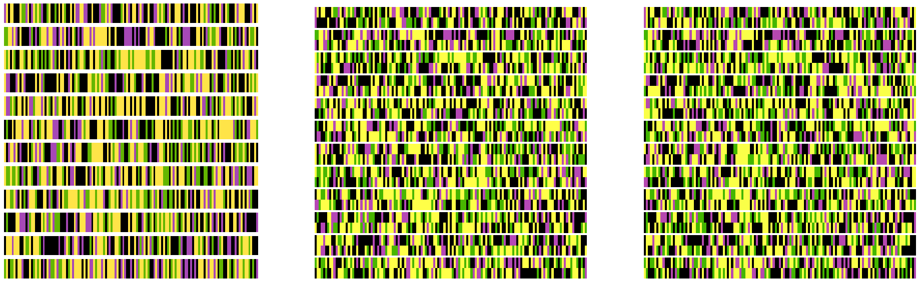
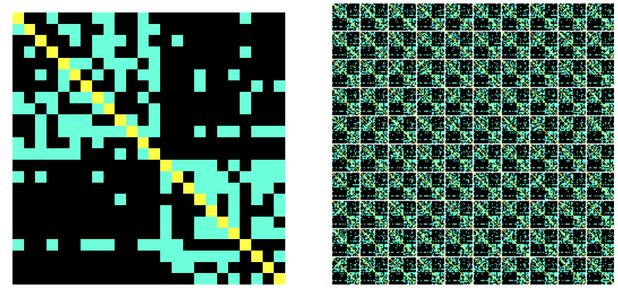
Markov chains for DNA sequences



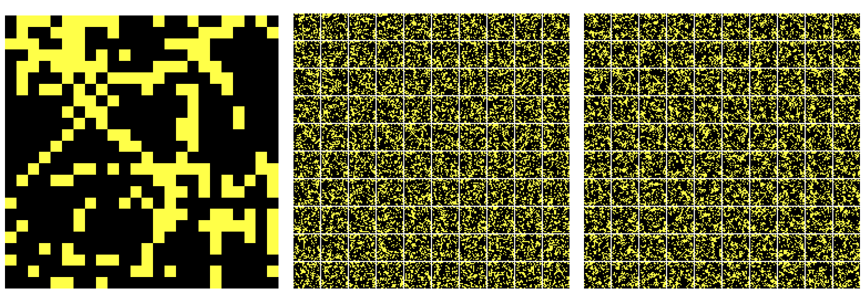
Markov random graphs for social networks



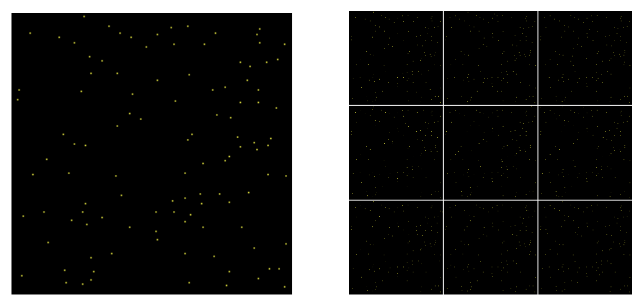
Exact Goodness–of–Fit Tests and Some Applications

JULIAN BESAG

Department of Statistics, University of Washington, Seattle, USA



Markov random fields in statistical ecology



Markov point processes

Outline

Simple Monte Carlo p–values.

Examples.

Markov chain Monte Carlo (MCMC) p–values.

Examples.

Green's method for irreducibility.

Is there any role for MC and MCMC tests in model choice and validation??

Frequentist goodness–of–fit tests

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

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

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 particular **test statistic** $u = u(x)$.

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

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

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

Suppose can draw **random sample** $x^{(2)}, \ldots, x^{(m)}$ 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 $m \quad \Rightarrow \quad$ 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. Testing for independence in contingency tables

$x^{(1)}$ is observed $r \times s$ contingency table

Standard multinomial assumptions, with unknown probability θ_{ij} in cell (i, j).

Test for independence of row and column categorizations: $\theta_{ij} = \phi_i \psi_j$.

Conditioning on row and column totals to eliminate ϕ_i's and the ψ_j's $\Rightarrow$

$$\pi(x) = \frac{\prod_i x_{i+}! \prod_j x_{+j}!}{x_{++}! \prod_i \prod_j x_{ij}!}, \qquad x \in S,$$

where S is corresponding constrained space.

Generate random sample $x^{(2)}, \ldots, x^{(m)}$ from π, use any test statistic $u(x)$, ...

OK for testing independence in **multiway** tables. Sparseness OK.

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

Ex. Goodness of fit for Markov chains

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

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

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

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

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.

Notes on goodness–of–fit tests for Markov chains

Standard asymptotics do not apply to short sequences or multiple short sequences.

Ordinary Monte Carlo test requires sampling from uniform distribution on $\mathcal{S}$.

Easy for binary sequences.

MCMC tests require proposals that maintain the n_{ij}'s and move freely around $\mathcal{S}$.

Candidates include **swaps** between **single elements** or **subsequences**.

In fact, **ordinary Monte Carlo** test is **always** available ... and very fast !!

Draw samples from $\mathcal{S}$ using **Euler tours**; see Aldous (1990), Kandel et al (1996).

Applies also to tests for **higher–order chains**, reformulating as 1st–order.

Can compare **asymptotics** with **exact** values.

Binary data for 77 schizophrenics

Presence / absence of a particular response over 12 months

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 ?

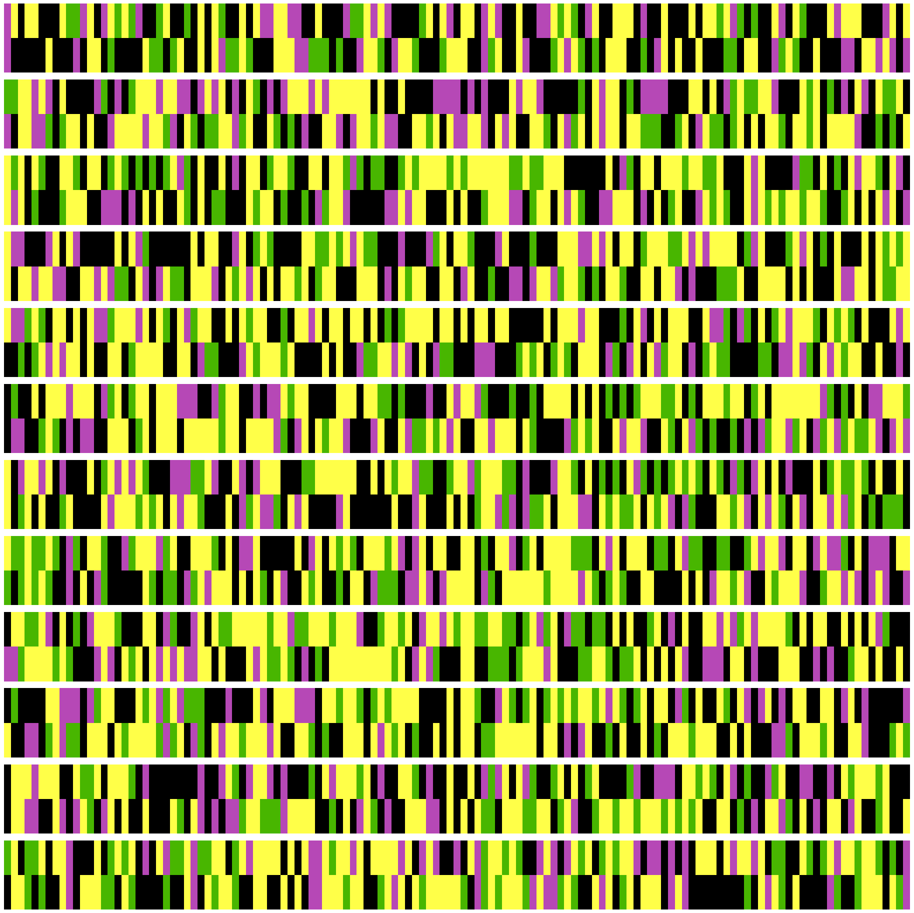
Deviance v. 2nd–order chains = 47.32 on 154 d.fr. but ordinary Monte Carlo and MCMC (proposing swaps within rows) p–values = 0.03.

DNA sequence data Avery & Henderson (1999)

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

DNA sequence data

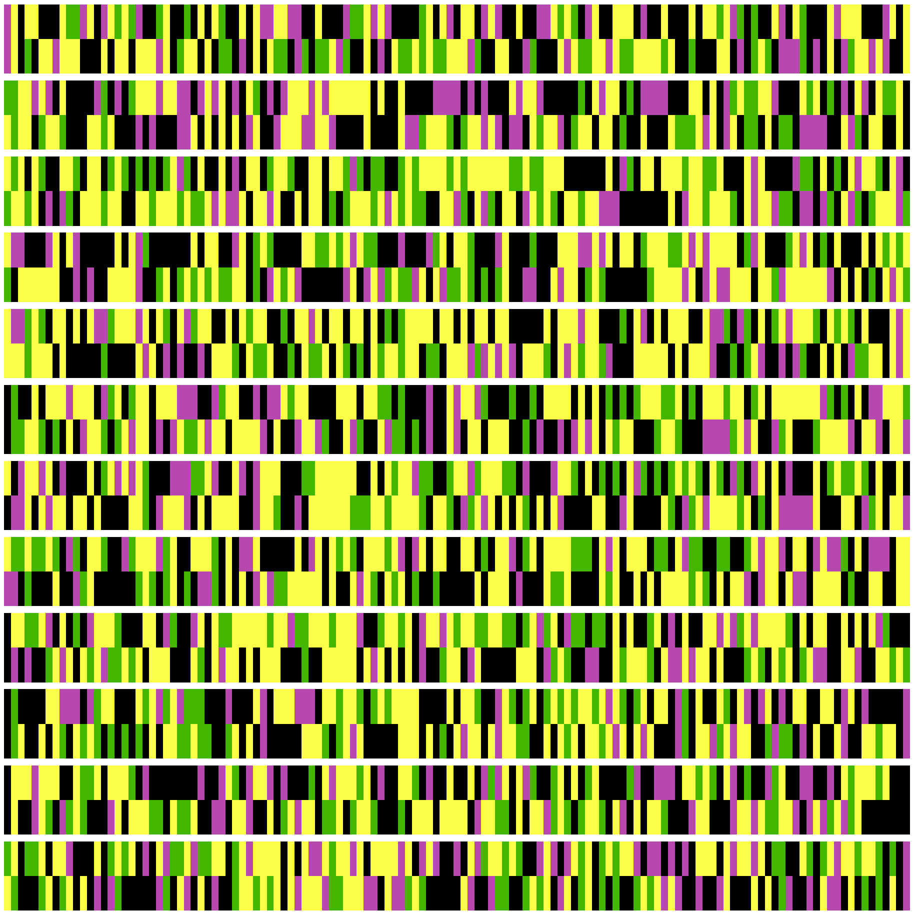
Data and sample with same frequencies of pairs.



Deviance = 55.66 on 36 d.fr. χ^2 : 0.02, MCMC : 0.03

DNA sequence data

Data and sample with same frequencies of triples.



Deviance = 150.31 on 144 d.fr. χ^2 : 0.34, MCMC : 0.43

However, what if random sampling from $\{\pi(x) : x \in \mathcal{S}\}$ is not available?

e.g. Rasch model, hierarchical models for k–way tables, random graph models, ...

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

Metropolis algorithm

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

Consider *any* **symmetric transition matrix** R with elements $R(x, x')$

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

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

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

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

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

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

The Times : daily Sudoku puzzle ("fiendish")

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 : S = set of all 9×9 tables x with feasible subtables.

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

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

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

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

Algorithm :

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 $v(x) = 0$ is reached.

Ex. LSAT data from Boch & Lieberman (1970, Psychometrika)

Two datasets, each with responses to 5 questions from 1000 individuals.

Correct (1) or Incorrect (0) — Frequencies

Q1	Q2	Q3	Q4	Q5	I	II
0	0	0	0	0	3	12
0	0	0	0	1	6	19
0	0	0	1	0	2	1
0	0	0	1	1	11	7
0	0	1	0	0	1	3
0	0	1	0	1	1	19
0	0	1	1	0	3	3
0	0	1	1	1	4	17
0	1	0	0	0	1	10
0	1	0	0	1	8	5
0	1	0	1	0	0	3
0	1	0	1	1	16	7
0	1	1	0	0	0	7
0	1	1	0	1	3	23
0	1	1	1	0	2	8
0	1	1	1	1	15	28
1	0	0	0	0	10	7
1	0	0	0	1	29	39
1	0	0	1	0	14	11
1	0	0	1	1	81	34
1	0	1	0	0	3	14
1	0	1	0	1	28	51
1	0	1	1	0	15	15
1	0	1	1	1	80	90
1	1	0	0	0	16	6
1	1	0	0	1	56	25
1	1	0	1	0	21	7
1	1	0	1	1	173	35
1	1	1	0	0	11	18
1	1	1	0	1	61	136
1	1	1	1	0	28	32
1	1	1	1	1	298	308

Rasch model : overall coincidences as test statistic $\Rightarrow$ p–values 0.857 & 0.004.

Ex. Darwin's finches

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

$x^{(1)}$ 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 $\theta_{ij} = \phi_i \, \psi_j \quad \Rightarrow$

Table has probability

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

Conditioning on row and column totals eliminates ϕ_i's and ψ_j's

$\Rightarrow$ uniform distribution $\pi(x)$ on constrained space 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 :

$$\begin{array}{ccccc} & \vdots & & \vdots & \\ \cdots & a & \cdots & b & \cdots \\ & \vdots & & \vdots & \\ \cdots & b & \cdots & a & \cdots \\ & \vdots & & \vdots & \end{array} \quad \rightarrow \quad \begin{array}{ccccc} & \vdots & & \vdots & \\ \cdots & b & \cdots & a & \cdots \\ & \vdots & & \vdots & \\ \cdots & a & \cdots & b & \cdots \\ & \vdots & & \vdots & \end{array}$$

where a and $b = 0$ or 1. Can prove **irreducibility**.

E.g. choose **two rows** and **two columns** at random, propose corresponding **swap** and **accept swap if valid** else retain existing configuration. OK **Metropolis**.

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

Exact Markov chain Monte Carlo p–values

Besag (1983)

Besag & Clifford (1989)

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

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

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

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

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

Need another fix ...

Note that corresponding **backwards** transition matrix Q has (x', x) element

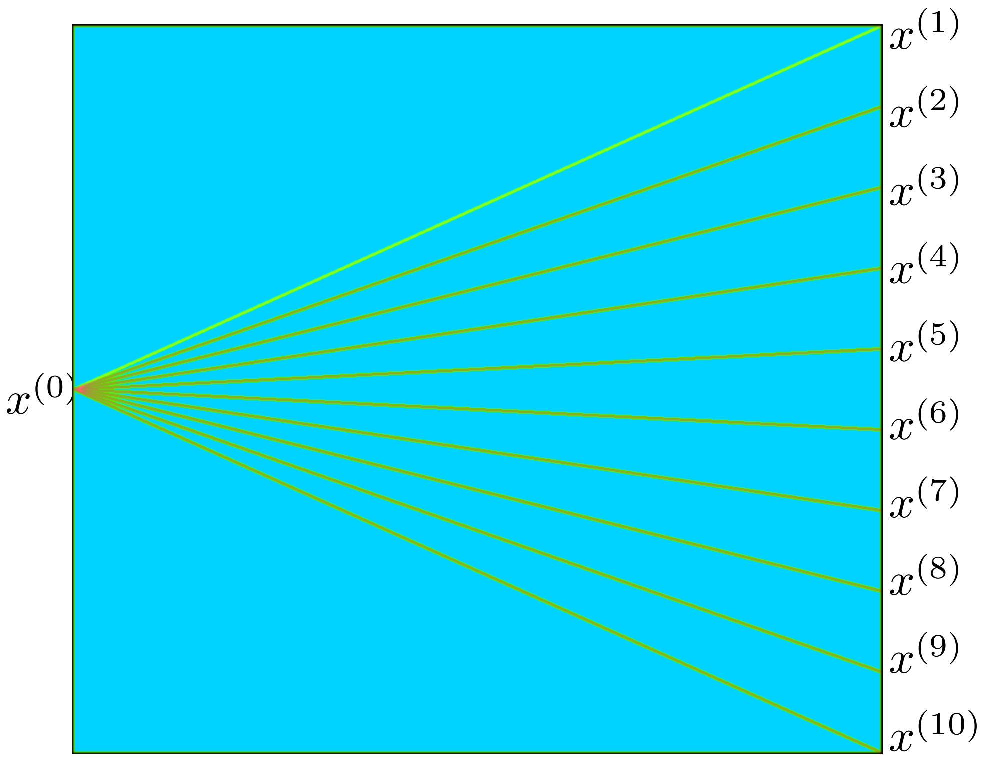
$$Q(x', x) = \pi(x)\, P(x, x') \,/\, \pi(x')$$

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 $x^{(1)} \Rightarrow x^{(0)}$.

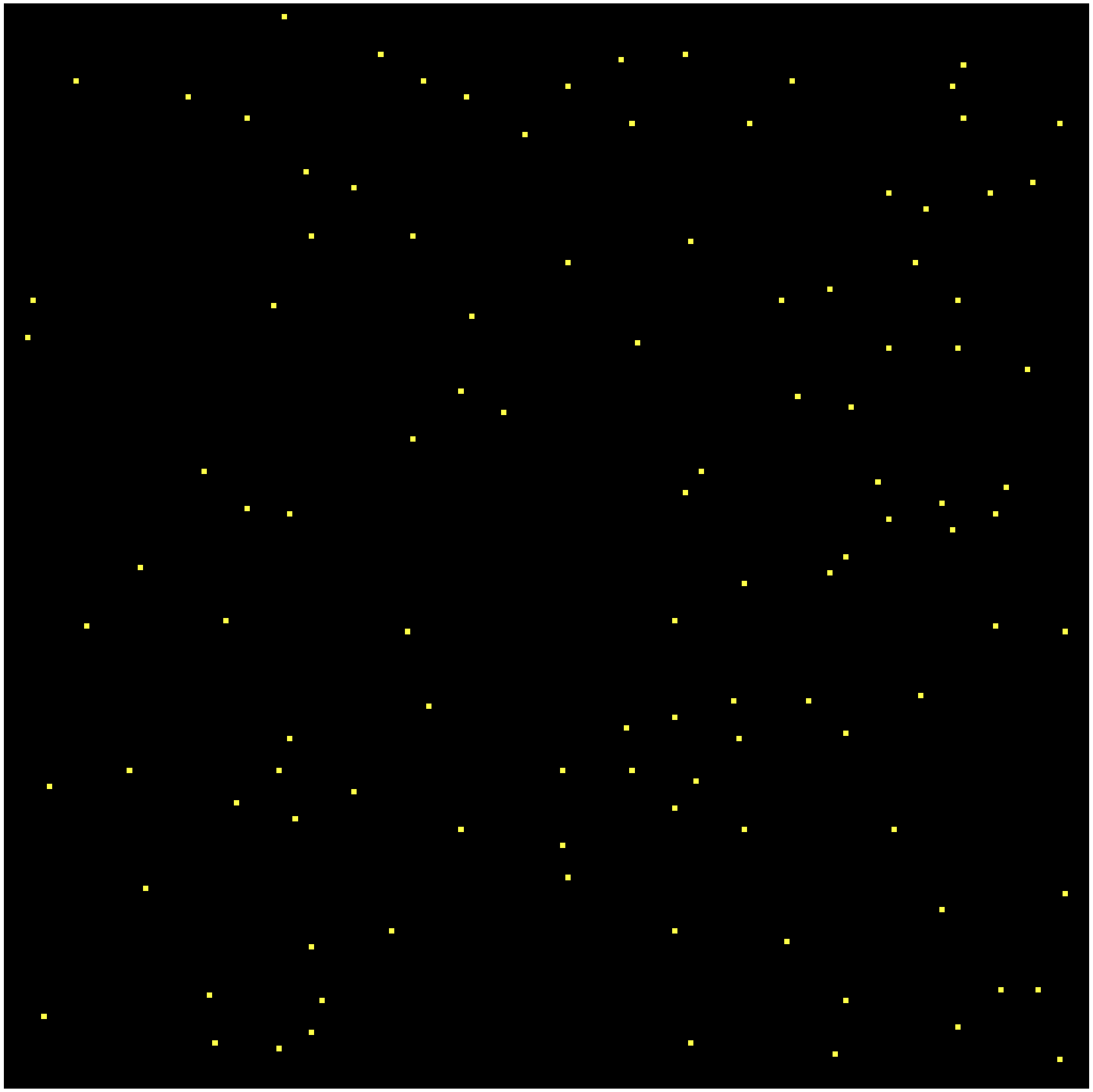
Use P to run chain forwards t steps from $x^{(0)}$, $m-1$ times $\Rightarrow x^{(2)}, \ldots, x^{(m)}$.



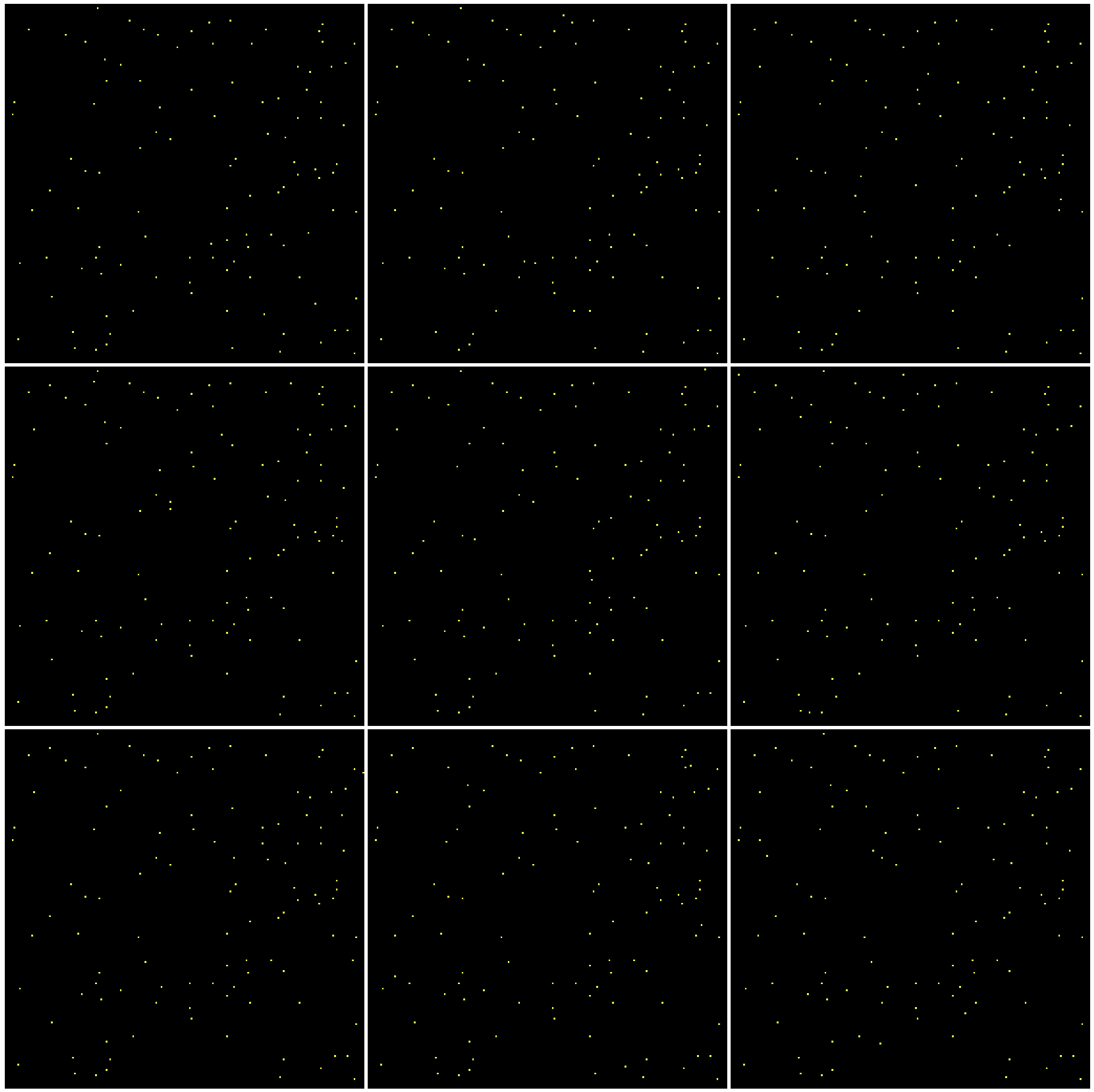
$\mathcal{H}_0 \Rightarrow x^{(1)}, \ldots, x^{(m)}$ are **exchangeably** from π; and so are $u^{(1)}, \ldots, u^{(m)}$.

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

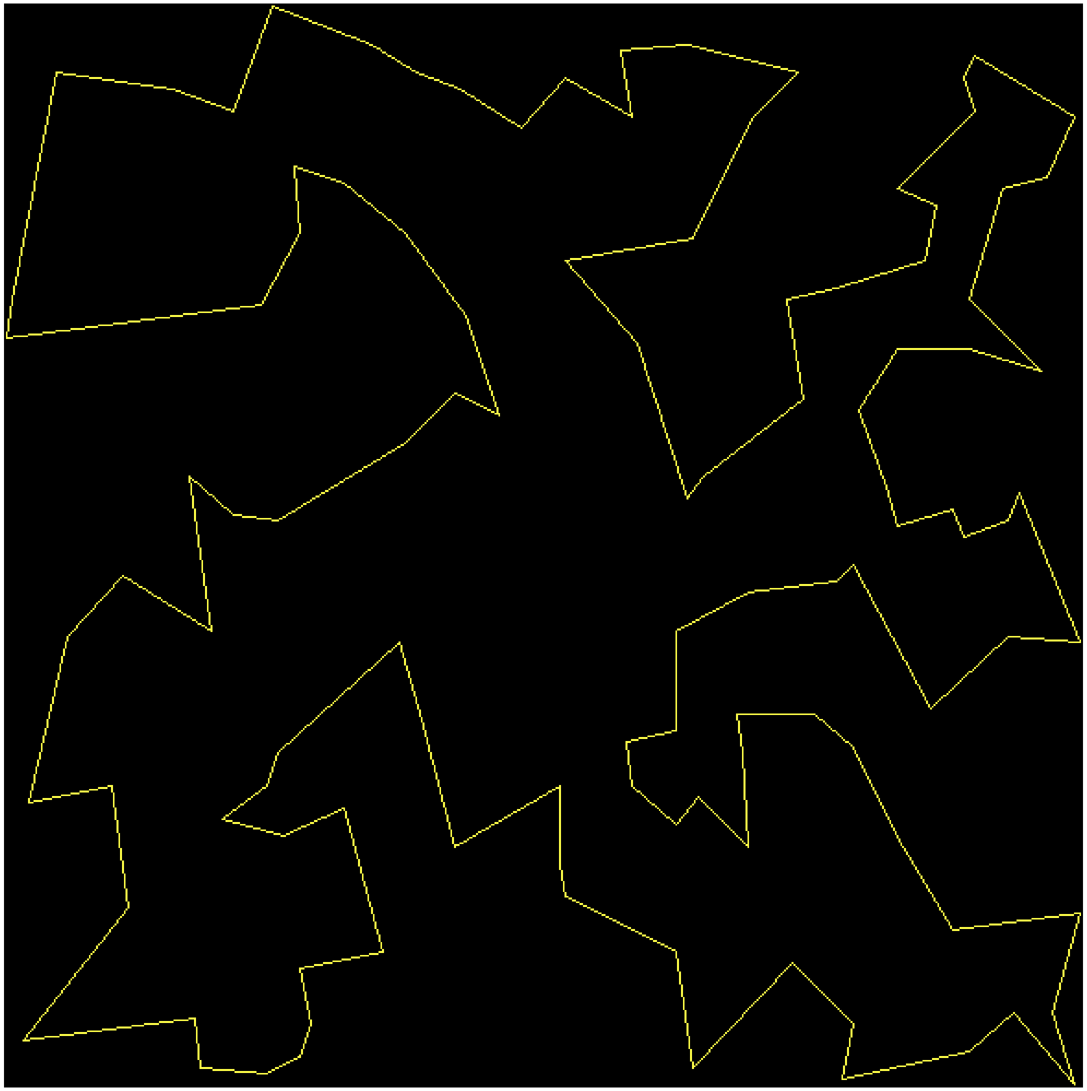
Ex. Nesting sites of 110 Gray Gulls in Northern Chile



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

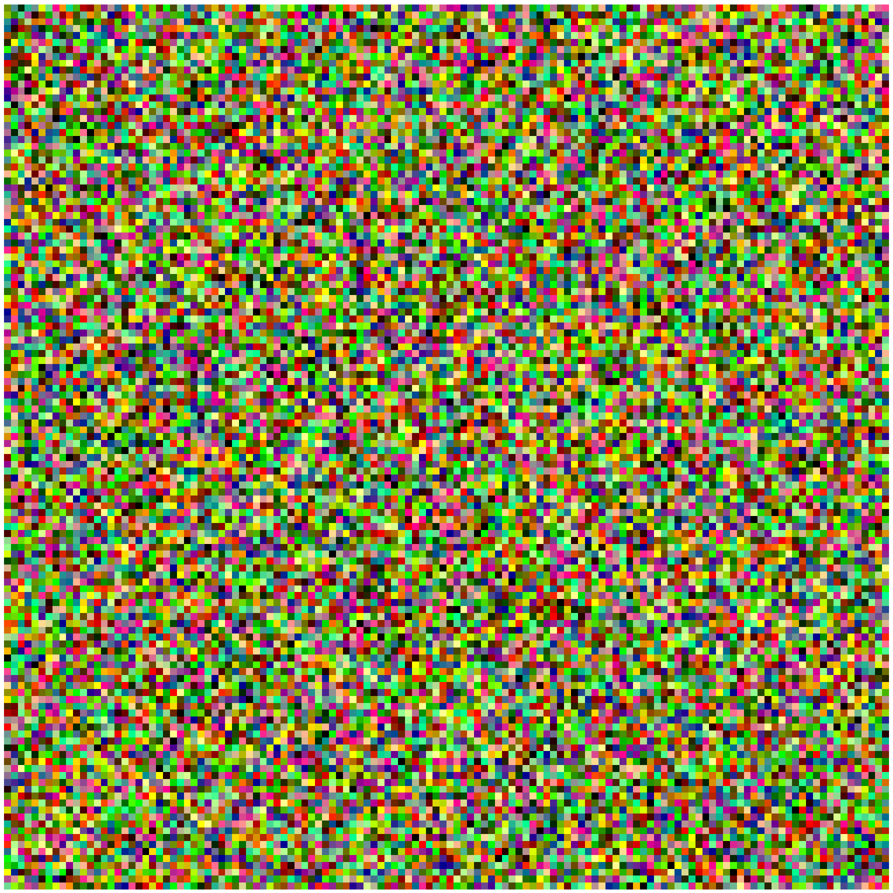


Gray Gulls: a fast route around the nests



Constructing large Latin squares

Can be done similarly ...



128×128 **Latin square**

Smoking and Lung Cancer in China

Agresti (1996)

			Lung Cancer?		Odds Ratio
			Y	N	
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	

General Metropolis algorithm for testing k–way tables

To test appropriateness of a hierarchical model for a k–way table:

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

NB. If x^* has any negative entries, 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:

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

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

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

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.

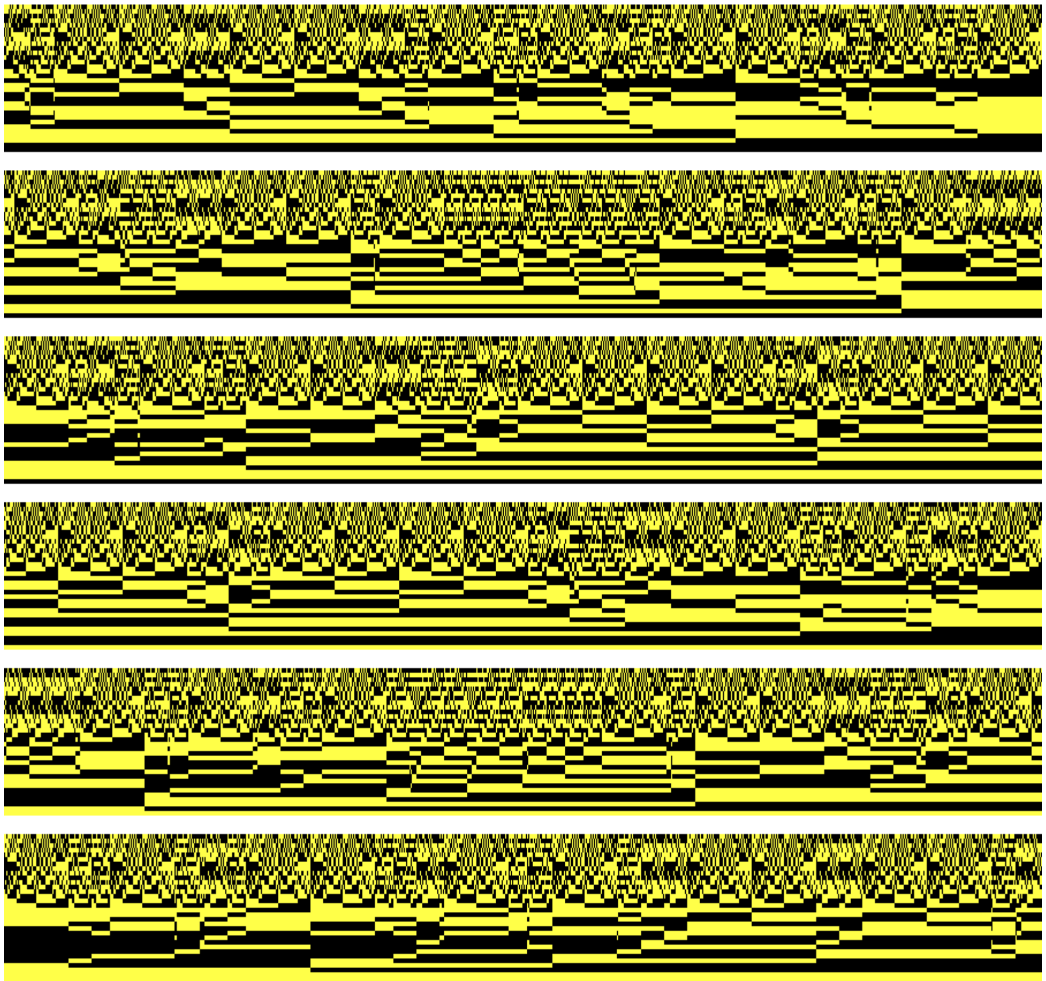
Explanatory variables: Gender (G) & Race (R).

	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

Model 6 allows pairwise interactions AC, AM, AG, AR, CM, MG, GR :

Deviance = 19.91 on 19 d.fr. $\left\{ \begin{array}{lcl} \text{Conventional p–value} & = & 0.40 \\ \text{MCMC exact p–value} & = & 0.19 \end{array} \right.$

32



5406 proposals

black = +1

yellow = −1

Model 7 allows pairwise interactions AC, AM, AG, AR, CM, GR :

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

Model 7^{+} is the graphical (conditional independence) model ACM, AGR :

$$\text{Deviance} = 26.33 \text{ on } 18 \text{ d.fr.} \qquad \begin{cases} \text{Conventional p–value} & = & 0.09 \\ \text{MCMC exact p–value} & = & 0.09 \\ \text{Simple MC exact p-value} & = & 0.08 \end{cases}$$

NB. Model 7, adding 1 to all observations :

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

Markov random fields

Observations x_i on "arbitrary" r.v.'s X_i for $i = 1, \ldots, n$.

Definition. A **cliquo** is a set of sites all of which are mutual **neighbors**.

Hammersley–Clifford theorem $\Rightarrow$

$$\pi(x) \,/\, \pi(0) \;=\; \exp\{\textstyle\sum_{A \in \mathcal{C}} (\prod_{i \in A} x_i)\, G_A(x_A)\}$$

where $\mathcal{C}$ is the set of all cliquos.

Markov random fields for binary data

Observations $x_i = 0$ or 1 on r.v.'s X_i for $i = 1, \ldots, n$.

so $G_A(x_A) = G_A(1_A)$ w.l.o.g. and

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

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

Markov random fields for binary data

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

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

Homogeneity assumptions produce schemes of the form

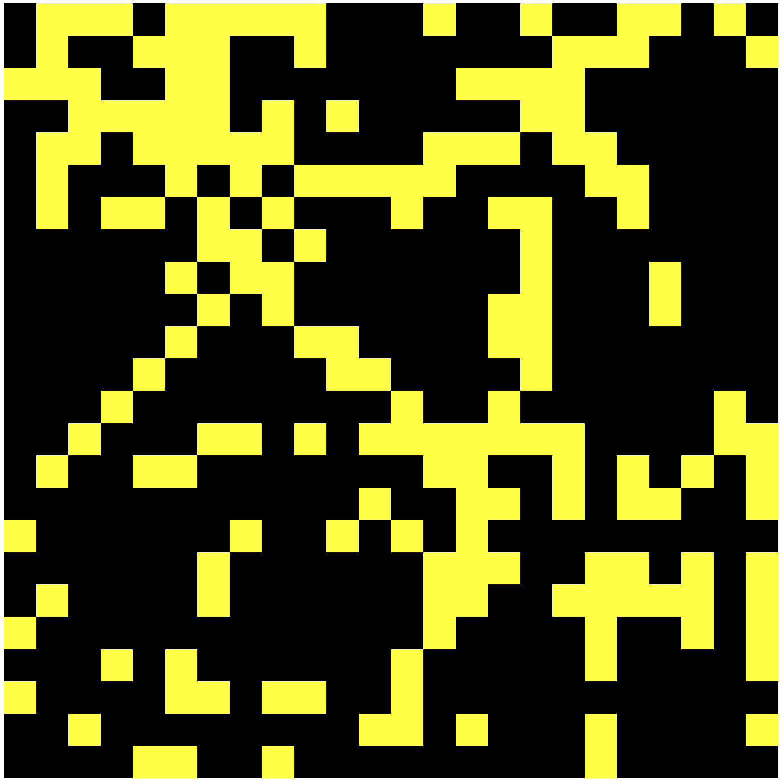
$$\pi(x;\theta) \;=\; \frac{\exp\{\theta_1 t_1(x) + \ldots + \theta_q t_q(x)\}}{c(\theta)}$$

where $\theta_1, \ldots, \theta_q$ are parameters, $c(\theta)$ is a normalizing constant and

$t_1(x), \ldots, t_q(x)$ are **sums of products–over–cliquos** of the x_i's.

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

Presence (yellow = 1), absence (black = 0) of Carex arenaria



Q. Can we find a model for which the interior data are a typical realization?

Autologistic models on rectangular arrays $\quad i = (r, s)$

Homogeneous 1st–order autologistic model

$$\pi(x; \alpha, \beta) \;\propto\; \exp\left(\alpha \textstyle\sum_{r,s} x_{r,s} + \beta_1 \sum_{r,s} x_{r,s}\, x_{r,s+1} + \beta_2 \sum_{r,s} x_{r,s}\, x_{r+1,s}\right)$$

No higher–order terms possible.

Fully–symmetric Ising model

$$\pi(x; \beta) \;\propto\; \exp\{\beta\, t(x)\}, \quad \text{where} \;\; t(x) = \text{ number of like–valued direct adjacencies.}$$

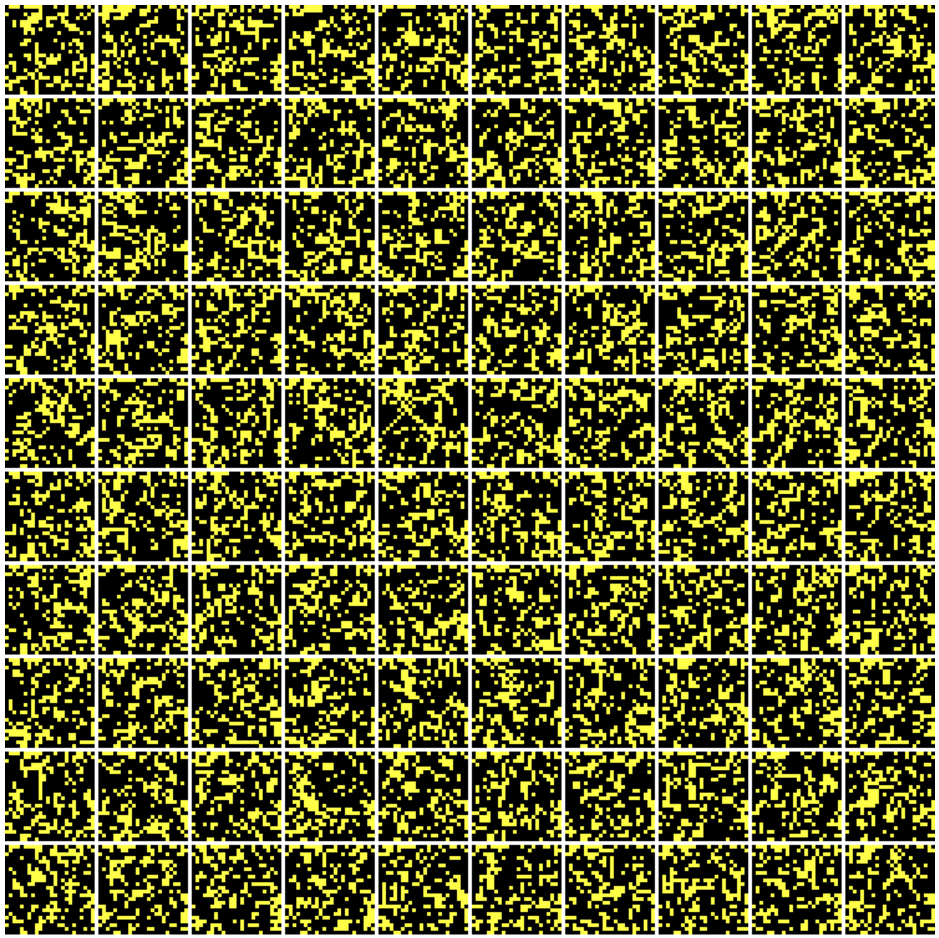
Homogeneous 2nd–order autologistic model

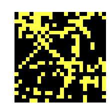
$$\pi(x; \alpha, \beta, \gamma) \;\propto\; \exp\big(\alpha \textstyle\sum_{r,s} x_{r,s} + \beta_1 \sum_{r,s} x_{r,s}\, x_{r,s+1} + \beta_2 \sum_{r,s} x_{r,s}\, x_{r+1,s}$$

$$+\gamma_1 \textstyle\sum_{r,s} x_{r,s}\, x_{r+1,s+1} + \gamma_2 \sum_{r,s} x_{r,s}\, x_{r+1,s-1}\big)$$

Can allow further terms for cliquos of triples and quadruples.

MCMC goodness of fit : 100 samples from 1st–order autologistic scheme ?

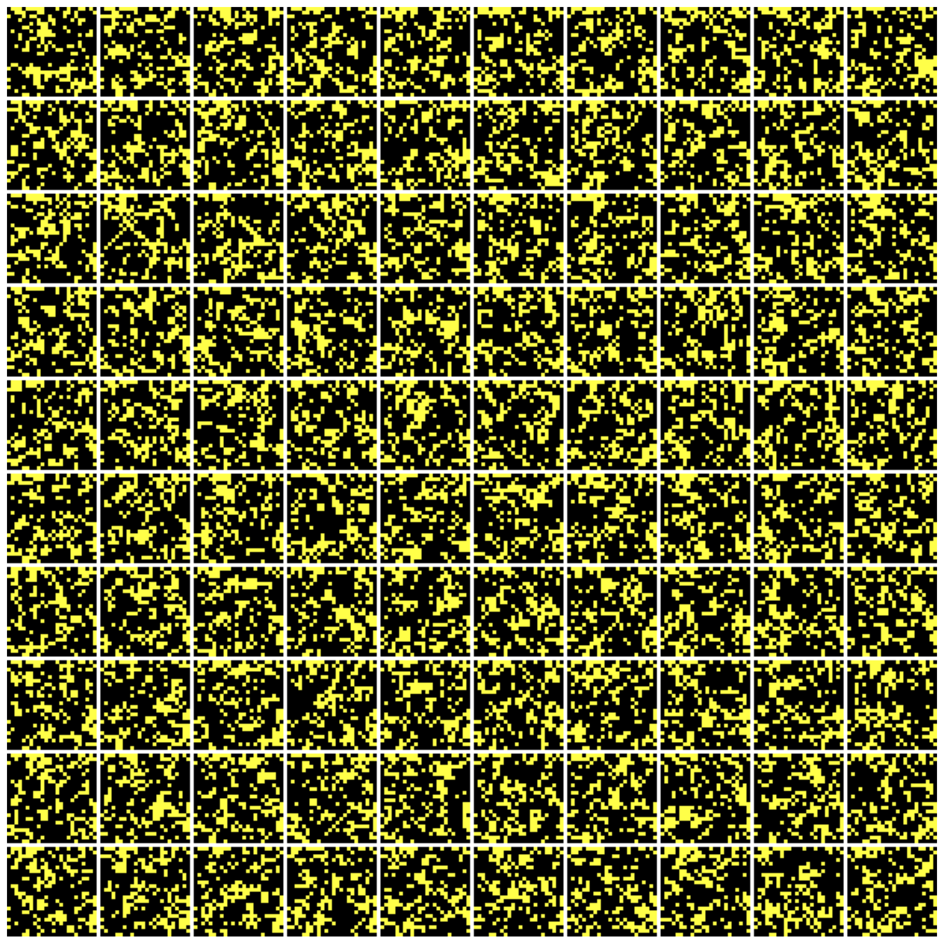


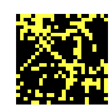


data

p–value based on horizontal, vertical & diagonal 111's = 0.0013 to 0.0016

MCMC goodness of fit : 100 samples from 2nd–order autologistic scheme ?





data

p–value based on horizontal, vertical & diagonal 111's = 0.0149 to 0.0226

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.

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

Markov property of Frank & Strauss, 1986

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

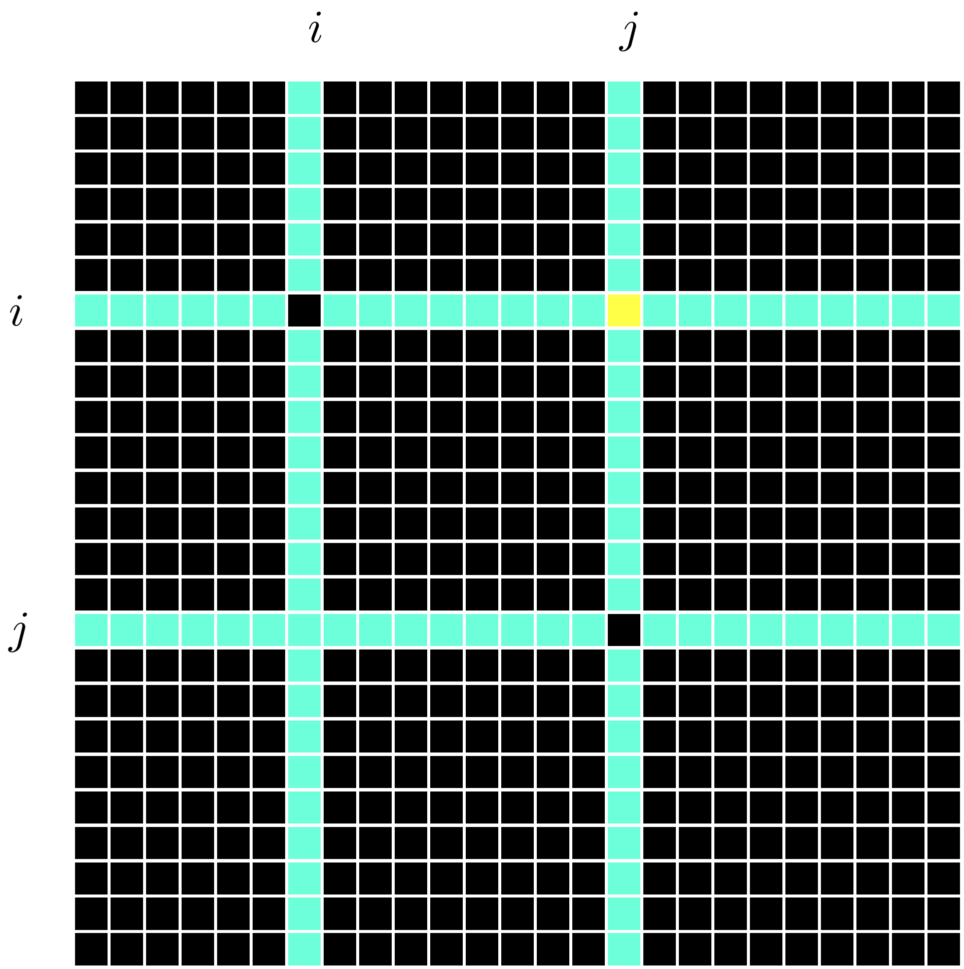
e.g. X_{12} is conditionally independent of X_{34}

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

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

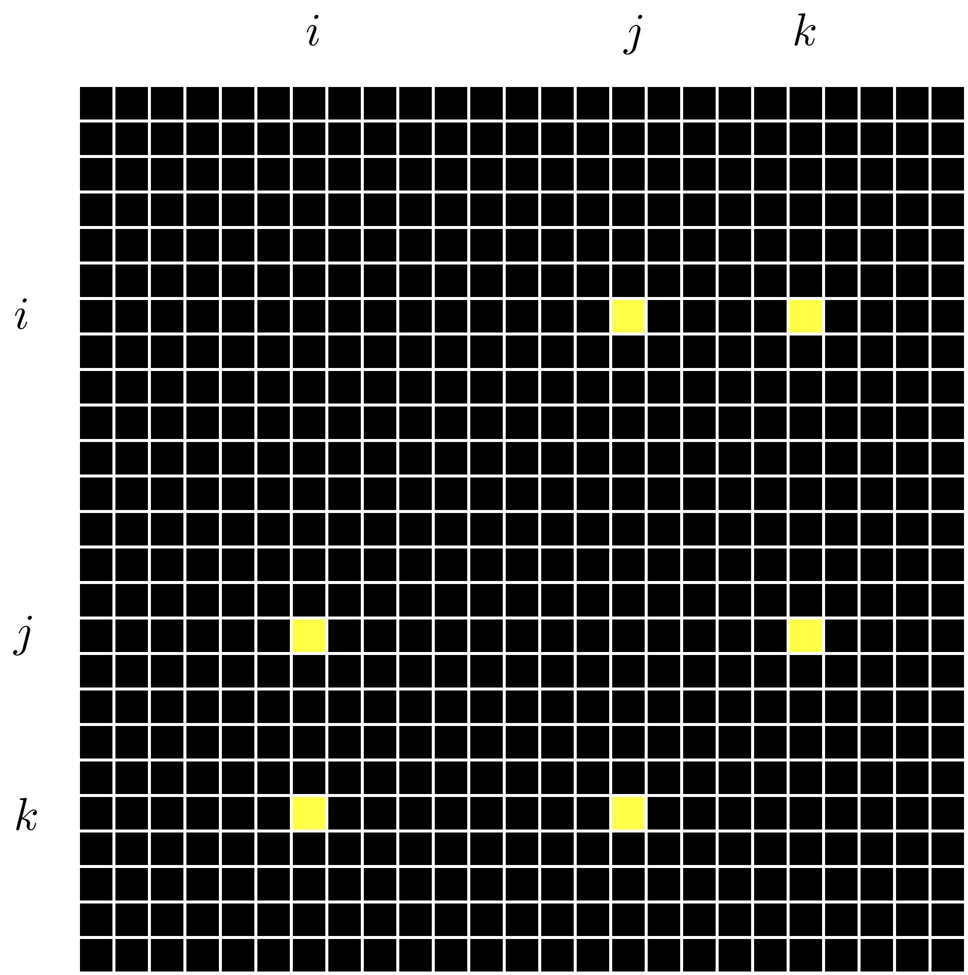
Type II : $\{\,(i,j),\,(j,i),\,(i,k),\,(k,i),\,(i,l),\,(l,i),\,\ldots\}$ $\qquad i, \, j, \, k, \, l, \, \ldots$ distinct

Markov property



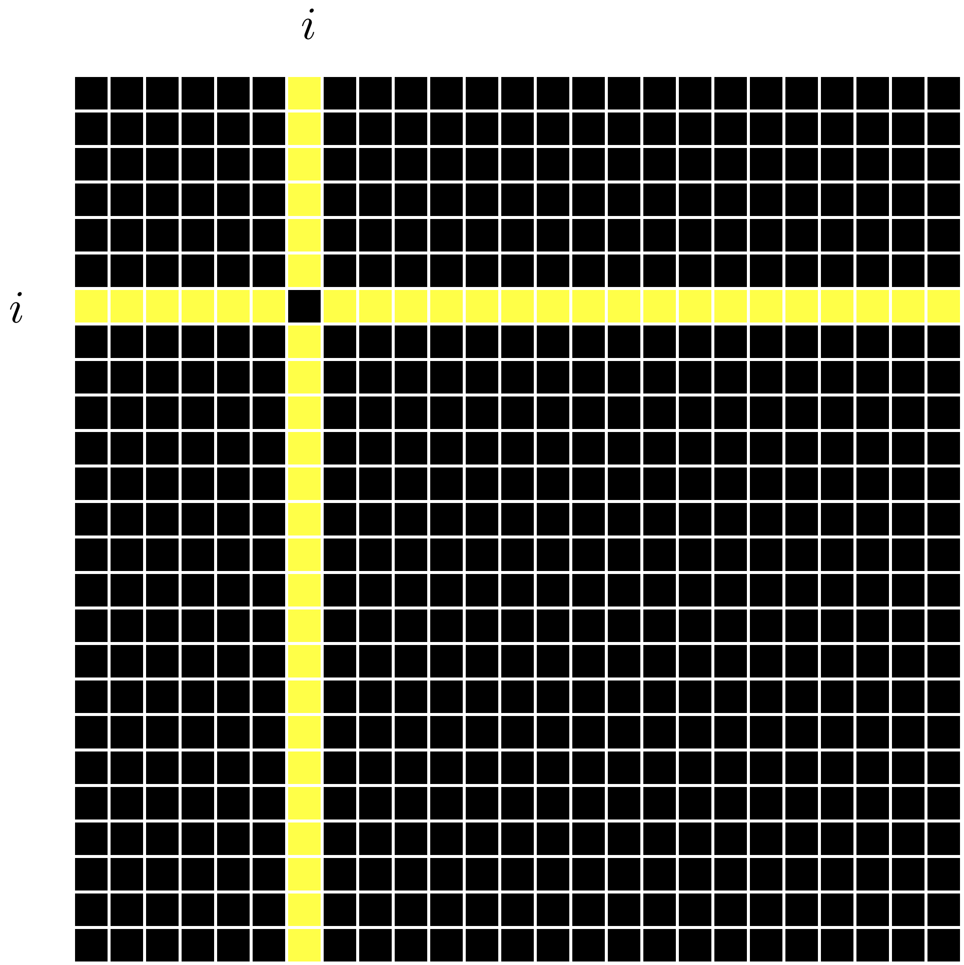
Neighbors of yellow site (i, j) in blue.

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

Social network for 24 school kids

13 boys, 11 girls.

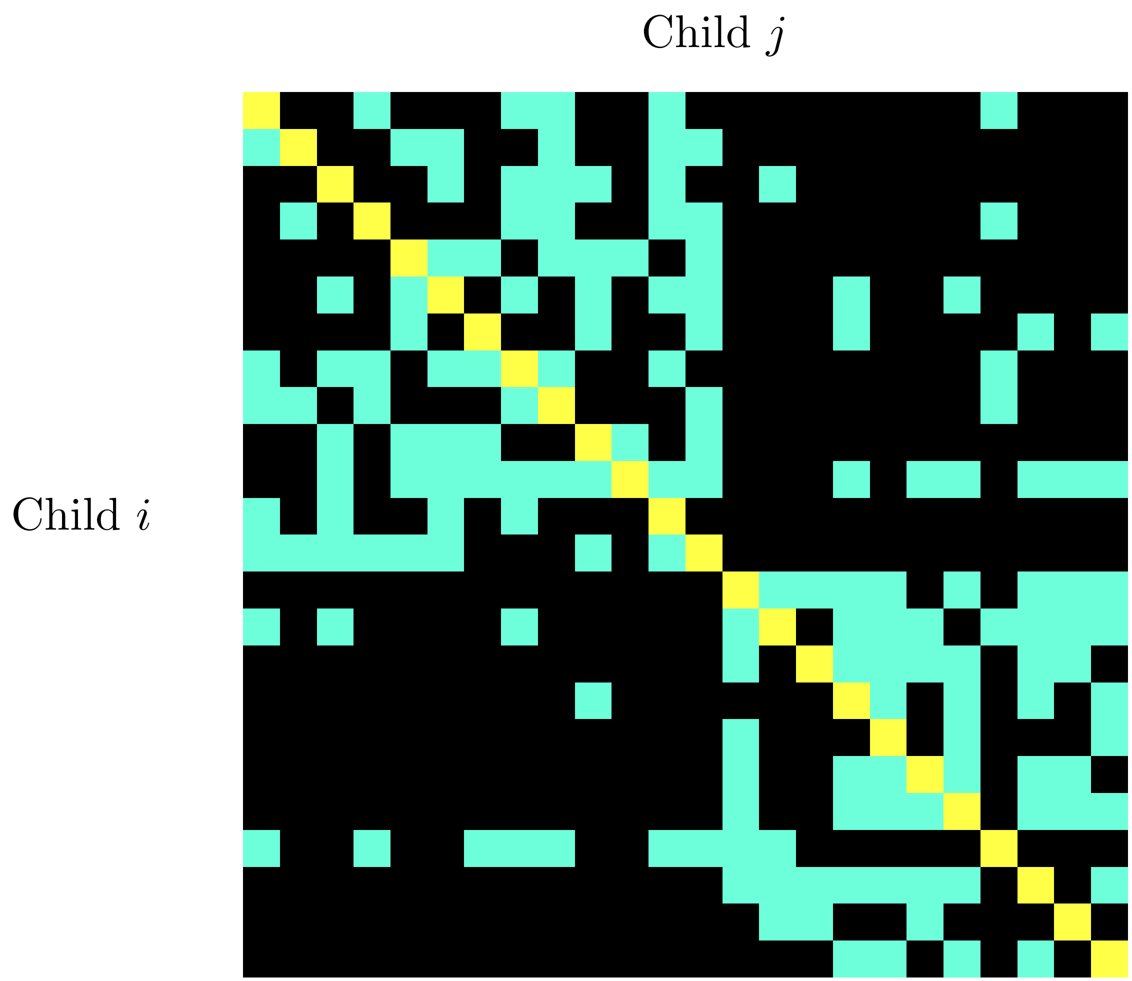
$X_{ij} = 1$ if i claims j is a friend, else $X_{ij} = 0$.

j

i

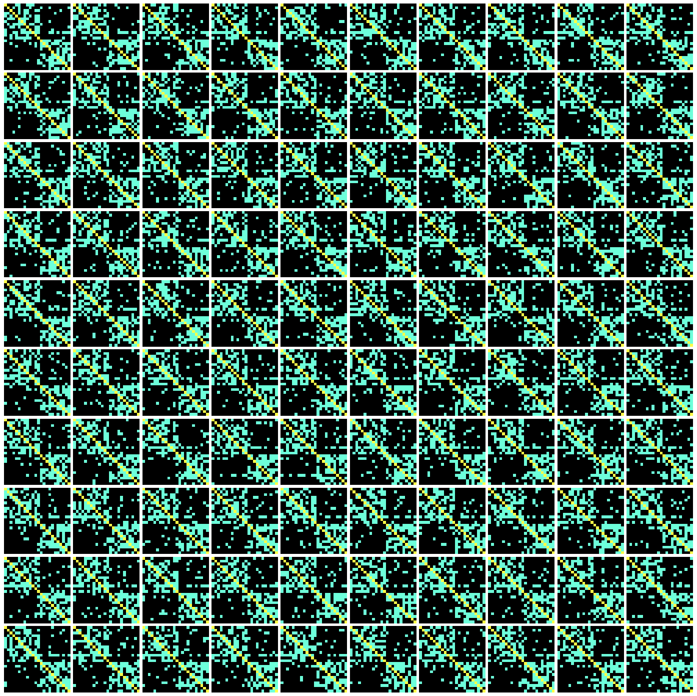
	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
1		0	0	1	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
0	0		0	0	1	0	1	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0
0	1	0		0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0		1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	1		0	1	0	1	0	1	1	0	0	0	1	0	0	1	0	0	0	0
0	0	0	0	1	0		0	0	1	0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

Exact goodness–of–fit tests for 24 school kids



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

Exact goodness–of–fit tests for 24 school kids

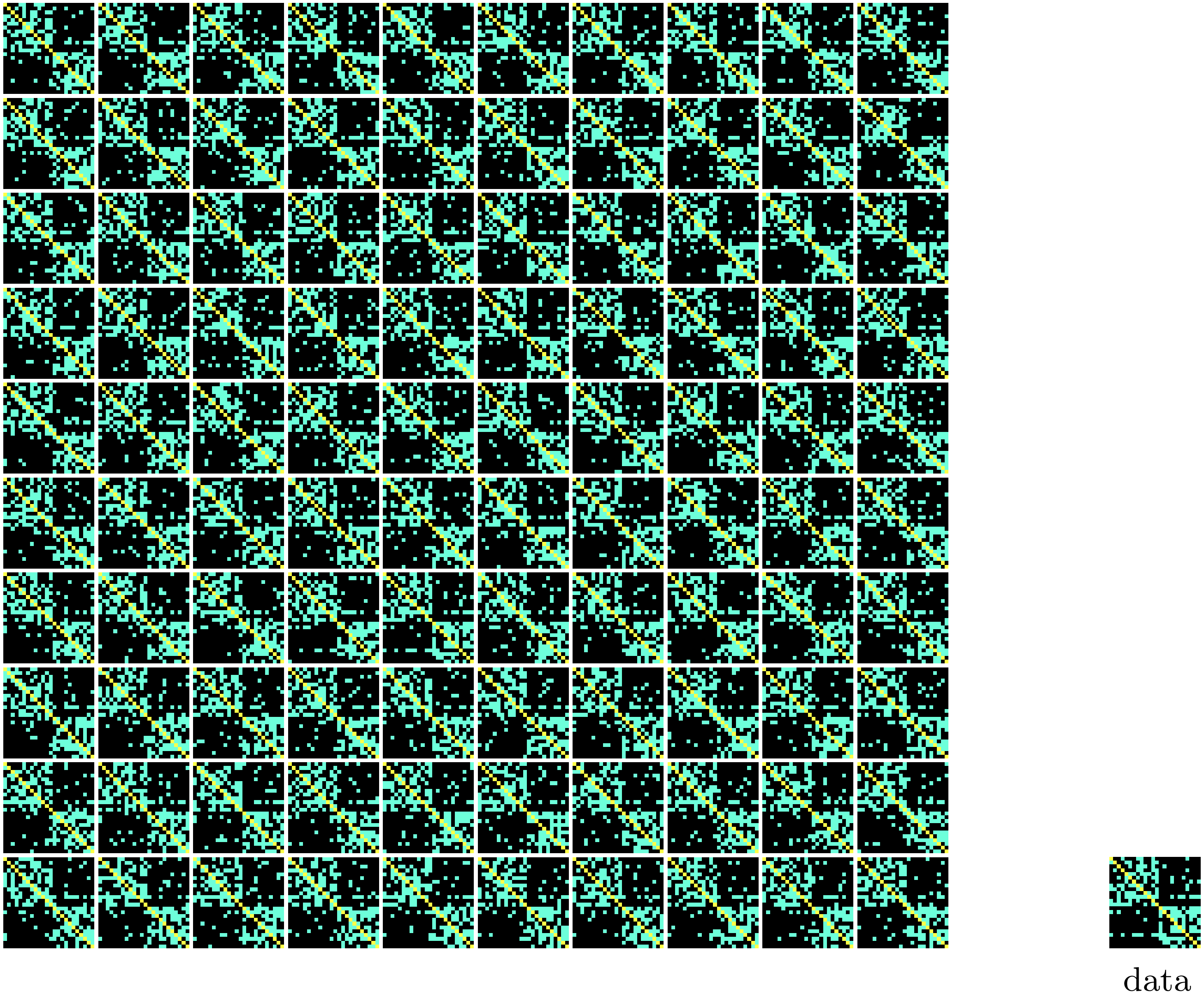


Model 18 : Markov random graph formulation with 49 parameters.

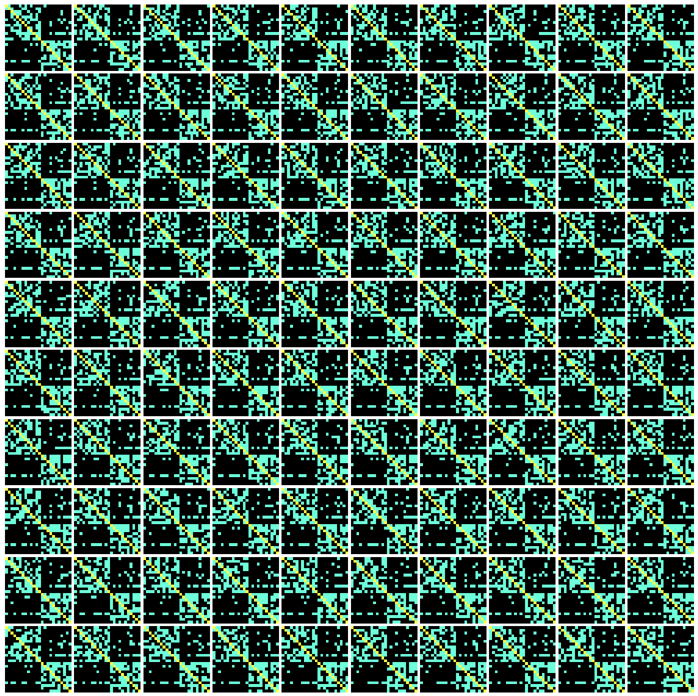
BB/GG + BG/GB choice + mutuality + expansiveness + attractiveness.

100 of 1000 realizations with same 49 image statistics, of which one is the data.

Two–sided p–value 0.002 based on either t_3 (transitivity) or t_4 (cyclicity).


data

Exact goodness–of–fit tests for 24 school kids

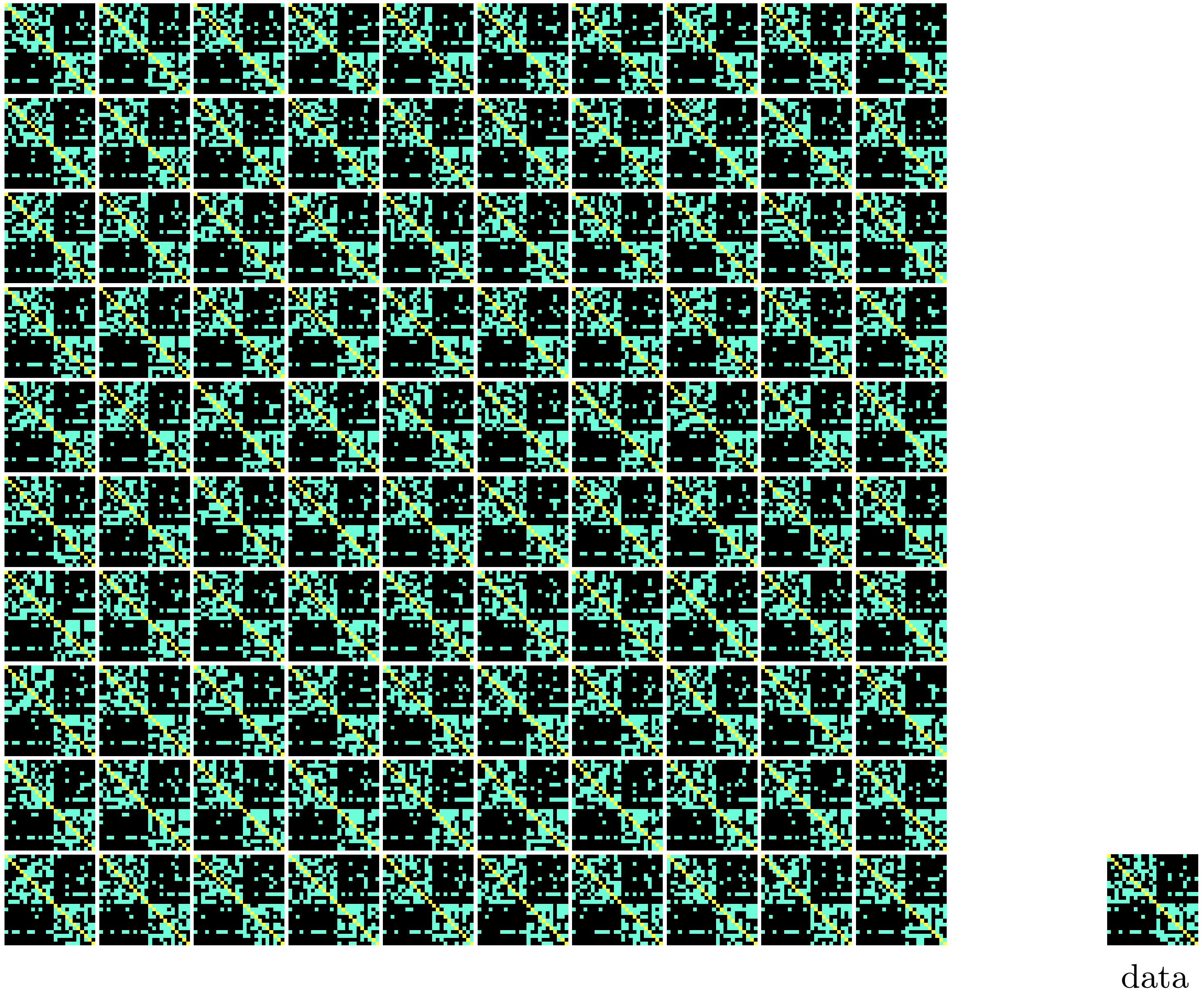


Model 18 + 3 mutualities, with 51 parameters.

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

Irreducibility in MCMC goodness of fit tests

Exact MCMC p–values do **not** require irreducibility for validity !!

Toy example : Testing for 1st–order autologistic scheme

Elementwise swaps usually OK : reject all proposals that change like–like adjacencies.

Not necessarily irreducible : e.g. if boundary is fixed :

0	0	0	0		0	0	0	0
0	0	1	0	cannot be	0	1	0	0
0	1	0	0	obtained from	0	0	1	0
0	0	0	0		0	0	0	0

Irreducibility is generally preferable (though "mobility" is more important).

Green's (1986) method for irreducibility

Target distribution $\{\pi(x) : x = (x_1, \ldots, x_n) \in \mathcal{S}\}$, where $\mathcal{S}$ is **minimal**.

Corresponding **product space** $\mathcal{S}^* = \mathcal{S}_1 \times \ldots \mathcal{S}_n$, where $\mathcal{S}_i$ is minimal for X_i.

Define new distribution :

$$\pi^*(x) \;\propto\; \pi(x) \exp\{-\lambda \,|\, t(x) - t(x^{(1)}) \,|\}, \qquad x \in \mathcal{S}^*,$$

where $\lambda > 0$ is fixed and $t(x)$ is any particular statistic.

Run MCMC algorithm for $\{\pi^*(x) : x \in \mathcal{S}^*\}$, deleting all samples x with $x \notin \mathcal{S}$

$\Rightarrow$ thinned samples are from Markov chain with limit distribution $\{\pi(x) : x \in \mathcal{S}\}$.

Must choose λ carefully : if too large, rarely leave $\mathcal{S}$; if too small, rarely hit $\mathcal{S}$.

Might use intermediate sample space $\mathcal{S}^\dagger$ s.t. $\mathcal{S} \subset \mathcal{S}^\dagger \subset \mathcal{S}^*$.

Markov chains, stationary distributions and ergodicity

Let $X^{(0)}, X^{(1)}, \ldots$ denote a Markov chain with transition matrix P.

Let row vector $p^{(t)}$ denote the p.m.f. of $X^{(t)}$, so that

$$p^{(t)} = p^{(0)} P^t, \qquad t = 0, 1, \ldots,$$

Want to construct P s.t. $p^{(t)} \to \pi$ as $t \to \infty$.

$\Rightarrow$ P must satisfy $\pi P = \pi$.

Ex. Sudoku puzzles and a Metropolis algorithm

Task : Replace 0's such that each 3×3 subtable is an arrangement of the integers $1, 2, \ldots, 9$ and so are all the rows and columns of the 9×9 overall table.

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

Guardian #60 ("medium!")

Initial configuration

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

Eventual solution

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

Guardian #59 ("difficult")

Initial configuration

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

Eventual solution

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

Times #210 ("difficult")

Initial configuration

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

Eventual solution

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

Times #211 ("fiendish")

Initial configuration

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

Eventual solution

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

Times ("fiendish")

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

Sample space : S = set of all 9×9 tables x with feasible 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

Metropolis algorithm

Sample space : S = set of all 9×9 tables x with feasible subtables.

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

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

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

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

Algorithm :

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 $v(x) = 0$ is reached.

Metropolis algorithm

Let $\{\pi(x) : x \in S\}$ denote the target p.m.f.

Let x' denote current state and x'' the proposed next state.

Accept x'' as the next state with probability $\min\{1, \pi(x'')/\pi(x')\}$ else retain x'.

$\Rightarrow$ Corresponding Markov chain has limiting distribution $\{\pi(x)\}$.

Assumptions

1. Proposal matrix is symmetric; i.e. for each x', x'', the probability of proposing x' from x'' is the same as that of proposing x'' from x'.
2. Proposal matrix $\Rightarrow$ each $x'' \in S$ can eventually be reached from each $x' \in S$.

On the way . . .

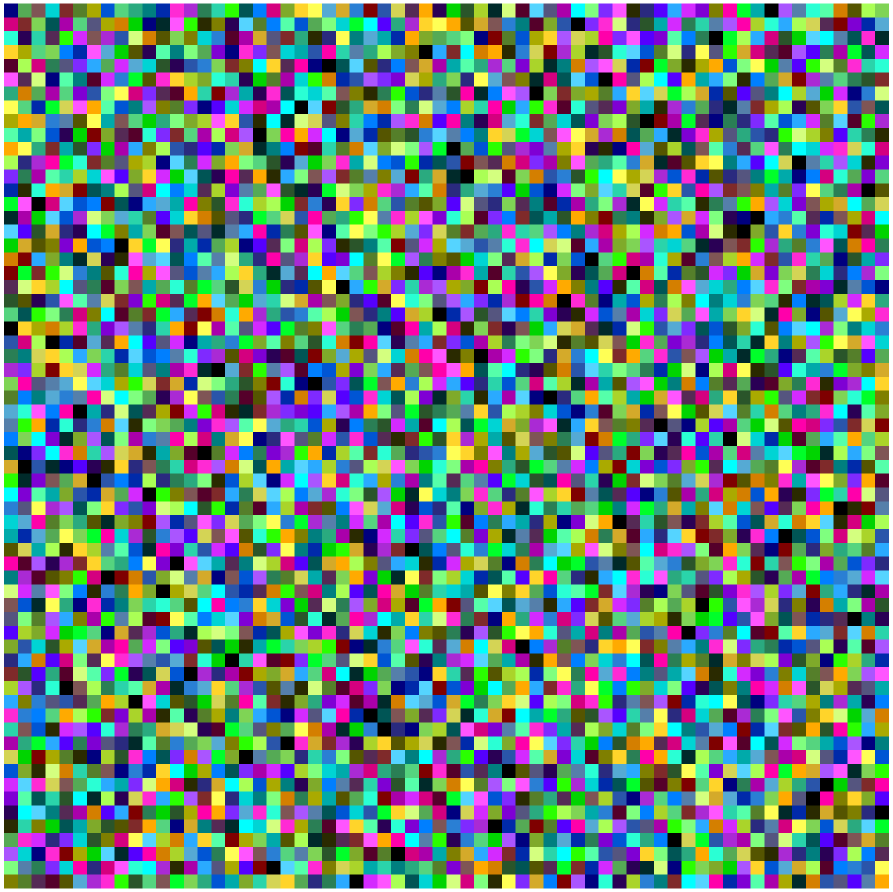
2	9	5	7	3	1	8	6	4
4	3	1	8	6	5	9	2	7
8	7	6	2	4	9	1	3	5
3	1	7	9	5	2	4	8	6
9	6	4	3	8	7	5	1	2
5	8	2	4	1	6	7	9	3
6	7	8	5	2	3	1	4	9
1	2	9	6	7	4	3	5	8
3	5	4	1	9	8	2	7	6

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

Constructing large Latin squares

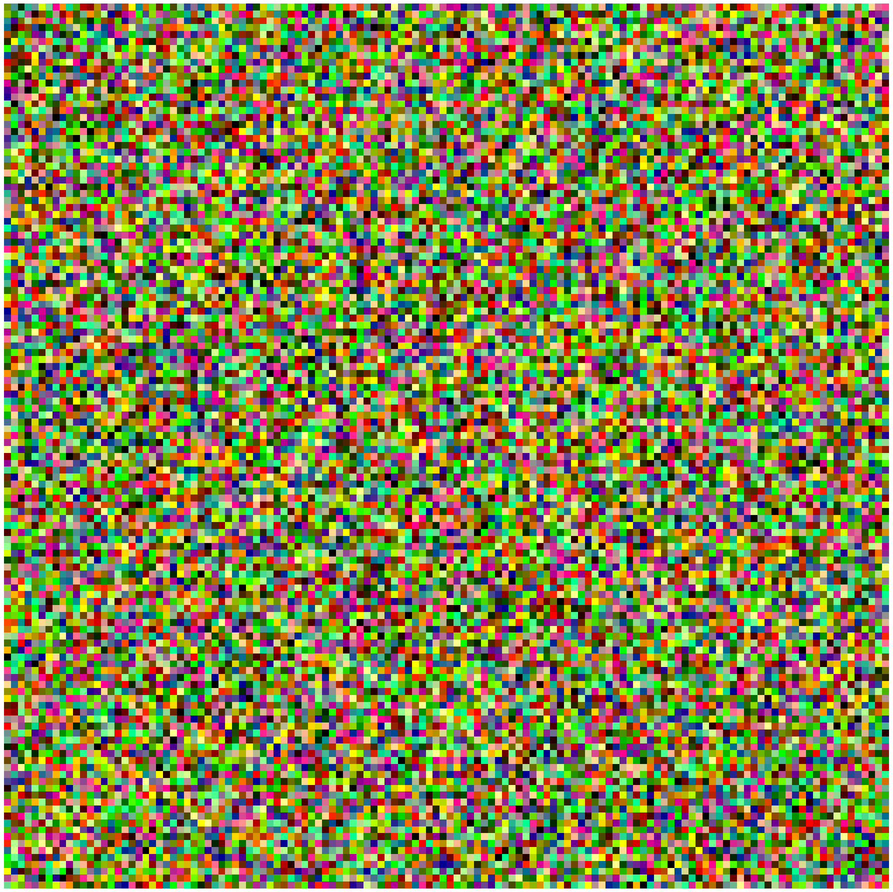
Can be done similarly . . .



64×64 **Latin square**

Constructing large Latin squares

Can be done similarly ...



128×128 **Latin square**