Random sampling of combinatorial structures via probabilistic divide-and-conquer

Stephen DeSalvo

University of California, Los Angeles

May 11, 2016

Overview

1. Combinatorial Structures and Sampling Methods
2. Integer Partitions
3. Contingency Tables
4. Future Work

But first!

Definition

A **random Bernoulli matrix** is an $n \times n$ matrix in which all entries are $+1$ or -1 with equal probability, independent of all other entries.

$$\begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$$

Question

What is the probability that a random Bernoulli matrix is singular?

Some history

Definition

Let S denote the event that a random Bernoulli matrix is singular.
Let $P_n := \mathbb{P}(S)$ denote the probability that a random Bernoulli matrix is singular.

Conjecture

$$P_n = 2^2 \binom{n}{2} \left(\frac{1}{2}\right)^n (1 + o(1)).$$

Some history

Definition

Let S denote the event that a random Bernoulli matrix is singular.
Let $P_n := \mathbb{P}(S)$ denote the probability that a random Bernoulli matrix is singular.

Conjecture

$$P_n = 2^2 \binom{n}{2} \left(\frac{1}{2}\right)^n (1 + o(1)).$$

Theorem (Komlós 1977)

$$P_n = O\left(\frac{1}{\sqrt{n}}\right).$$

Some history

Definition

Let S denote the event that a random Bernoulli matrix is singular.
Let $P_n := \mathbb{P}(S)$ denote the probability that a random Bernoulli matrix is singular.

Conjecture

$$P_n = 2^2 \binom{n}{2} \left(\frac{1}{2}\right)^n (1 + o(1)).$$

Theorem (Kahn, Komlós, Szemerédi 1995)

$$P_n = O\left(.999^n\right).$$

Some history

Definition

Let S denote the event that a random Bernoulli matrix is singular.
Let $P_n := \mathbb{P}(S)$ denote the probability that a random Bernoulli matrix is singular.

Conjecture

$$P_n = 2^2 \binom{n}{2} \left(\frac{1}{2}\right)^n (1 + o(1)).$$

Theorem (Tao, Vu 2007)

$$P_n \leq \left(\frac{3}{4} + o(1)\right)^n.$$

Some history

Definition

Let S denote the event that a random Bernoulli matrix is singular.
Let $P_n := \mathbb{P}(S)$ denote the probability that a random Bernoulli matrix is singular.

Conjecture

$$P_n = 2^2 \binom{n}{2} \left(\frac{1}{2}\right)^n (1 + o(1)).$$

Theorem (Bourgain, Vu, Wood 2012)

$$P_n \leq \left(\frac{1}{\sqrt{2}} + o(1)\right)^n.$$

Motivation

Quoting from Kahn, Komlós, Szemerédi:

Of course the guess ... may be sharpened, e.g., to

$$P_n - 2^2\binom{n}{2}\left(\frac{1}{2}\right)^n \sim 2^4\binom{n}{4}\left(\frac{3}{8}\right)^n,$$

the right-hand side being essentially the probability of having a minimal row or column dependency of length 4.

Templates

We think of null vectors of the form $v_i \pm v_j$ as coming from a *template* 11. Similarly, null vectors of the form $v_i \pm v_j \pm v_k \pm v_\ell$ we say correspond to template 1111. Null vectors of the form $\lambda_1 v_{i_1} \pm \cdots \pm \lambda_k v_{i_k}$ we say correspond to template $\lambda = (\lambda_1, \ldots, \lambda_k)$.

Conjectured list in decreasing order of r_λ

i	$\lambda(i)$	$len(\lambda)$	r_λ	$256 \cdot r_\lambda$
1	11	2	$1/2$	128.
2	1111	4	$3/8$	96.
3	111111	6	$5/16$	80.
4	11111111	8	$35/128$	70.
5	21111	5	$1/4$	64.
6	1111111111	10	$63/256$	63.
7	2111111	7	$15/64$	60.
8	111111111111	12	$231/1024$	57.75
9	221111	6	$7/32$	56.
10	211111111	9	$7/32$	56.
11	11111111111111	14	$429/2048$	53.625
12	21111111111	11	$105/512$	52.5
13	22111111	8	$13/64$	52.
14	1111111111111111	16	$6435/32768$	50.2734
15	2111111111111	13	$99/512$	49.5
16	2211111111	10	$49/256$	49.
17	2221111	7	$3/16$	48.
18	111111111111111111	18	$12155/65536$	47.4805
19	211111111111111	15	$3003/16384$	46.9219
20	221111111111	12	$93/512$	46.5
21	222111111	9	$23/128$	46.
22	11111111111111111111	20	$46189/262144$	45.1064
23	21111111111111111	17	$715/4096$	44.6875
24	22111111111111	14	$1419/8192$	44.3438
25	3211111	7	$11/64$	44.
26	22221111	8	$11/64$	44.
27	22211111111	11	$11/64$	44.
28	1111111111111111111111	22	$88179/524288$	43.0562
29	2111111111111111111	19	$21879/131072$	42.7324

Conjectured list in decreasing order of r_λ

i	$\lambda(i)$	$len(\lambda)$	r_λ	$256 \cdot r_\lambda$
30	2211111111111111	16	2717/16384	42.4531
31	2221111111111	13	675/4096	42.1875
32	31111111	8	21/128	42.
33	33111111	8	21/128	42.
34	321111111	9	21/128	42.
35	3111111111	10	21/128	42.
36	2222111111	10	21/128	42.
37	111111111111111111111111	24	676039/4194304	41.2621
38	211111111111111111111	21	20995/131072	41.0059
39	221111111111111111	18	10439/65536	40.7773
40	32111111111	11	81/512	40.5
41	222211111111	12	323/2048	40.375
42	31111111111111	14	1287/8192	40.2188
43	311111	6	5/32	40.
44	322111	6	5/32	40.
45	3222111	7	5/32	40.
46	32211111	8	5/32	40.
47	222221111	9	5/32	40.
48	11111111111111111111111111	26	1300075/8388608	39.6751
49	21111111111111111111111	23	323323/2097152	39.4681
50	22111111111111111111	20	20111/131072	39.2793
51	3211111111111	13	627/4096	39.1875
52	3111111111111111	16	5005/32768	39.1016
53	3221111111	10	39/256	39.
54	3311111111	10	39/256	39.
55	22222111111111	14	623/4096	38.9375
56	22222111111	11	155/1024	38.75
57	1111111111111111111111111111	28	5014575/33554432	38.2582
58	2111111111111111111111111	25	156009/1048576	38.0881
59	322211111	9	19/128	38.

Lower bound expansion

Theorem (Arratia, D 2013)

$$P_n \geq Q_1(n)\left(\frac{1}{2}\right)^n + Q_2(n)\left(\frac{3}{8}\right)^n + Q_3(n)\left(\frac{5}{16}\right)^n + Q_4(n)\left(\frac{35}{128}\right)^n$$
$$+ Q_5(n)\left(\frac{1}{4}\right)^n + Q_6(n)\left(\frac{63}{256}\right)^n + Q_7(n)\left(\frac{15}{64}\right)^n + Q_8(n)\left(\frac{231}{1024}\right)^n$$
$$+ o\left(\frac{7+\epsilon}{32}\right)^n,$$

where $Q_1(n) = 2^2\binom{n}{2}$, $Q_2(n) = 2^4\binom{n}{4}$ $Q_3(n) = 2^6\binom{n}{6}$, $Q_4(n) = 2^8\binom{n}{8}$

$$Q_5(n) = 2^5\binom{5}{1}\binom{n}{5} - 4\left(2\binom{n}{2}^2 + 8\binom{n}{4} + 5\binom{n}{3}\right),$$

$$Q_6(n) = 2^{10}\binom{n}{10}, \quad Q_7(n) = 2^7\binom{7}{1}\binom{n}{7}, \quad Q_8(n) = 2^{12}\binom{n}{12}.$$

Random combinatorial structures

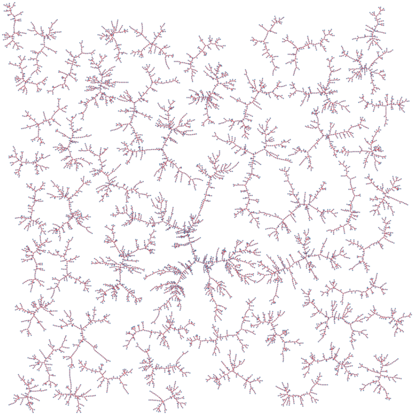
In this talk, we will consider sampling problems which can be described as

$$(X_1, X_2, \ldots, X_n) \qquad \text{independent random variables}$$

subject to a restriction which breaks the independence. We consider finite n, arbitrarily large, and the "combinatorial" comes from the fact that the entire collection of random variables will satisfy a restriction, represented by some measurable event E.

$$\mathcal{L}(\ (X_1, \ldots, X_n) \mid (X_1, \ldots, X_n) \in E\ ).$$

Tree



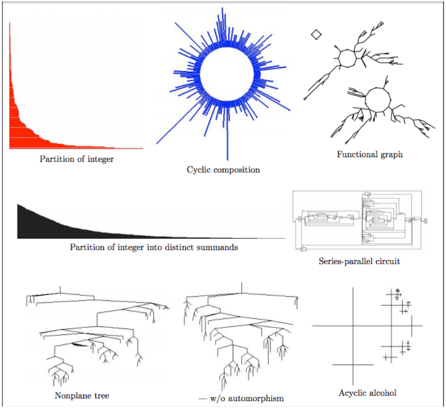
A tree is a connected graph with no cycles. A forest is a graph with no cycles.

$$\mathcal{L}(\ (X_{1,2}, X_{1,3}, \ldots, X_{n-1,n})\ |\ \text{no cycles}\ )$$

However, you wouldn't simply generate edges at random in a large graph and expect to generate a forest.

`http://ilan.schnell-web.net/prog/perfect-hash/algo.html`

Other structures



A Boltzmann model encodes recursive relationships between structures and provides a sampling algorithm, often in the form

$$\mathcal{L}\left((X_1, X_2, \ldots, X_n) \,\middle|\, \sum_{i=1}^{n} i\, X_i = n \right)$$

In many cases, the conditioning event is mild enough so that it can be effectively ignored without introducing an overwhelming bias.

Flajolet, Phillipe, Fusy, Eric, and Pivoteau, Carine. (2007) "Boltzmann sampling of unlabelled structures." Proceedings of the Meeting on Analytic Algorithmics and Combinatorics. Society for Industrial and Applied Mathematics.

Tables

19	1	6	5	0	6	37
34	13	7	12	12	5	83
7	4	0	1	7	7	26
26	2	5	16	0	7	56
86	20	18	34	19	25	

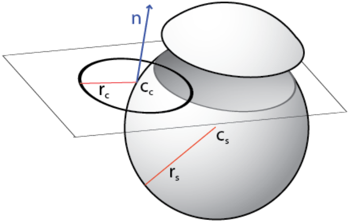
For contingency tables, the conditioning cannot be ignored.

Barvinok, Alexander. (2010) "What does a random contingency table look like?." Combinatorics, Probability and Computing 19.04 : 517-539.

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

Similarly with Latin squares and Sudoku matrices, the conditioning is an essential feature of the structure, and breaking it fundamentally changes various statistics like Shannon's entropy.

Newton, Paul K., DeSalvo, Stephen. (2009). "The Shannon entropy of Sudoku matrices". Proceedings of the Royal Society A. doi:10.1098/rspa.2009.0522.



A method should also be able to sample from reasonably tame sets as well, assuming we resolve any paradoxes.

$$\mathcal{L}\left((X_1, \ldots, X_n) \,\middle|\, \sqrt{\sum_{i=1}^{n} X_i^2} = r, \ \sum_{i=1}^{n} a_i X_i = k \right)$$

http://gamedev.stackexchange.com/questions/75756/sphere-sphere-intersection-and-circle-sphere-intersection

Some sampling methods

1. Recursive method of Nijenhuis and Wilf and other table methods.
2. Boltzmann sampling.
3. Run a forward Markov chain with stationary distribution which is uniform over a set which includes the set of objects of interest.
4. Run a Markov chain, coupling from the past, with stationary distribution which is uniform over the set of objects.
5. Importance sampling.
6. Ad hoc approaches.

Integer partitions: a model structure

Arratia, R. and DeSalvo, S. (2016) "Probabilistic divide-and-conquer: a new exact simulation method, with integer partitions as an example." Combinatorics, Probability, and Computing.

Integer Partitions

Definition

An integer partition of size n is a multiset of positive integers $\lambda_1, \lambda_2 \ldots \lambda_\ell$ such that $\sum_{i \geq 1} \lambda_i = n$. Alternatively, it is the set of all n-tuples $(c_1, \ldots, c_n)$ of nonnegative integers such that $\sum_{i=1}^{n} i\, c_i = n$.

Example (All integer partitions of 1, 2, 3, 4, 5)

λ	(c_1)
1	(1)

λ	(c_1, c_2)
2	(0,1)
11	(2,0)

λ	(c_1, c_2, c_3)
3	(0,0,1)
21	(1,1,0)
111	(3,0,0)

λ	(c_1, c_2, c_3, c_4)
4	(0,0,0,1)
31	(1,0,1,0)
22	(0,2,0,0)
211	(2,1,0,0)
1111	(4,0,0,0)

Fristedt's approach

Fristedt's conditioning device (see also Vershik, Kerov)

- Let $Z_1, Z_2, \ldots$ denote *independent* geometric random variables.
- Z_i has parameter $1 - x^i$ for *any* $0 < x < 1$, $i = 1, 2, \ldots$.
- Let c_i denote the number of parts of size i in an integer partition of size n (recall all partitions of n satisfy $\sum_{i=1}^n i\, c_i = n$).

$$\mathbb{P}(Z_1 = c_1, Z_2 = c_2 \ldots) = \prod_{i \geq 1} x^{i\, c_i}(1 - x^i) = x^n \prod_{i \geq 1}(1 - x^i).$$

Theorem (Fristedt 1993)

Let $C_i(n)$ denote the number of parts of size i in a uniformly random integer partition of n, $i = 1, 2, \ldots$. For any x between 0 and 1,

$$\left((Z_1(x), \ldots, Z_n(x)) \,\middle|\, \sum_{i=1}^n i\, Z_i(x) = n \right) =_D (C_1(n), \ldots, C_n(n)).$$

A sampling algorithm

Algorithm 1 Rejection Sampling of Integer Partitions

0. Let $x = e^{-\pi/\sqrt{6n}}$
1. Generate sample from $\mathcal{L}(Z_1, Z_2, \ldots, Z_n)$, call it $(z_1, \ldots, z_n)$.
2. If $\sum_{i=1}^{n} i\, z_i = n$, return $(z_1, \ldots, z_n)$; else restart.

Fristedt 1993

Let $x = e^{-\pi/\sqrt{6n}}$

$$\mathbb{P}\left(\sum_{i=1}^{n} i\, Z_i = n\right) \sim \frac{1}{\sqrt[4]{96 n^3}}.$$

The rejection sampling algorithm involves $O(n^{3/4})$ expected number of rejections before completing. This value of x is asymptotically optimal.

Probabilistic divide-and-conquer

Algorithm 2 Probabilistic Divide-and-Conquer Deterministic Second Half for Integer Partitions (Arratia, D. 2016)

0. Let $x = e^{-\pi/\sqrt{6n}}$.
1. Generate sample from $\mathcal{L}(Z_2, \ldots, Z_n)$, call it $(z_2, \ldots, z_n)$. Set $k := n - \sum_{i=2}^{n} i\, z_i$.
2. If $k \geq 0$ and $U < e^{-k\,\pi/\sqrt{6n}}$, return $(k, z_2, \ldots, z_n)$; else restart.

Theorem (Arratia, D. 2016)

Algorithm 2 samples integer partitions of size n uniformly at random, and the expected number of rejections before completing is $O(n^{1/4})$.

Probabilistic divide-and-conquer

Setup

We assume that the target distribution we wish to sample from can be expressed in the form $\mathcal{L}\big((A, B) \,\big|\, (A, B) \in E\big)$ for some $E \subset \mathcal{A} \times \mathcal{B}$, where

$$A \in \mathcal{A}, \;\; B \in \mathcal{B} \;\; \text{have given distributions,} \tag{1}$$

$$A, B \;\; \text{are independent,} \tag{2}$$

and that E is a measurable event of positive probability.

PDC Lemmas

Lemma ((PDC Lemma 1) Arratia, D. 2016)

Suppose E is a measurable event of positive probability. Suppose X is a random element of $\mathcal{A}$ with distribution

$$\mathcal{L}(X) = \mathcal{L}(\,A\,|\,(A,B) \in E\,), \tag{3}$$

and Y is a random element of $\mathcal{B}$ with conditional distribution

$$\mathcal{L}(Y\,|X = x) = \mathcal{L}(\,B\,|\,(x,B) \in E\,). \tag{4}$$

Then $\mathcal{L}(X,Y) = \mathcal{L}((A,B)|(A,B) \in E)$.

Comparison with rejection sampling

Algorithm 2 Rejection Sampling (von Neumann 1951)

1. Generate sample from $\mathcal{L}(A)$, call it a.
2. Generate sample from $\mathcal{L}(B)$, call it b.
3. Check if $(a, b) \in E$; if so, return (a, b), otherwise restart.

I.e., generate $(A_1, B_1), (A_2, B_2), \ldots$ until the first i such that $(A_i, B_i) \in E$.

Algorithm 3 Probabilistic Divide-and-Conquer (Arratia, D. 2016)

1. Generate sample from $\mathcal{L}(A \,|\, (A, B) \in E)$, call it x.
2. Generate sample from $\mathcal{L}(B \,|\, (x, B) \in E)$, call it y.
3. Return (x, y).

Von Neumann, J. (1951) *13. Various Techniques Used in Connection With Random Digits.* National Bureau of Standards, Applied Math. Series 12, pp. 36-38.

PDC sampling

Algorithm 4 Probabilistic Divide-and-Conquer via Rejection Sampling (Arratia, D. 2016)

1a. Generate sample from $\mathcal{L}(A)$, call it a.

1b. Accept a with probability $t(a)$, where $t(a)$ is a function of $\mathcal{L}(B)$ and E; otherwise, restart.

2. Generate sample from $\mathcal{L}(B|(a, B) \in E)$, call it y.
3. Return (a, y).

In many cases of interest,

1. $\mathcal{L}(B|(a, B) \in E)$ is deterministic; or
2. $\mathcal{L}(B|(a, B) \in E)$ has the same form as $\mathcal{L}(A|(A, B) \in E)$ with smaller inputs, i.e., recursive.
3. $\mathcal{L}(B)$ can be used instead! (Requires added assumptions)

Probabilistic divide-and-conquer

Algorithm 4 Self–Similar Probabilistic Divide-and-Conquer for Integer Partitions (Arratia, D. 2016)

procedure SS_PDC_IP(n)
0. If $n = 1$, return 1; otherwise,
1. Generate sample from $\mathcal{L}(Z_1, Z_3, Z_5, \ldots)$, call it $(z_1, z_3, z_5, \ldots)$.
2. Set $k := n - \sum_{i\ odd} i\, z_i$.
3. If $k < 0$ or k is odd, restart.
4. If $U < f_n(k/2)/\max_\ell f_n(\ell)$,
   let $(z_2, z_4, \ldots) = \text{SS\_PDC\_IP}(k/2)$;
   return $(z_1, z_2, \ldots, z_n, 0, 0, \ldots)$.
   Else restart.

end procedure

$$f_n(j) = \mathbb{P}\left(\sum_{i \le n/2} (2i) Z_{2i} = 2j\right) = p(j)\, x^{2j} \prod_{i \ge 1} 1 - x^{2i},$$

where $p(j)$ is the number of integer partitions of size j.

Probabilistic divide-and-conquer

Algorithm 5 Self–Similar Probabilistic Divide-and-Conquer for Integer Partitions (Arratia, D. 2016)

procedure SS_PDC_IP(n)
0. If $n = 1$, return 1; otherwise,
1. Generate sample from $\mathcal{L}(Z_1, Z_3, Z_5, \ldots)$, call it $(z_1, z_3, z_5, \ldots)$.
2. Set $k := n - \sum_{i\ odd} i\, z_i$.
3. If $k < 0$ or k is odd, restart.
4. If $U < f_n(k/2)/\max_\ell f_n(\ell)$,
   let $(z_2, z_4, \ldots) = \text{SS\_PDC\_IP}(k/2)$;
   return $(z_1, z_2, \ldots, z_n, 0, 0, \ldots)$.
   Else restart.

end procedure

Theorem (Arratia, D. 2016)

Algorithm 5 samples integer partitions of size n uniformly at random, with an expected overall rejection rate of at most $2\sqrt{2}$.

Punchline for integer partitions

$$(Z_1, Z_2, \ldots, Z_n), \qquad E = \left\{ \sum_{i=1}^{n} i\, Z_i = n. \right\}$$

Division	(A, B)	Expected # rejections
Rejection Sampling	$A = (Z_1, Z_2, \ldots, Z_n)$, $B = \emptyset$	$O(n^{3/4})$
PDC Deterministic second half	$A = (Z_2, Z_3, \ldots, Z_n)$, $B = (Z_1)$	$O(n^{1/4})$
PDC Self-similar 1	$A = (Z_1, Z_3, Z_5, \ldots)$, $B = (Z_2, Z_4, Z_6, \ldots)$	$O(1)$

Feel the Bern

Let $G(q), G'(q)$ denote independent geometric random variables with parameter $1-q$. Let $\mathrm{Bern}(p)$ denote a Bernoulli random variable with parameter p, independent of the other random variables. We have

$$G(q) =_D \mathrm{Bern}\left(\frac{q}{1+q}\right) + 2G'(q^2).$$

For integer partitions, we may consider the following PDC division,

$$A = \left(\mathrm{Bern}_1\left(\frac{x}{1+x}\right), \mathrm{Bern}_2\left(\frac{x^2}{1+x^2}\right), \ldots, \mathrm{Bern}_n\left(\frac{x^n}{1+x^n}\right)\right)$$

$$B = (2Z_1\left(x^2\right), 2Z_2\left(x^4\right), \ldots, 2Z_n\left(x^{2n}\right)).$$

The expected number of rejections is $O(1)$.

Contingency tables

(r, c)-Contingency Tables

DeSalvo, S. and Zhao, J. Y., *Random sampling of contingency tables via probabilistic divide-and-conquer*, `http://arxiv.org/abs/1507.00070`.

Contingency tables

Definition

Let $r = (r_1, \ldots, r_m)$ and $c = (c_1, \ldots, c_n)$ denote vectors of nonnegative integers. An $m \times n$ *nonnegative integer-valued* (r, c)*-contingency table* is an $m \times n$ table of nonnegative integers $X_{i,j}$, $1 \le i \le m$, $1 \le j \le n$, such that

$$\sum_{i=1}^{m} X_{i,j} = c_j, \qquad \forall j = 1, 2, \ldots, n.$$

$$\sum_{j=1}^{n} X_{i,j} = r_i, \qquad \forall i = 1, 2, \ldots, m.$$

Definition

The restriction $X_{i,j} \in \{0, 1\}$ is known as *binary* (r, c)-contingency table. We can also let $X_{i,j} \in \mathbb{R}_{\ge 0}$, and call them nonnegative real-valued (r, c)-contingency tables.

Uniform over nonnegative integer-valued tables

Uniform distribution

Let $\mathbb{X} = \{X_{i,j}\}_{i,j}$ denote a collection of independent geometric random variables with parameters $p_{i,j}$, $1 \leq i \leq m$, $1 \leq j \leq n$. Then for a particular (r, c)-table ξ we have

$$\mathbb{P}(\mathbb{X} = \xi) = \prod_{i,j} \mathbb{P}(X_{i,j} = \xi_{i,j}) = \prod_{i,j} p_{i,j}(1 - p_{i,j})^{\xi_{i,j}}.$$

If we also have $p_{i,j} = 1 - \alpha_i \beta_j$, i.e., $p_{i,j}$ factors, then we also have

$$\mathbb{P}(\mathbb{X} = \xi) = \prod_i \alpha_i^{r_i} \prod_j \beta_j^{c_j} \prod_{i,j} (1 - \alpha_i \beta_j).$$

Random sampling of contingency tables

Algorithm 6 Rejection sampling of nonnegative integer-valued (r, c)-contingency tables

0. Let $p_{i,j} = \frac{m}{m+c_j}$ for all $1 \le i \le m, 1 \le j \le n$.
1. Generate sample from $\mathcal{L}(X_{1,1}, \ldots, X_{m,n})$, call it x.
2. If x satisfies all row and column conditions, return x; else restart.

Algorithm 7 PDC random sampling of nonnegative integer-valued (r, c)-contingency tables

0. Let $p_{i,j} = \frac{m}{m+c_j}$ for all $1 \le i \le m, 1 \le j \le n$.

1a. Generate sample from $\mathcal{L}(\epsilon_{1,1})$, apply rejection $f(1,1)$.

1b. Generate sample from $\mathcal{L}(\epsilon_{2,1})$, apply rejection $f(2,1)$.

etc.

1z. Generate sample from $\mathcal{L}(\epsilon_{m-1,n-1})$, apply rejection $f(m-1, n-1)$.

2. Fill in deterministic values, divide residual line sums by 2.
3. Repeat.

Enumerative rejection function

Let $\mathcal{O}$ denote an $m \times n$ matrix with entries in $\{0,1\}$. For any set of row sums and column sums (r,c), let $\Sigma(r,c,\mathcal{O})$ denote the number of (r,c)-contingency tables with entry (i,j) forced to be even if the (i,j)th entry of $\mathcal{O}$ is 1, and no restriction otherwise. Let $\mathcal{O}_{i,j}$ denote the matrix which has entries with value 1 in the first $j-1$ columns, and entries with value 1 in the first i rows of column j, and entries with value 0 otherwise.

Define for $1 \leq i \leq m-2, 1 \leq j \leq n-2,\ k \in \{0,1\}$,

$$f(i,j,k) := \frac{\Sigma\begin{pmatrix} (\ldots, r_i - k, \ldots), \\ (\ldots, c_j - k, \ldots), \\ \mathcal{O}_{i,j} \end{pmatrix}}{\Sigma\begin{pmatrix} (\ldots, r_i - 1, \ldots), \\ (\ldots, c_j - 1, \ldots), \\ \mathcal{O}_{i,j} \end{pmatrix} + \Sigma\begin{pmatrix} (\ldots, r_i, \ldots), \\ (\ldots, c_j, \ldots), \\ \mathcal{O}_{i,j} \end{pmatrix}}.$$

Probabilistic rejection function

$f(i,j,k) \propto$

$$\mathbb{P}\left(\begin{array}{llll}
\sum_{\ell=1}^{j-1} 2\xi''_{1,\ell}(q_\ell^2, c_\ell) & +\eta'_{1,j,i}(q_j, c_j) & +\sum_{\ell=j+1}^{n} \xi'_{1,j}(q_\ell, c_\ell) & = r_1 \\
\sum_{\ell=1}^{j-1} 2\xi''_{2,\ell}(q_\ell^2, c_\ell) & +\eta'_{2,j,i}(q_j, c_j) & +\sum_{\ell=j+1}^{n} \xi'_{2,j}(q_\ell, c_\ell) & = r_2 \\
\vdots & & & \\
\sum_{\ell=1}^{j-1} 2\xi''_{i-1,\ell}(q_\ell^2, c_\ell) & +\eta'_{i-1,j,i}(q_j, c_j) & +\sum_{\ell=j+1}^{n} \xi'_{i-1,j}(q_\ell, c_\ell) & = r_{i-1} \\
\sum_{\ell=1}^{j-1} 2\xi''_{i,\ell}(q_\ell^2, c_\ell) & +\eta''_{i,j,i}(q_j, c_j) & +\sum_{\ell=j+1}^{n} \xi'_{i,j}(q_\ell, c_\ell) & = r_i - k \\
\sum_{\ell=1}^{j-1} 2\xi''_{i+1,\ell}(q_\ell^2, c_\ell) & +\eta''_{i+1,j,i}(q_j, c_j) & +\sum_{\ell=j+1}^{n} \xi'_{i+1,j}(q_\ell, c_\ell) & = r_{i+1} \\
\vdots & & & \\
\sum_{\ell=1}^{j-1} 2\xi''_{m,\ell}(q_\ell^2, c_\ell) & +\eta''_{m,j,i}(q_j, c_j) & +\sum_{\ell=j+1}^{n} \xi'_{m,j}(q_\ell, c_\ell) & = r_m
\end{array}\right)$$

$$\times \quad \mathbb{P}\left(\sum_{\ell=1}^{i} 2\,\xi_{\ell,j}(q_j^2) + \sum_{\ell=i+1}^{m} \xi_{\ell,j}(q_j) = c_j - k\right).$$

Probabilistic rejection function

$$f(i,j,k) \propto$$

$$\mathbb{P}\left(\begin{array}{llll}
\sum_{\ell=1}^{j-1} 2\xi''_{1,\ell}(q_\ell^2, c_\ell) & +\eta'_{1,j,i}(q_j, c_j) & +\sum_{\ell=j+1}^{n} \xi'_{1,j}(q_\ell, c_\ell) & = r_1 \\
\sum_{\ell=1}^{j-1} 2\xi''_{2,\ell}(q_\ell^2, c_\ell) & +\eta'_{2,j,i}(q_j, c_j) & +\sum_{\ell=j+1}^{n} \xi'_{2,j}(q_\ell, c_\ell) & = r_2 \\
\vdots & & & \\
\sum_{\ell=1}^{j-1} 2\xi''_{i-1,\ell}(q_\ell^2, c_\ell) & +\eta'_{i-1,j,i}(q_j, c_j) & +\sum_{\ell=j+1}^{n} \xi'_{i-1,j}(q_\ell, c_\ell) & = r_{i-1} \\
\boldsymbol{\sum_{\ell=1}^{j-1} 2\xi''_{i,\ell}(q_\ell^2, c_\ell)} & \boldsymbol{+\eta''_{i,j,i}(q_j, c_j)} & \boldsymbol{+\sum_{\ell=j+1}^{n} \xi'_{i,j}(q_\ell, c_\ell)} & \boldsymbol{= r_i - k} \\
\sum_{\ell=1}^{j-1} 2\xi''_{i+1,\ell}(q_\ell^2, c_\ell) & +\eta''_{i+1,j,i}(q_j, c_j) & +\sum_{\ell=j+1}^{n} \xi'_{i+1,j}(q_\ell, c_\ell) & = r_{i+1} \\
\vdots & & & \\
\sum_{\ell=1}^{j-1} 2\xi''_{m,\ell}(q_\ell^2, c_\ell) & +\eta''_{m,j,i}(q_j, c_j) & +\sum_{\ell=j+1}^{n} \xi'_{m,j}(q_\ell, c_\ell) & = r_m
\end{array}\right)$$

$$\times \quad \mathbb{P}\left(\boldsymbol{\sum_{\ell=1}^{i} 2\, \xi_{\ell,j}(q_j^2) + \sum_{\ell=i+1}^{m} \xi_{\ell,j}(q_j) = c_j - k}\right).$$

Probabilistic rejection function

Approximate sampling rejection function

$$f(i,j,k) \propto$$

$$\mathbb{P}\left(\sum_{\ell=1}^{\mathbf{j}-1} 2\xi''_{\mathbf{i},\ell}(\mathbf{q}_\ell^2, \mathbf{c}_\ell) + \eta''_{\mathbf{i},\mathbf{j},\mathbf{i}}(\mathbf{q}_\mathbf{j}, \mathbf{c}_\mathbf{j}) + \sum_{\ell=\mathbf{j}+1}^{\mathbf{n}} \xi'_{\mathbf{i},\mathbf{j}}(\mathbf{q}_\ell, \mathbf{c}_\ell) = \mathbf{r}_\mathbf{i} - \mathbf{k}\right)$$

$$\times \quad \mathbb{P}\left(\sum_{\ell=1}^{\mathbf{i}} 2\,\xi_{\ell,\mathbf{j}}(\mathbf{q}_\mathbf{j}^2) + \sum_{\ell=\mathbf{i}+1}^{\mathbf{m}} \xi_{\ell,\mathbf{j}}(\mathbf{q}_\mathbf{j}) = \mathbf{c}_\mathbf{j} - \mathbf{k}\right).$$

These functions are sums of i.i.d. random variables with explicitly known distributions, and hence can be computed using convolutions.

Current Developments

1. Random sampling of numerical tables, e.g., Latin squares.

   DeSalvo, S. *Exact, uniform sampling of Latin squares and Sudoku matrices*, `http://arxiv.org/abs/1502.00235`

   DeSalvo, S. *Random sampling of numerical tables via probabilistic divide-and-conquer, in preparation*.

2. DeSalvo, S. *Improvements to exact Boltzmann samplers using probabilistic divide-and-conquer and the recursive method*. GASCOM 2016

3. What are the most general conditions on continuous–random variables which allow more general PDC algorithms?

   DeSalvo, S. *Probabilistic divide-and-conquer: deterministic second half*, `http://arxiv.org/abs/1411.6698`

Thanks!

Questions?