathcal{H}_0$ if $u^{(1)}$ is extreme w.r.t. draws from π.

Assume π cannot be simulated directly, so ordinary Monte Carlo is not available.

Need a fix ...

Construct **transition matrix** $\{P(\mathbf{x}, \mathbf{x}')\}$ that has π as **stationary** distribution.

Under $\mathcal{H}_0$ corresponding Markov chain initiated by $\mathbf{x}^{(1)}$ is **stationary** !!

However, successive observations are (highly) **dependent**.

Need another fix ...

Note that corresponding **backwards** transition matrix Q has $(\mathbf{x}', \mathbf{x})$ element

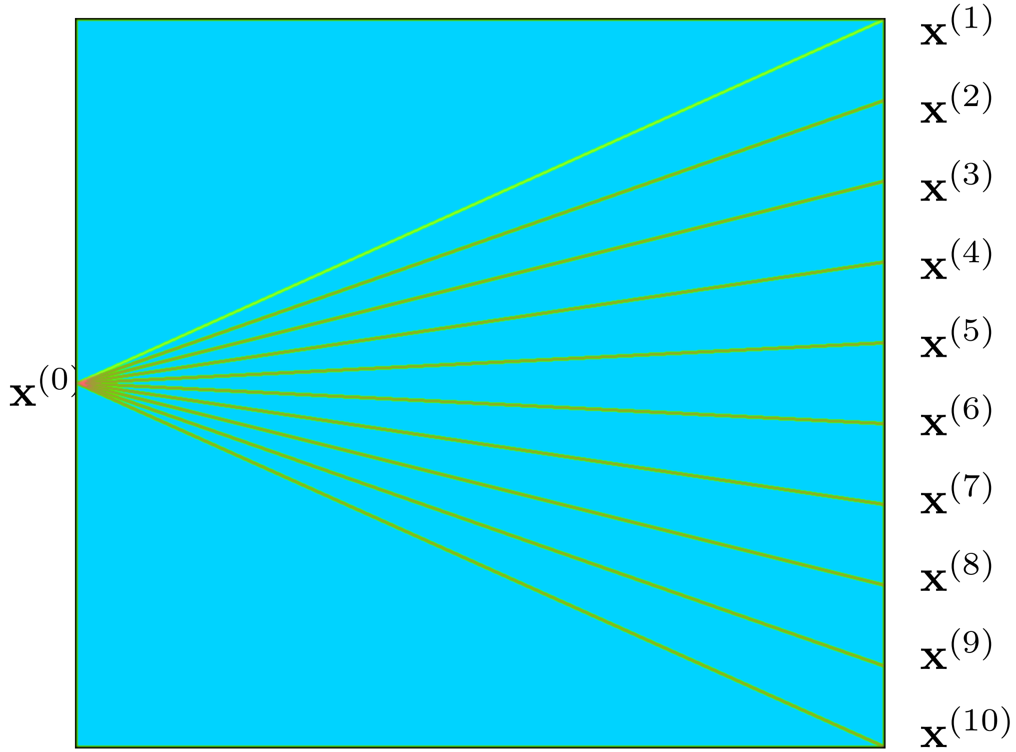
$$Q(\mathbf{x}', \mathbf{x}) = \pi(\mathbf{x})\, P(\mathbf{x}, \mathbf{x}') \,/\, \pi(\mathbf{x}')$$

so Q also **maintains** π and, if P is **time reversible** (e.g. **Metropolis**), $Q = P$.

Parallel runs

Use Q to run chain backwards t steps from data $\mathbf{x}^{(1)} \Rightarrow \mathbf{x}^{(0)}$.

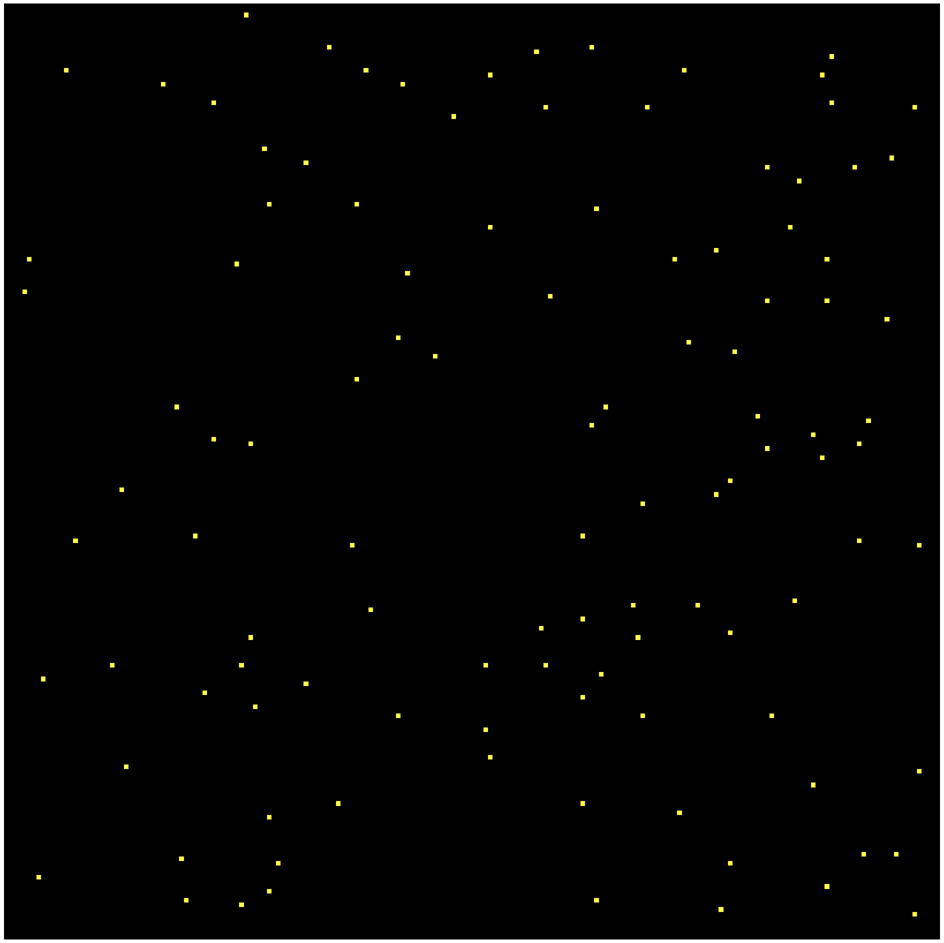
Use P to run chain forwards t steps from $\mathbf{x}^{(0)}$, $m-1$ times $\Rightarrow \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}$.



$\mathcal{H}_0 \Rightarrow \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$ are **exchangeably** from π; and so are $u^{(1)}, \ldots, u^{(m)}$.

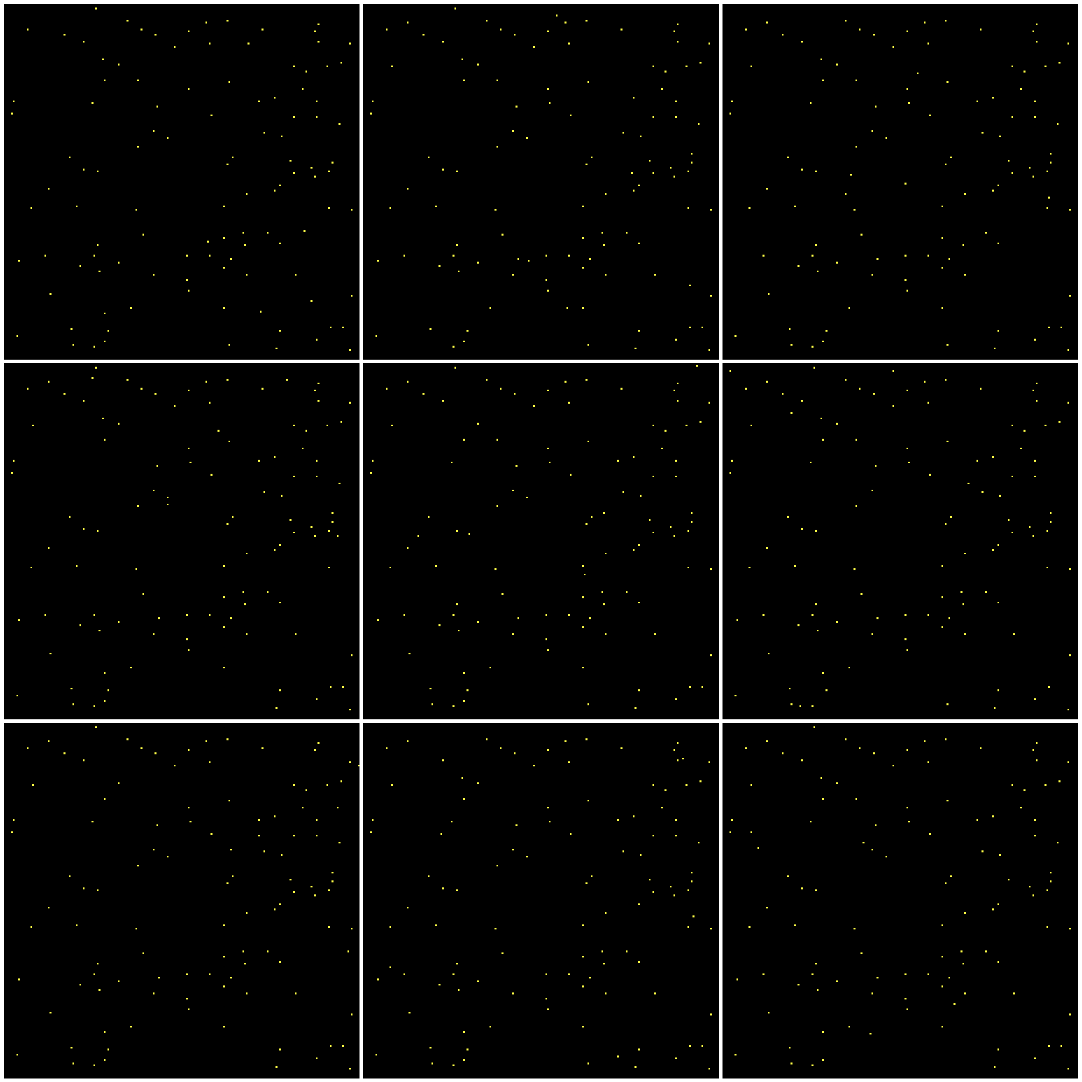
Rank of $u^{(1)}$ among $u^{(1)}, \ldots, u^{(m)}$ provides an **exact** p–value for $\mathcal{H}_0$.

Ex. Nesting sites of 110 Gray Gulls in Northern Chile



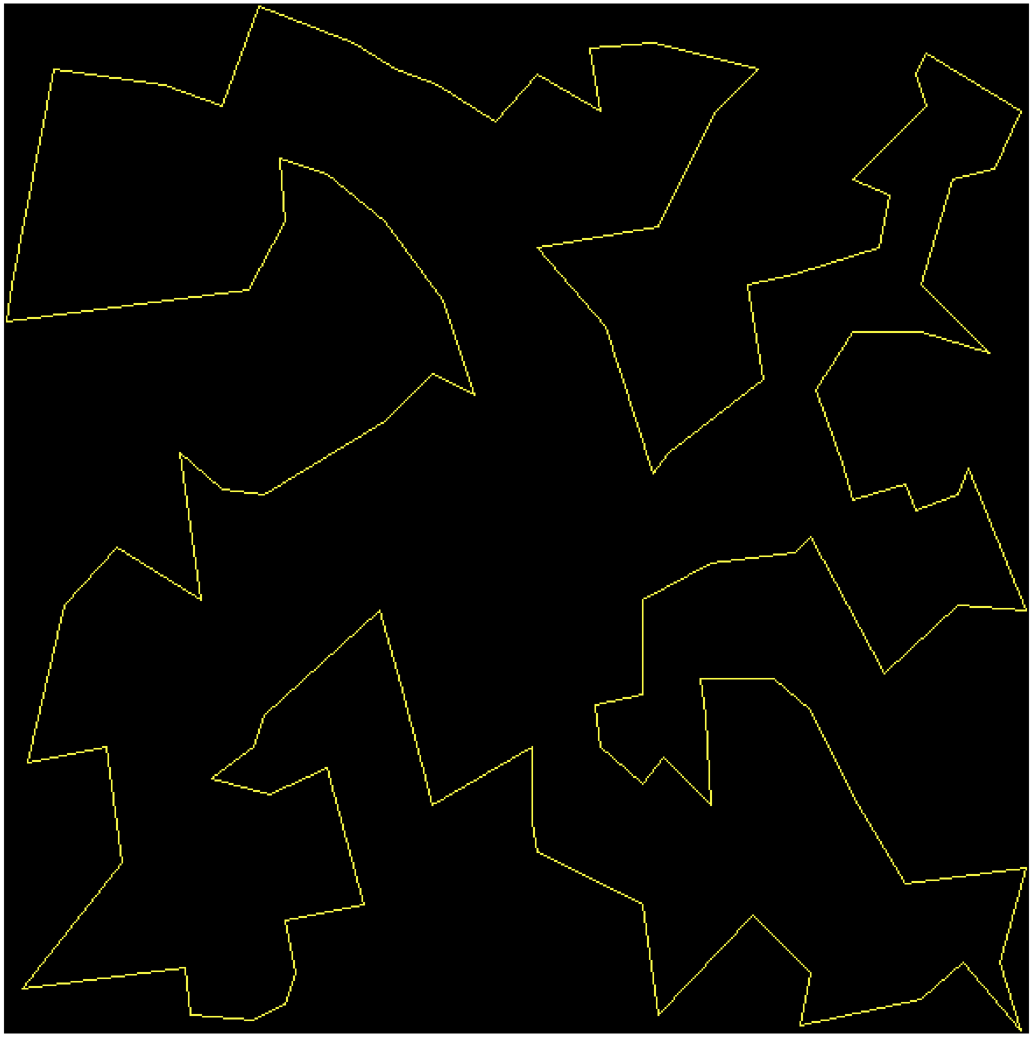
$\mathcal{H}_0$: observed configuration $\mathbf{x}^{(1)}$ is a realization of a Strauss process.

Short runs in conditional simulations for goodness–of-fit tests :



$\mathbf{x}^{(0)}$ in the center, surrounded by $\mathbf{x}^{(1)}$ in NW and $\mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(8)}$.

Gray Gulls: a fast route around the nests!



... using Metropolis algorithm for simulated annealing.

Ex. Darwin's finches

Sanderson (2000), Manly (1995)

Presence (1) or absence (0) of 13 species of finch on 17 Galapagos Islands.

Presence (1) or absence (0) of 13 species of finch on 17 Galapagos Islands.

Island identifier

Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	14
2	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0	13
3	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	14
4	0	0	1	1	1	0	0	1	0	1	0	1	1	0	1	1	1	10
5	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1	0	0	12
6	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	2
7	0	0	1	1	1	1	1	1	1	0	0	1	0	1	1	0	0	10
8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
9	0	0	1	1	1	1	1	1	1	1	0	1	0	0	1	0	0	10
10	0	0	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0	11
11	0	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0	6
12	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2
13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	17
Total	4	4	11	10	10	8	9	10	8	9	3	10	4	7	9	3	3	

Define F_{ij} = number of islands on which **species** i and j **co–exist**.

p–value 0.0001 for the **Rasch model**, based on overall **co–occurrences**.

Rasch model

$\mathbf{x}^{(1)}$ is observed $r \times s$ **binary** table. Can be very large in educational testing.

e.g. $x_{ij} = 1$ or 0 is correct or incorrect response of candidate i to item j.

e.g. $x_{ij} = 1$ or 0 is presence or absence of species i in location j.

Most common statistical formulation for binary tables is Rasch model $\Rightarrow$

x_{ij}'s independent & odds of 1 to 0 in cell (i, j) are $\theta_{ij} : 1$, where $\theta_{ij} = \phi_i \, \psi_j$ $\Rightarrow$

Any binary table $\mathbf{x}$ has probability

$$\pi(\mathbf{x}\,; \boldsymbol{\theta}) \;=\; \prod_{i=1}^{r} \prod_{j=1}^{c} \frac{\theta_{ij}^{x_{ij}}}{1 + \theta_{ij}} \;=\; \frac{\prod_i \phi_i^{x_{i+}} \, \prod_j \psi_j^{x_{+j}}}{\prod_i \prod_j \, (1 + \phi_i \, \psi_j)}$$

Conditioning on observed row and column totals $x_{i+}^{(1)}$ & $x_{+j}^{(1)}$ eliminates ϕ_i's and ψ_j's.

$\Rightarrow$ uniform distribution $\pi(\mathbf{x})$ on constrained space $\mathcal{S}$.

Very nasty counting problem, no methods of random sampling, require MCMC, ...

Equivalent to testing for **no 3-way interaction** in a $2 \times r \times c$ contingency table.

Proposals used in simulating the Rasch model

Margins are **preserved** by moves of the form :

$$\begin{array}{cccccc}
 & \vdots & & \vdots & \\
\cdots & 0 & \cdots & 1 & \cdots \\
 & \vdots & & \vdots & \\
\cdots & 1 & \cdots & 0 & \cdots \\
 & \vdots & & \vdots &
\end{array}
\quad
\begin{array}{c}
\rightarrow \\
\text{or} \\
\leftarrow
\end{array}
\quad
\begin{array}{cccccc}
 & \vdots & & \vdots & \\
\cdots & 1 & \cdots & 0 & \cdots \\
 & \vdots & & \vdots & \\
\cdots & 0 & \cdots & 1 & \cdots \\
 & \vdots & & \vdots &
\end{array}$$

Can prove **irreducibility** e.g. Jennison (1983).

E.g. choose **two rows** and **two columns** at random, propose corresponding **swap** and **accept swap if valid** else retain existing configuration. OK **Metropolis**.
For more efficient alternatives, see Besag & Clifford (1989).

Can also modify moves to cater for **structural zeros**.

Bayesian inference for incomplete $2 \times I \times K$ tables

Geographical epidemiology for non–communicable disease

Zones (e.g. counties) $i = 1, \ldots, I$

Risk (e.g. ethnic) groups $k = 1, \ldots, K$

Known number at risk n_{ik} in zone i, group k.

Unknown number of cases y_{ik} among n_{ik} at risk.

Known total number of cases y_{i+} in each zone i.

Excellent Poisson approximation for rare disease, else ...

Voting behaviour in two–party system Wakefield (2004)

Poisson approximation is no longer appropriate.

More informative version ...

Known total number of cases y_{+k} in each group k?? Cressie & Chan (1989)

Bayesian inference for incomplete $2 \times I \times K$ tables

Besag (2006)

Known numbers at risk n_{ik}

$$
\begin{array}{ll}
 & \text{Risk group } k \\
 & \\
 & \begin{array}{ccccccc}
n_{11} & \dots & n_{1k'} & \dots & n_{2k''} & \dots & n_{1K} \\
\vdots & & \vdots & & \vdots & & \vdots \\
n_{i1} & \dots & n_{ik'} & \dots & n_{ik''} & \dots & n_{iK} \\
\vdots & & \vdots & & \vdots & & \vdots \\
n_{I1} & \dots & n_{Ik'} & \dots & n_{Ik''} & \dots & n_{IK}
\end{array}
\end{array}
$$

Zone i

Unknown numbers of cases y_{ik} **:** y_{i+} **known**

$$
\begin{array}{ccccccccc}
 & & \text{Risk group } k & & & & & & \text{Total} \\
 & & & & & & & & \\
y_{11} & \dots & y_{1k'} & \dots & y_{2k''} & \dots & y_{1K} & & y_{1+} \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
y_{i1} & \dots & y_{ik'} & \dots & y_{ik''} & \dots & y_{iK} & & y_{i+} \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
y_{I1} & \dots & y_{Ik'} & \dots & y_{Ik''} & \dots & y_{IK} & & y_{I+}
\end{array}
$$

Zone i

Assume that y_{ik}'s are binomial with parameters n_{ik} and p_{ik}, say.

For a Bayesian justification, see Knorr–Held & Besag (1998).

Then the posterior–predictive distribution for the y_{ik}'s and the p_{ik}'s is

$$\begin{aligned}
\pi(\mathbf{y},\mathbf{p}\,|\,\mathbf{y}_{?+},\mathbf{n}) &= \mathcal{L}(\mathbf{y}\,|\,\mathbf{y}_{?+},\mathbf{p},\mathbf{n})\,\pi(\mathbf{p}\,|\,\mathbf{y}_{?+},\mathbf{n}) \\
&\propto \mathcal{L}(\mathbf{y}\,|\,\mathbf{y}_{?+},\mathbf{p},\mathbf{n})\,\mathcal{L}(\mathbf{y}_{?+}\,|\,\mathbf{p},\mathbf{n})\,\pi(\mathbf{p}\,|\,\mathbf{n}) \\
&\propto \mathcal{L}(\mathbf{y},\mathbf{y}_{?+}\,|\,\mathbf{p},\mathbf{n})\,\pi(\mathbf{p}\,|\,\mathbf{n}),
\end{aligned}$$

in which the role of $\mathbf{y}_{?+}$ is to constrain the y_{ik}'s to sum to y_{i+} for each i.

The above factorization decouples the likelihood from the prior in the MCMC.

Propose $(y_{ik'}, y_{ik''}) \to (y_{ik'}+r, y_{ik''}-r)$, with r uniform on $1, 2, \ldots, r_0$.

Irreducible with Metropolis ratio for valid proposals satisfying constraints

$$\frac{\pi(y_{ik'}+r, y_{ik''}-r\,|\,\ldots)}{\pi(y_{ik'}, y_{ik''}\,|\,\ldots)} = \frac{y_{ik'}!\,(n_{ik'}-y_{ik'})!\,y_{ik''}!\,(n_{ik''}-y_{ik''})!\;\exp(r\xi_{ik'k''})}{(y_{ik'}+r)!\,(n_{ik'}-y_{ik'}-r)!\,(y_{ik''}-r)!\,(n_{ik''}-y_{ik''}+r)!},$$

where $\xi_{ik'k''} = \ln(p_{ik'}/q_{ik'}) - \ln(p_{ik''}/q_{ik''})$ and $q_{ik} = 1 - p_{ik}$.

Bayesian inference for less incomplete $2 \times I \times K$ tables

Unknown numbers of cases y_{ik} : y_{i+}, y_{+k} known

$$
\begin{array}{ll|ccccccc|c}
 & & & & \text{Risk group } k & & & & & \text{Total} \\
 & & y_{11} & \cdots & y_{1k'} & \cdots & y_{2k''} & \cdots & y_{1K} & y_{1+} \\
 & & \vdots & & \vdots & & \vdots & & \vdots & \vdots \\
 & & y_{i'1} & \cdots & y_{i'k'} & \cdots & y_{i'k''} & \cdots & y_{i'K} & y_{i'+} \\
\text{Zone } i & & \vdots & & \vdots & & \vdots & & \vdots & \vdots \\
 & & y_{i''1} & \cdots & y_{i''k'} & \cdots & y_{i''k''} & \cdots & y_{i''K} & y_{i''+} \\
 & & \vdots & & \vdots & & \vdots & & \vdots & \vdots \\
 & & y_{I1} & \cdots & y_{Ik'} & \cdots & y_{Ik''} & \cdots & y_{IK} & y_{I+} \\
\text{Total} & & y_{+1} & \cdots & y_{+k'} & \cdots & y_{+k''} & \cdots & y_{+K} &
\end{array}
$$

Propose $(y_{i'k'}, y_{i'k''}, y_{i''k'}, y_{i''k''}) \rightarrow (y_{i'k'} + r, y_{i'k''} - r, y_{i''k'} - r, y_{i''k''} + r)$

Irreducible with Metropolis ratio ...

Three–dimensional contingency tables

Smoking and Lung Cancer in China : $8 \times 2 \times 2$ (Agresti, 1996)

			Lung Cancer? Y	Lung Cancer? N	Odds Ratio
BEIJING	Smoker?	Y	126	100	2.20
		N	35	61	
SHANGHAI	Smoker?	Y	908	688	2.14
		N	497	807	
SHENYANG	Smoker?	Y	913	747	2.18
		N	336	598	
NANJINK	Smoker?	Y	235	172	2.85
		N	58	121	
HARBIN	Smoker?	Y	402	308	2.32
		N	121	215	
ZHENGZHOU	Smoker?	Y	182	156	1.59
		N	72	98	
TAIYUAN	Smoker?	Y	60	99	2.37
		N	11	43	
NANCHANG	Smoker?	Y	104	89	2.00
		N	21	36	

Test for no 3–way interaction in a 3–dimensional table

In multinomial sampling, with sample size n, the probability of a table $\mathbf{x}$ is

$$\pi_0(\mathbf{x}) \;=\; \frac{x_{+++}!}{\prod_{ijk} x_{ijk}!} \prod_{ijk} p_{ijk}^{x_{ijk}}$$

where $x_{+++} = n$ and the p_{ijk}'s are the cell probabilities.

No 3–way interaction $\;\Rightarrow\;$ $p_{ijk} = a_{ij}\, b_{ik}\, c_{jk}$ and hence

$$\pi_0(\mathbf{x}) \;=\; \frac{x_{+++}!}{\prod_{ijk} x_{ijk}!} \prod_{ij} a_{ij}^{x_{ij+}} \prod_{ik} b_{ik}^{x_{i+k}} \prod_{jk} c_{jk}^{x_{+jk}},$$

so the sufficient statistics are x_{ij+}, x_{i+k}, x_{+jk} over all valid i, j, k.

Let $\mathcal{S}$ be the sample space induced by conditioning on the sufficient statistics.

$$\Rightarrow \qquad \frac{\pi(\mathbf{x}')}{\pi(\mathbf{x})} \;=\; \frac{\pi_0(\mathbf{x}' \,|\, \mathbf{x}' \in \mathcal{S})}{\pi_0(\mathbf{x} \,|\, \mathbf{x} \in \mathcal{S})} \;=\; \prod_{ijk} \frac{x_{ijk}!}{x'_{ijk}!}\,, \qquad \mathbf{x}, \mathbf{x}' \in \mathcal{S},$$

which is especially easy to evaluate when $|x'_{ijk} - x_{ijk}|$ is small for all i, j, k.

Exact test for no 3–way interaction in a 3–dimensional table $\mathbf{x}^{(1)}$

1. **Sufficient statistics** are $x^{(1)}_{ij+}$, $x^{(1)}_{i+k}$, $x^{(1)}_{+jk}$ over all valid i, j, k.
2. Construct the two cubes $\mathcal{C}$ of alternating ± 1's s.t. sufficient statistics preserved.

   $\mathcal{C}$ defines the set of **proposals**.
3. Let $\mathbf{x} = \mathbf{x}^{(1)}$, the observed table.
4. Choose 8 vertices to form a **cuboid** in current table $\mathbf{x}$.
5. Apply random draw from $\mathcal{C}$ to chosen vertices of $\mathbf{x} \quad \Rightarrow \quad$ new table $\mathbf{x}^*$.
6. Set $\mathbf{x} = \mathbf{x}^*$ with probability $\min\{1, \pi(\mathbf{x}^*)/\pi(\mathbf{x})\}$, else retain $\mathbf{x}$.
7. Return to 4

NB. If $\mathbf{x}^*$ has any negative entries, $\mathbf{x}^*$ is rejected by 6.

OK **except** must fix exactness by **parallel** or **forwards / backwards** versions.

Testing homogeneity of odds ratios for smoking and lung cancer

Test of no three–way interaction :

$$\text{Deviance} = 5.20 \text{ on 7 d.fr.} \qquad \left\{ \begin{array}{lcl} \text{Conventional } p\text{–value} & = & 0.64 \\ \text{MCMC exact p–value} & = & 0.63 \end{array} \right.$$

Test based on Mantel–Haenszel estimate of common odds ratio :

$$\text{Breslow–Day statistic} = 5.20 \text{ on 7 d.fr.} \qquad \left\{ \begin{array}{lcl} \text{Conventional p–value} & = & 0.64 \ ? \\ \text{MCMC exact p–value} & = & 0.63 \end{array} \right.$$

General Metropolis algorithm for testing q–dimensional tables

To test appropriateness of a hierarchical model for a q–dimensional table:

1. Identify **sufficient statistics** for corresponding model parameters.
2. Construct collection $\mathcal{C}$ of q–d hypercubes of ± 1's (and 0's?) symmetrically s.t. sufficient statistics are preserved: $\mathcal{C}$ defines the **proposals**.
3. Set $\mathbf{x} =$ observed table.
4. Choose 2^q vertices of a q–d hypercube in current table $\mathbf{x}$.
5. Apply random draw from $\mathcal{C}$ to chosen vertices of $\mathbf{x} \quad \Rightarrow \quad$ new table $\mathbf{x}^*$.
6. Set $\mathbf{x} = \mathbf{x}^*$ with probability $\min\{1, \pi(\mathbf{x}^*)/\pi(\mathbf{x})\}$, else retain $\mathbf{x}$.
7. Return to 4

NB. If $\mathbf{x}^*$ has any negative entries, $\mathbf{x}^*$ is rejected by 6.

OK **except** must fix exactness by **parallel** or **forwards / backwards** versions.

Alcohol, Cigarette and Marijuana use by Gender and Race

from A. Agresti (1996), *An Introduction to Categorical Data Analysis*, Wiley.

Response variables: Alcohol (A), Cigarette (C) & Marijuana (M) use.

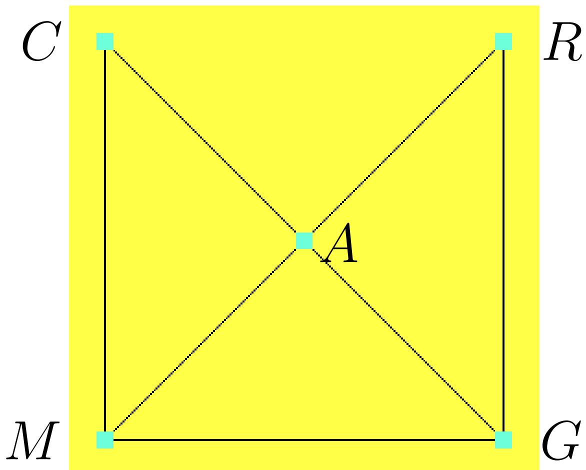
Explanatory variables: Gender (G) & Race (R).

	Gender		Female				Male			
	Race		White		Other		White		Other	
	Marijuana?		Y	N	Y	N	Y	N	Y	N
Alcohol? Y	Smoker?	Y	405	268	23	23	453	228	30	19
		N	13	218	2	19	28	201	1	18
Alcohol? N	Smoker?	Y	1	17	0	1	1	17	1	8
		N	1	117	0	12	1	133	0	17

2^5 contingency table

Alcohol, Cigarette and Marijuana use by Gender and Race

Agresti's model 6 allows pairwise interactions $\{AC, AM, AG, AR, CM, GM, GR\}$



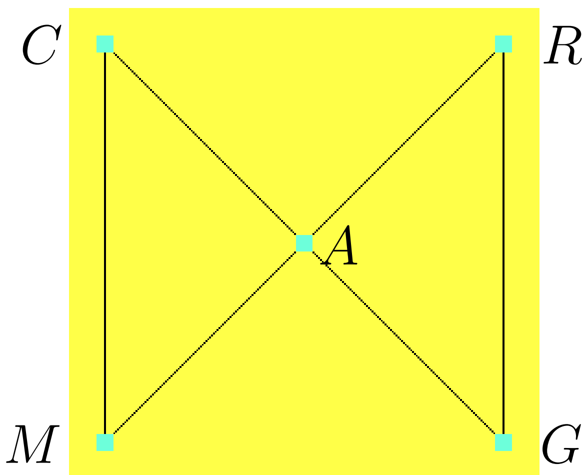
pairwise interactions graph

Deviance = 19.91 on 19 d.fr.

$$\begin{cases} \text{Conventional p–value} & = & 0.40 \\ \text{MCMC exact p–value} & = & 0.18 \end{cases}$$

Alcohol, Cigarette and Marijuana use by Gender and Race

Agresti's model 7 allows $\{AC, AM, AG, AR, CM, GR\}$ 2^5 contingency table



pairwise interactions graph

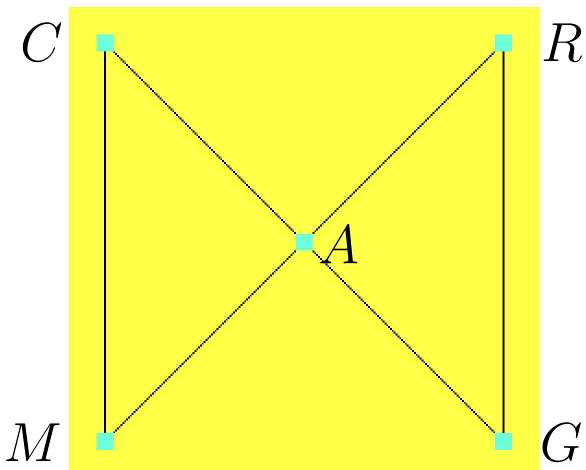
Deviance = 28.81 on 20 d.fr. $\left\{ \begin{array}{lcl} \text{Conventional p–value} & = & 0.09 \\ \text{MCMC exact p–value} & = & 0.02 \end{array} \right.$

NB. Model 7, adding 1 to all observations :

Deviance = 33.82 on 20 d.fr. $\left\{ \begin{array}{lcl} \text{Conventional p–value} & = & 0.03 \\ \text{MCMC exact p–value} & = & 0.02 \end{array} \right.$

Alcohol, Cigarette and Marijuana use by Gender and Race

JB's model 7^+ is the graphical (conditional independence) model $\{ACM, AGR\}$



conditional dependence graph

Deviance = 26.33 on 18 d.fr.

$$\left\{ \begin{array}{ccc} \text{Conventional p–value} & = & 0.09 \\ \text{MCMC exact p–value} & = & 0.08 \\ \text{Simple MC exact p-value} & = & 0.08 \end{array} \right.$$

NB. Proposal list contains 5621641 2^5 hypercube configurations of ± 1's & 0's.

MCMC in an extended space $\mathcal{S}^\dagger$ for q–dimensional tables

Consider any particular hierarchical model for an observed table $\mathbf{x}^{(1)}$.

Let $\{t_k(\mathbf{x}) : k \in \mathcal{K}\}$ denote corresponding jointly sufficient statistics.

Let $\mathcal{S}$ denote the set of non–negative tables $\mathbf{x}$ s.t. $t_k(\mathbf{x}) = t_k(\mathbf{x}^{(1)}),\ k \in \mathcal{K}$.

Algorithms based on ± 1 hypercubes and limited to $\mathcal{S}$ are rarely irreducible.

Suppose relax non–negativity: e.g. allow a single negative element.

Define $\mathcal{S}^\dagger$ to be the corresponding space and extend $\pi(\mathbf{x})$ to

$$\pi^\dagger(\mathbf{x}) \ \propto \ \exp\{\lambda\eta(\mathbf{x})\} \,/ \prod_{x_{ijkl}\geq 0} x_{ijkl}! \qquad\qquad \mathbf{x} \in \mathcal{S}^\dagger,$$

where $\lambda > 0$ and $\eta(\mathbf{x}) = \sum_{ijkl} \min(0, x_{ijkl})$ penalizes negative x_{ijkl}'s.

NB1. Behaviour of $\pi^\dagger(\mathbf{x})$ is almost seamless between $\mathcal{S}$ and $\mathcal{S}^\dagger \backslash \mathcal{S}$.

NB2. Proposals not in $\mathcal{S}^\dagger$ are always rejected in Metropolis but ...

... can be avoided using an appropriate Hastings correction.

Solving Diaconis and Sturmfels' $4 \times 4 \times 6$ example

Suppose $\mathbf{x}^{(1)}$ is the $4 \times 4 \times 6$ table

0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0
0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

The 3-d hypercubes are $\mathbf{z}$ and $-\mathbf{z}$, where $\mathbf{z}$ has elements

$$(z_{000}, z_{001}, z_{010}, z_{011}, z_{100}, z_{101}, z_{110}, z_{111}) \;=\; (+1, -1, -1, +1, -1, +1, +1, -1)$$

Conditioning on the sufficient statistics, $\mathcal{S}$ has two states, $\mathbf{x}^{(1)}$ and

0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0

If states are restricted to $\mathcal{S}$, the increments $\pm\mathbf{z}$ do not permit any moves.

However, $\pm\mathbf{z}$ increments connect the states if negative entries are allowed.

E.g. with $\lambda = 3$, each of the two states appears about 18% of the time.

Markov random fields for binary data

$$\pi(\mathbf{x}) \;\propto\; \exp\Big\{\sum_i \alpha_i x_i + \sum_{i<j} \alpha_{ij} x_i x_j + \sum_{i<j<k} \alpha_{ijk} x_i x_j x_k + \ldots + \alpha_{12\ldots n} x_1 x_2 \ldots x_n\Big\}$$

where $\alpha_{ij\ldots l} = 0$ unless $(i, j, \ldots, l)$ is a cliquo.

Homogeneity assumptions produce schemes of the form

$$\pi(\mathbf{x}; \boldsymbol{\theta}) \;=\; \frac{\exp\{\theta_1 t_1(\mathbf{x}) + \ldots + \theta_q t_q(\mathbf{x})\}}{c(\boldsymbol{\theta})}$$

where $\theta_1, \ldots, \theta_q$ are parameters, $c(\boldsymbol{\theta})$ is a normalizing constant and $t_1(\mathbf{x}), \ldots, t_q(\mathbf{x})$ are **sums of products–over–cliquos** of the x_i's.

$t_1(\mathbf{x}), \ldots, t_q(\mathbf{x})$ are **jointly sufficient statistics** for $\theta_1, \ldots, \theta_q$.

Markov random graphs for social networks

Social network for class of 24 school kids : 13 boys, 11 girls.

$X_{ij} = 1$ if i claims j is a friend, else $X_{ij} = 0$.

j

		0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
	1		0	0	1	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
	0	0		0	0	1	0	1	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0
	0	1	0		0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0
	0	0	0	0		1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	1		0	1	0	1	0	1	1	0	0	0	1	0	0	1	0	0	0	0
	0	0	0	0	1	0		0	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	1
	1	0	1	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	
	1	1	0	1	0	0	0	1		0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
	0	0	1	0	1	1	1	0	0		1	0	1	0	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	1	1	1	1	1	1		1	1	0	0	0	1	0	1	1	0	1	1	1
	1	0	1	0	0	1	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0
i	1	1	1	1	1	1	0	0	0	1	0	1		0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0		1	1	1	1	0	1	0	1	1	1
	1	0	1	0	0	0	0	1	0	0	0	0	0	1		0	1	1	1	0	1	1	1	1
	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0		1	1	1	1	0	1	1	0
	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0		1	0	1	0	1	0	1
	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0		0	1	0	0	0	1
	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1		1	0	1	1	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1		0	1	1	1
	1	0	0	1	0	0	1	1	1	0	0	1	1	1	1	0	0	0	0	0		0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0		0	1
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0		0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	

Social networks

Individuals : $i,\ j,\ k,\ \ldots$

Ordered pairs ("sites") : (i,j) for $i \neq j$.

Relations : $X_{ij} = 1$ if i is "tied" to j.

$X_{ij} = 0$ if i is not tied to j.

$$\mathbf{X} = \{X_{ij}\}, \qquad \Pr(\mathbf{X} = \mathbf{x}) = \pi(\mathbf{x}) = \ ???$$

Markov property of Frank & Strauss, 1986

$$\Pr(x_{ij} \mid \ldots) \equiv \Pr(x_{ij} \mid x_{ji},\, x_{ik},\, x_{ki},\, x_{jk},\, x_{kj},\ k \neq i,j)$$

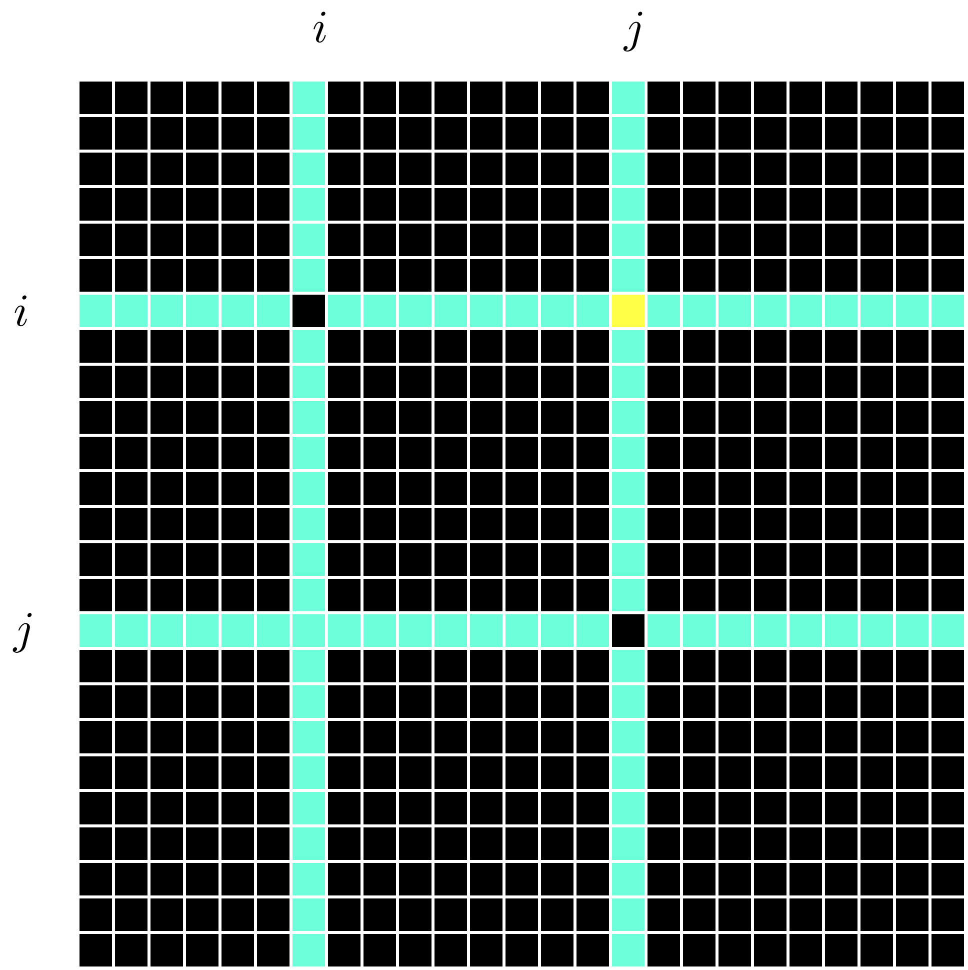
e.g. X_{12} is conditionally independent of X_{34}

$\Rightarrow$ cliques (i.e. maximal cliquos) :

Type I : $\{\,(i,j),\,(j,i),\,(i,k),\,(k,i),\,(j,k),\,(k,j)\,\}$ $\quad i,\ j,\ k$ distinct

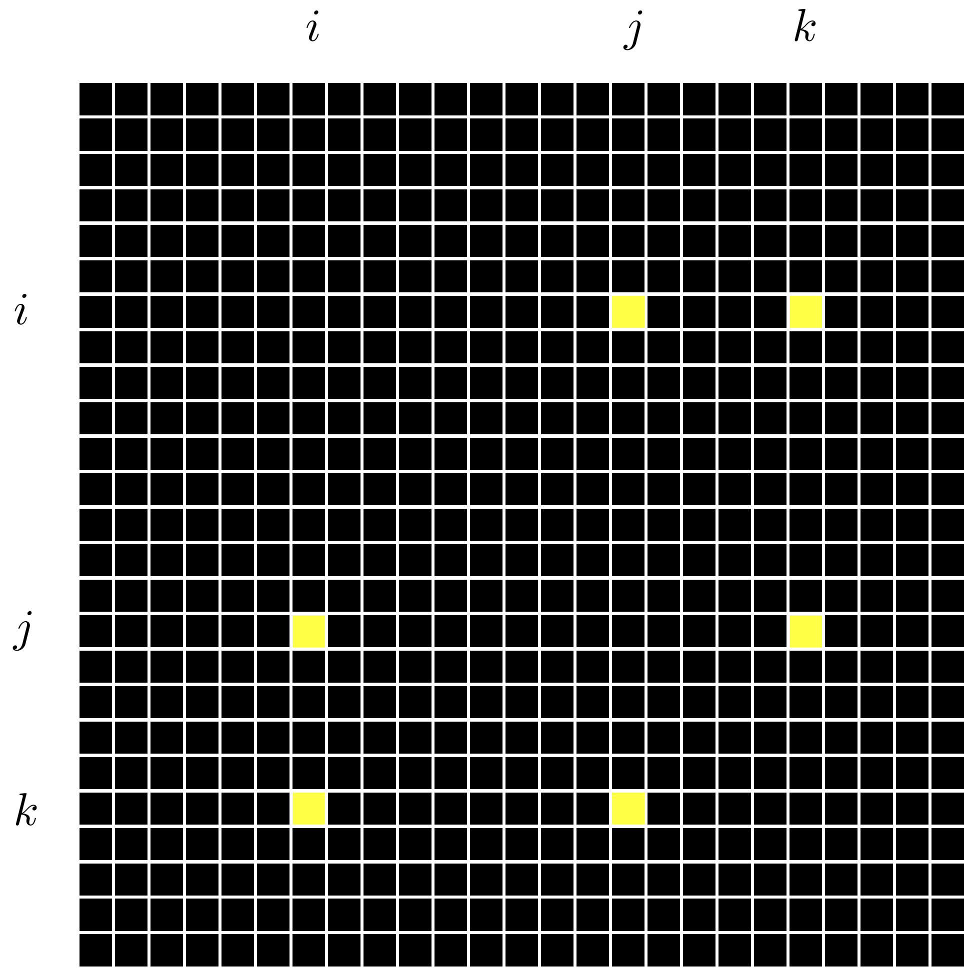
Type II : $\{\,(i,j),\,(j,i),\,(i,k),\,(k,i),\,(i,l),\,(l,i),\,\ldots\}$ $\quad i,\ j,\ k,\ l,\ \ldots$ distinct

Markov property



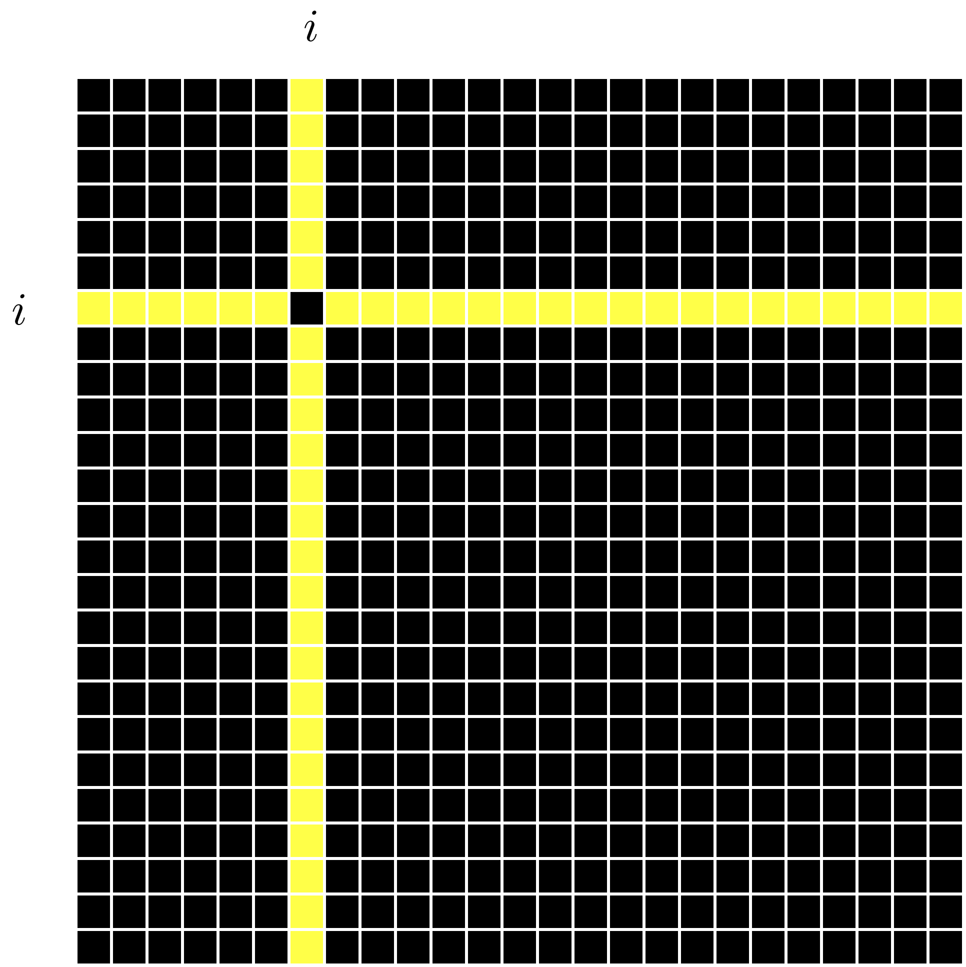
Neighbors of yellow site (i, j) in blue.

Type I clique



Clique $\{(i, j), (j, i), (i, k), (k, i), (j, k), (k, j)\}$ in yellow.

Type II clique



Clique in yellow.

Wasserman & Pattison (1996) models for school kids

Definitions

$t_1(x)$	$=$	$\sum_{i,j} x_{ij}$	"choice"
$t_2(\mathbf{x})$	$=$	$\sum_{i,j} x_{ij}x_{ji}$	"mutuality"
$t_3(\mathbf{x})$	$=$	$\sum_{i,j,k} x_{ij}x_{jk}x_{ik}$	"transitivity"
$t_4(\mathbf{x})$	$=$	$\sum_{i,j,k} x_{ij}x_{jk}x_{ki}$	"cyclicity"
x_{i+}	$=$	$\sum_{j} x_{ij}$	"expansiveness" of i
x_{+i}	$=$	$\sum_{j} x_{ji}$	"attractiveness" of i
$t_5(\mathbf{x})$	$=$	$\sum_{i} x_{+i}^2$	"2–in–stars"
$t_6(\mathbf{x})$	$=$	$\sum_{i} x_{i+}^2$	"2–out–stars"
$t_7(\mathbf{x})$	$=$	$\sum_{i} x_{+i}x_{i+}$	"2–mixed–stars"

Differential "choice" also available ("block" models): e.g. girl–girl (GG)

Homogeneous and block–homogeneous models

Model 2: choice + mutuality (2 parameters)

Model 3: choice + mutuality + transitivity (3)

Model 4: choice + mutuality + cyclicity (3)

Model 10: choice + mutuality + transitivity + cyclicity (4)

Model 30: BB + BG + GB + GG choice + mutuality + transitivity (6)

Model 30h: BB + BG + GB + GG choice + mutuality + transitivity + cyclicity (7)

Individual–level models

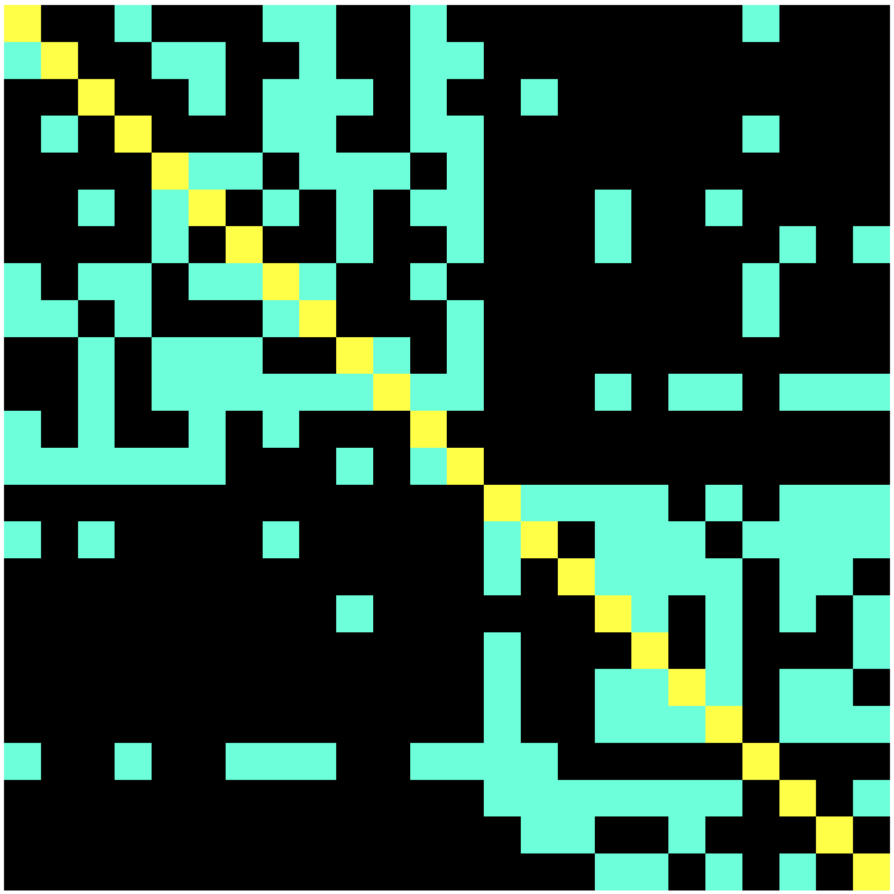
Model 18: BB/GG + BG/GB choice + mutuality + expansiveness + attractiveness

Model 23: BB/GG + BG/GB choice + mut + trans + expansiveness + attractiveness

Exact goodness–of–fit tests for class of 24 school kids

Child j

Child i

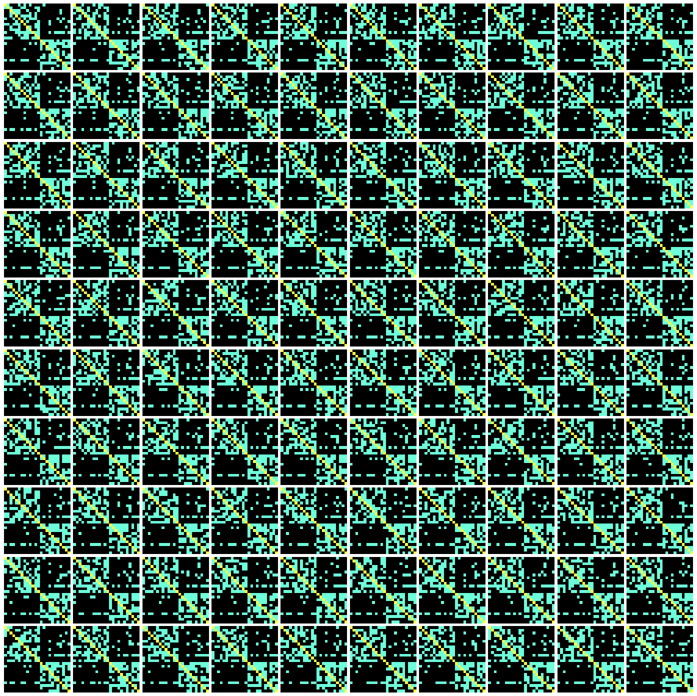


13 boys

11 girls

24 school children : pixel (i, j) is blue if i claims j is a friend.

Exact goodness–of–fit tests for 24 school kids

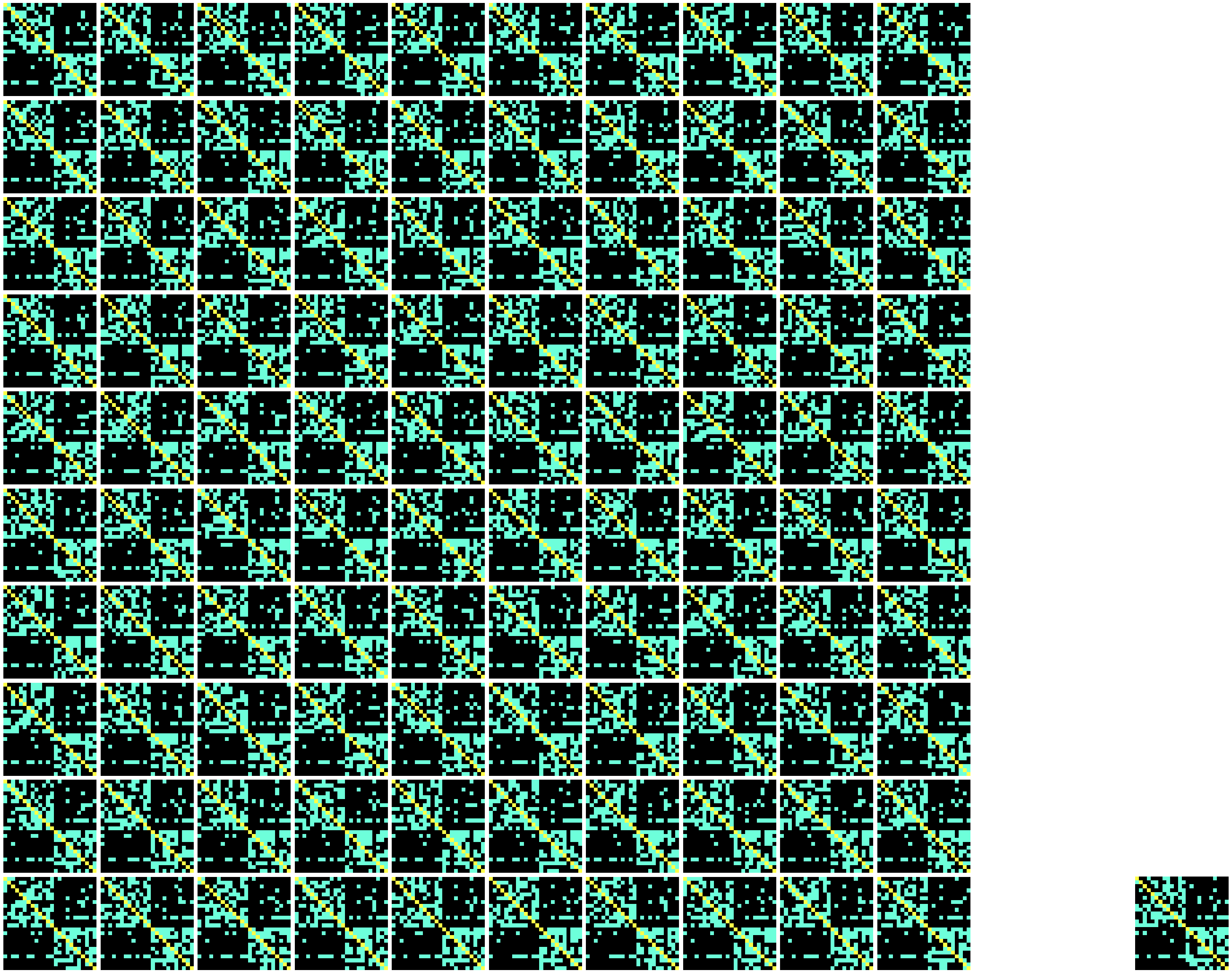


Model 18 + 3 mutualities, with 51 parameters.

BB/GG + BG/GB choice + mutuality + expansiveness + attractiveness.

100 of 1000 realizations with same 51 image statistics, of which one is the data.

Two–sided p–values 0.002 & 0.004 based on t_3 (transitivity) & t_4 (cyclicity).



data