Four lectures on importance sampling

Lecture 4: Sequential importance sampling in large finite state spaces

Henrik Hult

Department of Mathematics
KTH Stockholm

KTH, June 2010

Outline

Number of non-self intersecting paths

- Consider the integer lattice $L_n = \{(i,j) : 0 \leq i,j \leq n\}$. Compute the number of non-self intersecting paths from $(0,0)$ to (n,n).
- For $n = 10$ the answer is 1.5687×10^{24}.

Number of non-self intersecting paths

- Consider the integer lattice $L_n = \{(i,j) : 0 \leq i,j \leq n\}$. Compute the number of non-self intersecting paths from $(0,0)$ to (n,n).
- For $n = 10$ the answer is 1.5687×10^{24}.

Number of non-self intersecting paths

A sequential importance sampling algorithm for this problem

- Generate N paths denoted γ_j according to the uniform distribution on open nodes (show example).
- Let $p(\gamma)$ be the probability of a path γ and Γ_n be the set of non-self intersecting paths that reach (n, n).
- Estimate the number of paths by

$$|\hat{\Gamma}_n| = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{p(\gamma_j)} I\{\gamma_j \in \Gamma_n\}.$$

Number of non-self intersecting paths

Picture

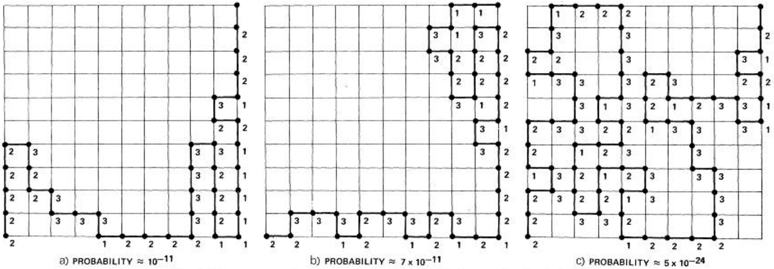

Fig. 4. Three more randomly generated paths, with their associated probabilities.

From D.E. Knuth, Mathematics and computer science: coping with finiteness. *Science,* 194(4271), 1235–1242, 1976.

Number of non-self intersecting paths

Bias and variance

- The estimate is unbiased

$$E\left[\frac{1}{N}\sum_{j=1}^{N}\frac{1}{p(\gamma_j)}I\{\gamma_j \in \Gamma_n\}\right] = \sum_{\gamma\in\Gamma_n}\frac{1}{p(\gamma)}p(\gamma) = |\Gamma_n|.$$

- The variance is

$$\text{Var}(|\hat{\Gamma}_n|) = \frac{1}{N}\Big(\sum_{\gamma\in\Gamma_n}\frac{1}{p(\gamma)^2}p(\gamma) - |\Gamma_n|^2\Big) = \frac{1}{N}\Big(\sum_{\gamma\in\Gamma_n}\frac{1}{p(\gamma)} - |\Gamma_n|^2\Big).$$

- Would like to find $p(\gamma)$ such that $\text{Var}(|\hat{\Gamma}_n|)$ to grow roughly like $|\Gamma_n|^2$.

Outline

1. Number of non-self intersecting paths
   - The number of binary tables
2. A simpler problem – monotone paths

Some related problems

Binary tables

- Consider a binary $0/1$ contingency table of size $m \times n$.
- Given a set of row sums $r = (r_1, \ldots, r_m)$ and column sums $c = (c_1, \ldots, c_n)$ let $\Gamma_{r,c}$ be the set of tables with row sums r and column sums c.
- **How many such tables are there?**

Darwin's Finch data

Picture

Table 1. Occurrence Matrix for Darwin's Finch Data

	Island																
Finch	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*
Large ground finch	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Medium ground finch	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0
Small ground finch	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0
Sharp-beaked ground finch	0	0	1	1	1	0	0	1	0	1	0	1	1	0	1	1	1
Cactus ground finch	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1	0	0
Large cactus ground finch	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
Large tree finch	0	0	1	1	1	1	1	1	1	0	0	1	0	1	1	0	0
Medium tree finch	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Small tree finch	0	0	1	1	1	1	1	1	1	1	0	1	0	0	1	0	0
Vegetarian finch	0	0	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0
Woodpecker finch	0	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0
Mangrove finch	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Warbler finch	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

NOTE: Island name code: A = Seymour, B = Baltra, C = Isabella, D = Fernandina, E = Santiago, F = Rábida, G = Pinzón, H = Santa Cruz, I = Santa Fe, J = San Cristóbal, K = Española, L = Floreana, M = Genovesa, N = Marchena, O = Pinta, P = Darwin, Q = Wolf.

From Chen et al., Sequential Monte Carlo for statistical analysis of tables. *JASA,* 100(469), 109–120, 2005.

A related problem

Binary tables

- Strategy: sample a table sequentially according to a probability $p(T_i)$.
- Estimate $|\Gamma_{r,c}|$ by

$$|\hat{\Gamma}_{r,c}| = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{p(T_j)} I\{T_j \in \Gamma_n\}.$$

- How to sample the tables?

A related problem

Binary tables

- The first column contains c_1 number of 1's.
- Sample the position of the 1's $i_1, \ldots, i_{c_1}$ from a conditional Poisson $\mathrm{CP}(p_1, \ldots, p_n)$, $p_i = r_i/n$.
- If $(Z_1, \ldots, Z_m)$ is independent $\mathrm{Ber}(p_1, \ldots, p_m)$ and $S = Z_1 + \ldots Z_m$, then the distribution of $(Z_1, \ldots, Z_m) \mid S$ is $\mathrm{CP}(p_1, \ldots, p_m)$.
- Update the table by computing remaining row sums.
- A modification of the algorithm avoids getting stuck.

Number of monotone paths[1]

- Return to the non-self intersecting paths. This time take a much simpler problem with a similar spirit.
- Consider the integer lattice $L_n = \{(i, j) : 0 \leq i, j \leq n\}$. Let Γ_n the number of monotone paths from $(0, 0)$ to (n, n). That is, paths that can only move up or to the right.
- Pretend you cannot compute the number of such paths. How would you simulate?

[1] This material comes partly from Bassetti and Diaconis (2005)

Number of monotone paths[1]

- Return to the non-self intersecting paths. This time take a much simpler problem with a similar spirit.
- Consider the integer lattice $L_n = \{(i, j) : 0 \leq i, j \leq n\}$. Let Γ_n the number of monotone paths from $(0, 0)$ to (n, n). That is, paths that can only move up or to the right.
- Pretend you cannot compute the number of such paths. How would you simulate?

[1] This material comes partly from Bassetti and Diaconis (2005)

Number of monotone paths

Naive generation of paths

A first sequential importance sampling algorithm for this problem

- Generate each path by taking $p^{\uparrow}(i,j) = 1 - p^{\rightarrow}(i,j) = 1/2$ for $i,j \neq n$ and 1 on the boundary.
- Let $p(\gamma)$ be the probability of a path γ.
- Estimate

$$|\hat{\Gamma}_n| = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{p(\gamma_j)}.$$

- Will this be efficient?

Problem session

Monotone paths and no-down moves

- **Problem 1a** Let $n = 5$. How many monotone paths are there from $(0, 0)$ to $(5, 5)$?
- **Problem 1b** If you were to simulate to find the answer, how would you simulate 'optimally' from position $(1, 3)$ to compute the number of paths? Find $p^{\uparrow}(1, 3)$.
- **Problem 2a** Suppose you can move up, left, or right (but not down) without intersecting your path. How many such paths are there from $(0, 0)$ to $(5, 5)$?
- **Problem 2b** If you were to simulate to find the answer, how would you simulate 'optimally' from position $(1, 3)$ if you reached $(1, 3)$ from below? Find $p^{\leftarrow}(1, 3), p^{\uparrow}(1, 3), p^{\rightarrow}(1, 3)$.

Number of monotone paths

Naive generation of paths – analysis

How efficient is the naive generation of paths with probability $1/2$ to go up?

- The number of monotone paths is

$$|\Gamma_n| = \binom{2n}{n} \approx \frac{4^n}{\sqrt{n}}, \text{ and } \binom{2n}{n}^2 \approx \frac{16^n}{n},$$

- Letting j be the number of steps until the path hits the boundary, the second moment of the estimator is

$$\sum_{\gamma} \frac{1}{p(\gamma)} = 2 \sum_{j=n}^{2n-1} 2^j \binom{j-1}{n-1} \sim 2^{2n} \binom{2n-2}{n-1} \approx \frac{16^n}{\sqrt{n}}$$

- So second moment divided by squared number of paths grows like $\sqrt{n}$.

Summary

I tried to provide an introduction to importance sampling and recent developments with a bias towards my own interests/work.

- Efficient importance sampling can be constructed by approximating the zero-variance change of measure.
- Large deviations/asymptotic analysis can be used to come up with the design of efficient algorithms.
- A stochastic control formulation of sequential importance sampling problems provides a basis for the efficiency analysis.

Summary

I tried to provide an introduction to importance sampling and recent developments with a bias towards my own interests/work.

- Efficient importance sampling can be constructed by approximating the zero-variance change of measure.
- Large deviations/asymptotic analysis can be used to come up with the design of efficient algorithms.
- A stochastic control formulation of sequential importance sampling problems provides a basis for the efficiency analysis.

Summary

I tried to provide an introduction to importance sampling and recent developments with a bias towards my own interests/work.

- Efficient importance sampling can be constructed by approximating the zero-variance change of measure.
- Large deviations/asymptotic analysis can be used to come up with the design of efficient algorithms.
- A stochastic control formulation of sequential importance sampling problems provides a basis for the efficiency analysis.

Summary

I tried to provide an introduction to importance sampling and recent developments with a bias towards my own interests/work.

- Efficient importance sampling can be constructed by approximating the zero-variance change of measure.
- Large deviations/asymptotic analysis can be used to come up with the design of efficient algorithms.
- A stochastic control formulation of sequential importance sampling problems provides a basis for the efficiency analysis.

References I

Bassetti, F. and Diaconis, P.
Examples comparing Importance Sampling and the Metropolis algorithm.
Illinois J. Math. 50:67–91, 2006.

Blanchet, J. and Glynn, P.
Efficient rare-event simulation for the maximum of heavy-tailed random walks.
Ann. Appl. Prob., 18:1351–1378, 2008.

Blanchet, J. and Liu, J.C.
State-dependent Importance Sampling for Regularly Varying Random Walks.
Adv. in Appl. Probab. 40: 1104-1128, 2008.

Blanchet, J., Glynn, P., and Liu, J.
Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue.
Queueing Systems 57:99-113, 2007.

Bucklew, J.A.
Introduction to rare event simulation.
Springer, 2004.

Chen Y., Diaconis, P., Holmes, S.P., and Liu, J.S.
Sequential Monte Carlo for statistical analysis of tables.
JASA, 100(469):109–120, 2005.

References II

Dupuis, P., Leder, K., and Wang, H.
Importance sampling for sums of random variables with regularly varying tails.
ACM Trans. Model. Comput. Simul. 17(3) Article 14, July 2007.

Dembo, A. and Zeitouni, O.
Large deviations techniques and applications.
Springer, 1998.

Dupuis, P. and Wang, H.
Subsolutions of an Isaacs equation and efficient schemes of importance sampling.
Math. OR. 32:723-757, 2007.

Dupuis, P. and Wang, H.
Dynamic importance sampling for uniformly recurrent Markov chains.
Ann. Appl. Probab. 15:1ÃÂÂ38, 2005.

Dupuis, P. and Wang, H.
Importance Sampling, Large Deviations, and Differential Games.
Stoch. and Stoch. Rep. 76:481-508, 2004

Hult, H. and Svensson, J.
On importance sampling with mixtures for random walks with heavy tails.
Preprint, 2009.
http://arxiv.org/PS_cache/arxiv/pdf/0909/0909.3333v1.pdf

References III

Hult, H. and Svensson, J.
Efficient calculation of risk measures by importance sampling - the heavy tailed case.
Preprint, 2009.
http://arxiv.org/PS_cache/arxiv/pdf/0909/0909.3335v1.pdf

Knuth, D.E.
Mathematics and computer science: coping with finiteness.
Science, 194(4271):1235-1242, 1976.

van der Vaart, A. and Wellner, J.
Weak convergence and empirical processes: with applications to statistics.
Springer, 1996.