Importance Sampling for Counting Sudoku's

Ad Ridder

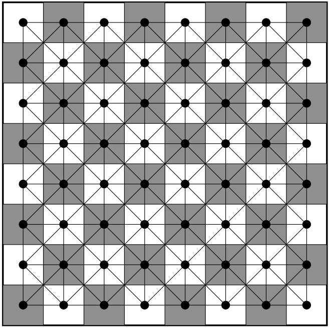
Department EOR
Vrije Universiteit Amsterdam
Homepage: `http://personal.vu.nl/a.a.n.ridder/`

RMC, 6 April 2013

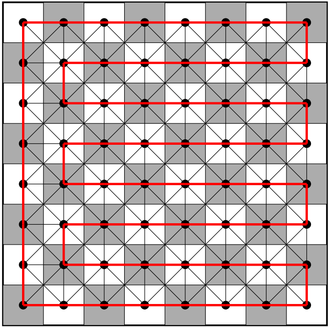
Counting Combinatorial Problems

- Number of Hamilton cycles in a graph;
- Number of perfect matchings in balanced bipartite graph;
  - Equivalently; the permanent of its associated biadjacency matrix;
- Number of binary contingency tables with given row and column totals;
- Number of feasible assignments to a 3-SAT formula;
- Number of random graphs with prescribed degrees;
- Etc.

Illustration Hamilton Cycles

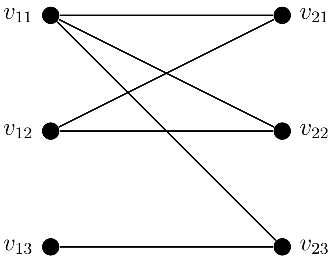


A graph

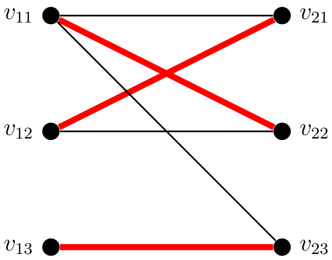


With a Hamilton cycle.

Illustration Matching



A bipartite graph.



With a perfect matching.

Illustration Contingency Table

							4
							2
							3
							5
							3
3	4	2	1	2	2	3	

A (empty) table with required row and column totals

1	1			1		1	4
	1	1					2
1	1				1		3
	1	1		1	1	1	5
1			1			1	3
3	4	2	1	2	2	3	

A solution.

Recreational Counting Problem: The Sudoku

Sudoku *puzzles* are published in newspapers, internet, etc. For instance

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

The Sudoku

Solving the puzzle you get the *Sudoku*.

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

The Sudoku Counting Problem

vrije Universiteit amsterdam

Question

How large is the number of distinct Sudoku's?

- Basically, this will answer whether there are enough puzzles before repeating.
- You find Sudoku's on Internet, in newspapers, magazines, special puzzle books,etc.
- Daily.
- We like to know whether these are all different.
- Not taking into account reduction by symmetries such as rotations, swapping, etc.

Formulation of a Counting Problem

- Given is a *population* Ω of elements.
- Interest is in some *subset* $A \subset \Omega$ of elements defined by some *property*:

$$A = \{x \in \Omega : \text{property}(x)\}.$$

- The *objective* is: compute the size of A; i.e. $|A|$ or $\#A$.

A Probabilistic Formulation

- Consider the probability space $(\Omega, \mathbb{P})$;
- Where $\mathbb{P} = \mathbb{P}_u$ is the probability measure on Ω induced by the *uniform* probability density function (pdf) u:

$$\mathbb{P}(B) = \mathbb{P}_u(B) = u(B) = \frac{|B|}{|\Omega|}, \quad \text{for all } B \subset \Omega;$$

- Specifically for the singleton subsets: $u(x) = 1/|\Omega|$;
- A random element of Ω is represented by the identity map $X : \Omega \to \Omega$;
- Thus, we get the natural interpretation of picking *at random* elements of Ω:

$$\mathbb{P}(X \in B) = u(B) = \frac{|B|}{|\Omega|}, \quad \text{for all } B \subset \Omega;$$

- The counting problem is reformulated by the computation problem

$$|A| = \mathbb{P}(X \in A)\,|\Omega| = u(A)\,|\Omega|.$$

Assumptions

1. The size $|\Omega|$ is known, but too large to check each element for its property;
2. The size $|A|$ is unknown;
3. It is easy to generate samples $x \in \Omega$ from the uniform distribution;
4. It is easy to test whether a sample $x \in \Omega$ lies in the target set A or not.

Easy means that the associated algorithm runs in polynomial time.

Counting by Simulation

- Generate iid copies $X_1, \ldots, X_n$ of X using the uniform pdf;
- Implement an *unbiased* estimator of $\mathbb{P}(X \in A)$ by

$$Y = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}\{X_k \in A\}$$

- Return unbiased estimator of $|A|$ by

$$\widehat{|A|} = Y\,|\Omega|$$

- Also called the *crude Monte Carlo* (CMC) estimator.

Next

- Apply to Sudoku counting problem;
- Analyse performance of the counting estimator.

Definition

A *9-permutation lattice* is a 9×9-lattice with each row a 9-permutation.

Example

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

The population Ω

- Population Ω is the set of all 9-permutation lattices;
- Clearly, its size is

$$|\Omega| = (9 \times 8 \times 7 \times \cdots \times 1)^9 = (9!)^9 = 1.091 \times 10^{50}$$

- Generating a random 9-permutation lattice is trivial since it is just a matter of placing 9 random 9-permutations under each other;
- The set A of all Sudoku's is a subset of Ω.

CMC Fails

- The set A of Suduko's is a *rare event* in the probability space $(\Omega, \mathbb{P})$;
- Running the simulation algorithm for generating 100M 9-permutation lattices did not result in any Sudoku;
- This took about 5:45 minutes on my MacBook Pro 2.4 GHz;
- In fact, we will see later that on average only 1 in 1.6×10^{28} 9-permutation lattices is a Sudoku;
- It would take 1.8×10^{15} years on average to 'catch' a Sudoku.

Change of Measure

- Idea: simulate elements $x \in \Omega$ using a pdf $q(x)$;
- Such that Sudoku's have a higher chance of being generated; i.e.

$$\mathbb{P}_q(X \in A) \gg \mathbb{P}_u(x \in A)$$

- Apply

$$\mathbb{P}_u(X \in A) = \mathbb{E}_u[\mathbb{1}\{X \in A\}] = \mathbb{E}_q\Big[\underbrace{\frac{d\mathbb{P}_u}{d\mathbb{P}_q}(X)}_{\substack{\text{likelihood} \\ \text{ratio } L(X)}}\ \mathbb{1}\{X \in A\}\Big]$$

- This is called a *change of measure*.

Importance Sampling

- From the change of measure we get *importance sampling simulation*;

Algorithm

Repeat n times (independently)

1. *Simulate* a 9-permutation lattice X using new pdf q;
2. *Compute* the associated likelihood ration $L(X) = u(X)/q(X)$;
3. *Determine* whether sample $X \in A$ (is a sudoku).

- Return IS estimator of the probability $Y_{\mathrm{IS}} = \frac{1}{n} \sum_{k=1}^{n} L(X_k) \mathbb{1}\{X_k \in A\}$;
- Return IS estimator of the number $\widehat{|A|}_{\mathrm{IS}} = Y_{\mathrm{IS}} \, |\Omega|$.

Caveat

- Importance sampling estimator is *unbiased*;
- Variance reduction (compared to CMC) can be huge;
- However, a careless change of measure could result in variance increase!
- The importance sampling issue is: find a good change of measure that is implementable;
- Actually, goal is to deliver an optimal importance sampling estimator, given an optimality criteria; or within a class of estimators.

A Change of Measure for the Sudoku

- A random 9-permutation lattice under the $\mathbb{P}_q$ probability:
  - *(a).* Simulate 9-permutations row by row starting at the top;
  - *(b).* Simulate a row from left to right;
  - *(c).* Suppose box (i,j) (row i, column j) has to get a digit:
    - In CMC randomly one of the $9-j+1$ row-feasible digits would be drawn;
    - In IS, determine the number R_{ij} of row+column+block feasible digits, and choose one at random;
    - If $R_{ij}=0$, apply the CMC rule.
- Note that after the first $R_{ij}=0$, generating the current lattice can be stopped with output zero.

Example

vrije Universiteit amsterdam

- Suppose that we have reached $i = 3, j = 5$; and that we have so far

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

- In CMC, choose one of the feasible *1,2,5,8,9* at random;
- In IS; just *2,8* are feasible.

Importance Sampling Results

- It takes just about 0.77 sec. to generate 1M lattices in the IS simulations;
- There were more than 3000 Sudoku's among them;
- Repeating this 100 times gave an estimate of 6.6622e+21 for the number of distinct Sudoku's;
- With 95% confidence interval

$$6.2633\text{e+}21 < \widehat{|A|} < 7.0611\text{e+}21$$

- The number of Sudoku puzzles is of the same order;
- An enormous number; surely you will never encounter the same puzzle twice;
- NB: for the classic 9×9 Sudoku the exact number has been established to be 6.671e+21 (6670903752021072936960); the estimate has a relative error of 2%.

Analysis of IS Estimator

vrije Universiteit amsterdam

- First some implementation issues:
  - Because the original $\mathbb{P}$ is the uniform: $u(X) = 1/|\Omega|$:

$$\widehat{|A|}_{\text{IS}} = \frac{1}{n}\sum_{k=1}^{n} L(X_k)\mathbb{1}\{X_k \in A\}\,|\Omega| = \frac{1}{n}\sum_{k=1}^{n}\frac{u(X_k)}{q(X_k)}\mathbb{1}\{X_k \in A\}\,|\Omega|$$

$$= \frac{1}{n}\sum_{k=1}^{n}\frac{1}{q(X_k)}\mathbb{1}\{X_k \in A\}$$

  - Where $q(X) = \prod_{ij} q_{ij}(X) = \prod_{ij} 1/R_{ij}(X)$

- Consider the rare event probability $\mathbb{P}_u(X \in A)$;
  - A probability estimator Y is *efficient* if

$$\text{RAT}(Y) = \frac{|\log \mathbb{E}[Y^2]|}{|\log(\mathbb{E}[Y])^2|} \approx 1$$

  - CMC is not efficient: $\text{RAT}(Y_{\text{CMC}}) = 0.5$;
  - The IS estimator has $\text{RAT}(Y_{\text{IS}}) \approx 0.95$ (empirically from the 100 experiments).

Discussion

- This importance sampling algorithm is based on a *one-step-look-ahead* (OSLA) idea;
- OSLA has been applied also for counting
  - Hamilton cycles;
  - Perfect matchings;
  - Self-avoiding walks;
  - Etc.
- Still, it can be stucked quite often;
- See the Sudoku counting: 'only' 3000 out of one million trials turned out to be successful;

Conclusion and Outlook

- A randomized algorithm for counting problems;
- Specifically, counting Sudoku's;
- Importance sampling based on one-step-look-ahead;
- Easy to use for other counting problems; other versions of the Sudoku; larger dimensional Sudoku's ($16 \times 16, 25 \times 25, \ldots$);
- Ongoing research for more advanced importance sampling algorithms (stochastic enumeration); and other randomized counting algorithms (splitting; MCMC).

Some References

1. Blanchet, J., and Rudoy, D. 2009. Rare event simulation and counting problems. Chapter 8 in *Rare event simulation using Monte Carlo methods*, eds. Rubino G. and Tuffin, B., Wiley, pp. 171-192.
2. L'Ecuyer, P., Blanchet, J.H., Tuffin, B., and Glynn, P.W. 2010. Asymptotic robustness of estimators in rare-event simulation, *ACM Transactions on Modeling and Computer Simulation*, 20(1), article 6.
3. Felgenhauer, B., and Javis, F. 2005. Enumerating possible Sudoku grids. Available at `http://www.afjarvis.staff.shef.ac.uk/sudoku/sudoku.pdf`
4. Jerrum, M. 2001. *Counting, sampling and integrating: algorithms and complexity*. Birkhäuser.
5. Mitzenmacher, M., Upfal, E. 2005. *Probability and computing: randomized algorithms and probabilistic analysis.* Cambridge University Press.
6. Rubinstein, R., Ridder, A., and Vaisman, S. 2014. *Fast sequential Monte Carlo methods for counting and optimization.* Wiley (forthcoming).