Rare-event Simulation and Counting

Jose Blanchet

Columbia IEOR Department

RESIM

Agenda

1. General Overview: Approximate Counting and Rare-event Simulation

Agenda

1. General Overview: Approximate Counting and Rare-event Simulation
2. Bipartite Graphs: Motivation, Approximations and Algorithm Design

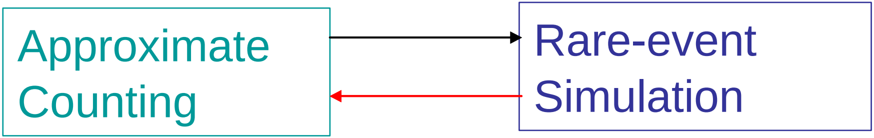
Agenda

1. General Overview: Approximate Counting and Rare-event Simulation
2. Bipartite Graphs: Motivation, Approximations and Algorithm Design
3. Bipartite Graphs: Complexity Analysis and Exact Sampling

Agenda

1. General Overview: Approximate Counting and Rare-event Simulation
2. Bipartite Graphs: Motivation, Approximations and Algorithm Design
3. Bipartite Graphs: Complexity Analysis and Exact Sampling
4. Conclusions

Approximate Counting and Rare-event Simulation



Goal: Discuss examples of properties / algorithms / insights that rare-event simulation techniques can provide in the context of approximate counting.

- **Applications:** Inference problems in graphical models: Psychology, Biology, etc. (Chen et al (2005), Blitzstein and Diaconis (2006)...)

General Overview

- **Applications:** Inference problems in graphical models: Psychology, Biology, etc. (Chen et al (2005), Blitzstein and Diaconis (2006)...)
- **Theory:** *Polynomial-time* randomized algorithms for approximate counting and / or sampling —> motivated by complexity theory (Valiant, Vazirini, Jerrum, Sinclair...)

General Overview

- **Applications:** Inference problems in graphical models: Psychology, Biology, etc. (Chen et al (2005), Blitzstein and Diaconis (2006)...)
- **Theory:** *Polynomial-time* randomized algorithms for approximate counting and / or sampling —> motivated by complexity theory (Valiant, Vazirini, Jerrum, Sinclair...)
- **Rare-event simulation:** Applications —> *efficient* algorithms whose complexity is tested in a meaningful (from an applied standpoint) asymptotic regime...

General Overview

- **Applications:** Inference problems in graphical models: Psychology, Biology, etc. (Chen et al (2005), Blitzstein and Diaconis (2006)...)
- **Theory:** *Polynomial-time* randomized algorithms for approximate counting and / or sampling —> motivated by complexity theory (Valiant, Vazirini, Jerrum, Sinclair...)
- **Rare-event simulation:** Applications —> *efficient* algorithms whose complexity is tested in a meaningful (from an applied standpoint) asymptotic regime...
- Differences / similarities?

Efficient Rare-event Simulation

- Want to compute $P(A_d) \searrow 0$ as $d \nearrow \infty$ (counting problems will be posed in this form)
- Assume that $EZ(d) = P(A_d)$
- **Definition: (Strong Efficiency)**

$$\sup_{d>1} \frac{Var(Z(d))}{P(A_d)^2} < \infty$$

- **Number of replications for relative precision is** $O(\varepsilon^{-2}\delta^{-1})$ **independent of** d...

$$P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\frac{Z_j(d)}{P(A_d)} - 1\right| \geq \varepsilon\right) \leq \frac{Var(Z(d))}{\varepsilon^2 n P(A_d)^2} \leq \delta$$

- Always keep in mind COST PER REPLICATION!

Importance Sampling for Counting

- Importance sampling algorithms: Chen, Diaconis, Holmes and Liu ‘05, Rubinstein ‘06, Botev and Kroese ’08
- Counting the number of simple undirected graphs with a given degree: Diaconis and Blitzstein ‘06,
- Rigorous complexity analysis of IS in counting settings: Bayati, Kim and Saberi ’07, Bezakova, Sinclair and Vigoda ’07, B. ’07.

Agenda

1. General Overview: Approximate Counting and Rare-event Simulation

Agenda

1. General Overview: Approximate Counting and Rare-event Simulation
2. **Bipartite Graphs: Motivation, Approximations and Algorithm Design**

Agenda

1. General Overview: Approximate Counting and Rare-event Simulation
2. **Bipartite Graphs: Motivation, Approximations and Algorithm Design**
3. Bipartite Graphs: Complexity Analysis and Exact Sampling

Agenda

1. General Overview: Approximate Counting and Rare-event Simulation
2. **Bipartite Graphs: Motivation, Approximations and Algorithm Design**
3. Bipartite Graphs: Complexity Analysis and Exact Sampling
4. Conclusions

Applications to IS to Counting Problems

- Bipartite graphs = binary tables with fixed magins...
- How many binary tables are there with given row and column sums?
- Motivation from Chen, Diaconis, Holmes and Liu (2005)

	Island																	
Finch	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	
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	1	0	1	0	1	1	0	0	**11**
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**
	3	**3**	**10**	**9**	**9**	**7**	**8**	**9**	**7**	**9**	**2**	**9**	**3**	**6**	**8**	**2**	**2**	**106**

Estimation Problems

- For inference applications really want to estimate quantiles and percentiles of a given statistic, say $f(\cdot)$, of the table given the margins

Estimation Problems

- For inference applications really want to estimate quantiles and percentiles of a given statistic, say $f(\cdot)$, of the table given the margins
- More precisely,

$$
\begin{aligned}
& P\left(f\left(Table\right) > z_q \middle| \, Table \text{ uniform given margins}\right) \\
= & \frac{P\left(f\left(Table\right) > z_q, \; Table \text{ uniform given margins}\right)}{P\left(Table \text{ uniform given margins}\right)}.
\end{aligned}
$$

MCMC Algorithms

- MCMC based algorithms (using the connection between approximate uniform sampling and approximate counting Jerrum, Valiant, Vazirini (1986))
- References: Sinclair (1993), Jerrum (2003) ...
- Best complexity known for **general sequences** (d = *total # ones*)

$$O^*(d^{11}\varepsilon^{-2}\delta^{-1})$$

due to Bezakova, Bhatnagar and Vigoda (2006)

IS for Counting Binary Contingency Tables

$$\begin{aligned} & P\,(\text{Table at random has required constrains}) \\ = \;& (\#\text{ Tables with constrains}) \;/\; (\#\text{ All tables}) \end{aligned}$$

- How to bias the distribution in the right way?
- How to analyze the complexity of such algorithm?

- Let $\mathbf{c}_k = (c_{k+1}, ..., c_n)$ $(n = \#$ *of columns*$)$, $\mathbf{r}_0 = (r_1, ..., r_m)$ $(m = \#$ *of rows*$)$
$$u_0(\mathbf{r}_0, \mathbf{c}_0) = P(\mathbf{r}_n = 0 | \mathbf{r}_0),$$
where $\mathbf{r}_{k+1} = \mathbf{r}_k - X_{k+1}$.
- $X_k \in \{0,1\}^m$ $(m = \#$ *of rows*$)$, *the components of* X_k are uniformly distributed over
$$\{x_1, ..., x_m : \sum_{i=1}^{m} x_i = c_k,\ x_i\text{'s} \in \{0,1\}\ \}.$$
- We have
$$\#\ of\ \ tables = u_0(\mathbf{r}_0, \mathbf{c}_0) \binom{m}{c_1} ... \binom{m}{c_n}$$

Applying Importance Sampling

- Note that

$$u_k(\mathbf{r}_k, \mathbf{c}_k) = Eu_{k+1}(\mathbf{r}_k - X_{k+1}, \mathbf{c}_{k+1}),$$

$$u_n(\mathbf{r}_n, *) = 1(\mathbf{r}_n = 0).$$

Applying Importance Sampling

- Note that
$$u_k(\mathbf{r}_k, \mathbf{c}_k) = Eu_{k+1}(\mathbf{r}_k - X_{k+1}, \mathbf{c}_{k+1}),$$
$u_n(\mathbf{r}_n, *) = 1(\mathbf{r}_n = 0)$.
- If we could use as importance sampler
$$q_k^*(\mathbf{r}_k, \mathbf{r}_{k+1}) = \frac{p_k(\mathbf{r}_k, \mathbf{r}_{k+1})\, u_{k+1}(\mathbf{r}_{k+1}, \mathbf{c}_{k+1})}{u_k(\mathbf{r}_k, \mathbf{r}_{k+1})}$$

Applying Importance Sampling

- Note that

$$u_k(\mathbf{r}_k, \mathbf{c}_k) = Eu_{k+1}(\mathbf{r}_k - X_{k+1}, \mathbf{c}_{k+1}),$$

$$u_n(\mathbf{r}_n, *) = 1(\mathbf{r}_n = 0).$$

- If we could use as importance sampler

$$q_k^*(\mathbf{r}_k, \mathbf{r}_{k+1}) = \frac{p_k(\mathbf{r}_k, \mathbf{r}_{k+1})\, u_{k+1}(\mathbf{r}_{k+1}, \mathbf{c}_{k+1})}{u_k(\mathbf{r}_k, \mathbf{r}_{k+1})}$$

- *The estimator would have* **zero variance**

$$Z = \frac{u_0(\mathbf{r}_0, \mathbf{c}_0)}{u_1(\mathbf{r}_1, \mathbf{c}_1)} \cdot \frac{u_1(\mathbf{r}_1, \mathbf{c}_1)}{u_2(\mathbf{r}_2, \mathbf{c}_2)} \cdots \frac{u_{n-1}(\mathbf{r}_1, \mathbf{c}_1)}{1}$$

Applying Importance Sampling

- Note that

$$u_k(\mathbf{r}_k, \mathbf{c}_k) = Eu_{k+1}(\mathbf{r}_k - X_{k+1}, \mathbf{c}_{k+1}),$$

$u_n(\mathbf{r}_n, *) = 1(\mathbf{r}_n = 0)$.

- If we could use as importance sampler

$$q_k^*(\mathbf{r}_k, \mathbf{r}_{k+1}) = \frac{p_k(\mathbf{r}_k, \mathbf{r}_{k+1})\, u_{k+1}(\mathbf{r}_{k+1}, \mathbf{c}_{k+1})}{u_k(\mathbf{r}_k, \mathbf{r}_{k+1})}$$

- *The estimator would have* **zero variance**

$$Z = \frac{u_0(\mathbf{r}_0, \mathbf{c}_0)}{u_1(\mathbf{r}_1, \mathbf{c}_1)} \cdot \frac{u_1(\mathbf{r}_1, \mathbf{c}_1)}{u_2(\mathbf{r}_2, \mathbf{c}_2)} \ldots \frac{u_{n-1}(\mathbf{r}_1, \mathbf{c}_1)}{1}$$

- Obviously $q^*(\cdot)$ is not a feasible choice... therefore

Applying importance sampling

- Use a change-of-measure of the form

$$q_k(\mathbf{r}_k, \mathbf{r}_{k+1}) = \frac{p_k(\mathbf{r}_k, \mathbf{r}_{k+1})\, v_{k+1}(\mathbf{r}_{k+1}, \mathbf{c}_{k+1})}{w_k(\mathbf{r}_k, \mathbf{c}_k)}$$

Applying importance sampling

- Use a change-of-measure of the form

$$q_k(\mathbf{r}_k, \mathbf{r}_{k+1}) = \frac{p_k(\mathbf{r}_k, \mathbf{r}_{k+1})\, v_{k+1}(\mathbf{r}_{k+1}, \mathbf{c}_{k+1})}{w_k(\mathbf{r}_k, \mathbf{c}_k)}$$

- In our current context (random walk for binary tables)

$$p_k(\mathbf{r}_k, \mathbf{r}_{k+1}) = \frac{1}{\binom{m}{c_k}}$$

if $0 \leq x_i - y_i \leq 1$ and $\sum_{i=1}^{m} (\mathbf{r}_k(i) - \mathbf{r}_{k+1}(i)) = c_k$ and $p_k(\mathbf{r}_k, \mathbf{r}_{k+1}) = 0$ otherwise.

Applying importance sampling

- Use a change-of-measure of the form

$$q_k(\mathbf{r}_k, \mathbf{r}_{k+1}) = \frac{p_k(\mathbf{r}_k, \mathbf{r}_{k+1})\, v_{k+1}(\mathbf{r}_{k+1}, \mathbf{c}_{k+1})}{w_k(\mathbf{r}_k, \mathbf{c}_k)}$$

- In our current context (random walk for binary tables)

$$p_k(\mathbf{r}_k, \mathbf{r}_{k+1}) = \frac{1}{\binom{m}{c_k}}$$

  if $0 \leq x_i - y_i \leq 1$ and $\sum_{i=1}^{m} (\mathbf{r}_k(i) - \mathbf{r}_{k+1}(i)) = c_k$ and $p_k(\mathbf{r}_k, \mathbf{r}_{k+1}) = 0$ otherwise.

- How to select $v_{k+1}(\cdot)$?

Constructing Good State-dependent Importance Samplers

- Approximate $\#$ *tables* $(\mathbf{r}, \mathbf{c}) \sim M(\mathbf{r}, \mathbf{c})$ as $\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j = d \nearrow \infty$

Constructing Good State-dependent Importance Samplers

- Approximate $\#$ *tables* $(\mathbf{r},\mathbf{c}) \sim M(\mathbf{r},\mathbf{c})$ as $\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j = d \nearrow \infty$
- Have $\{\varrho_1, ..., \varrho_m\}$ CELLS of type R and $\{\sigma_1, ..., \sigma_n\}$ CELLS of type C.

Constructing Good State-dependent Importance Samplers

- Approximate $\#$ *tables* $(\mathbf{r},\mathbf{c}) \sim M(\mathbf{r},\mathbf{c})$ as $\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j = d \nearrow \infty$
- Have $\{\varrho_1, ..., \varrho_m\}$ CELLS of type R and $\{\sigma_1, ..., \sigma_n\}$ CELLS of type C.
- Cell ϱ_i has r_i TOKENS, cell σ_j has c_j TOKENS

Constructing Good State-dependent Importance Samplers

- Approximate $\#$ *tables* $(\mathbf{r}, \mathbf{c}) \sim M(\mathbf{r}, \mathbf{c})$ as $\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j = d \nearrow \infty$
- Have $\{\varrho_1, ..., \varrho_m\}$ CELLS of type R and $\{\sigma_1, ..., \sigma_n\}$ CELLS of type C.
- Cell ϱ_i has r_i TOKENS, cell σ_j has c_j TOKENS
- All TOKENS are LABELED

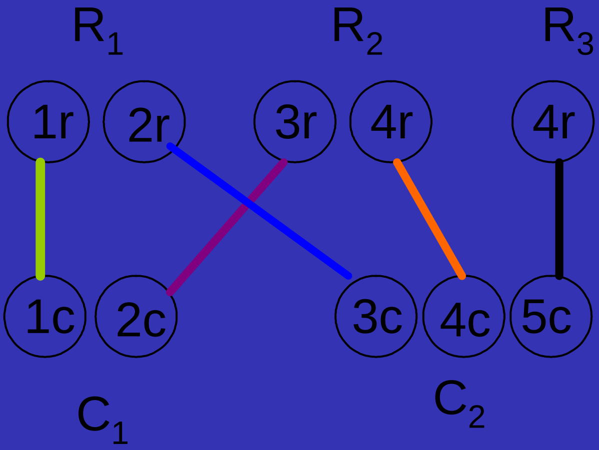
Constructing Good State-dependent Importance Samplers

- Approximate # *tables* $(\mathbf{r}, \mathbf{c}) \sim M(\mathbf{r}, \mathbf{c})$ as $\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j = d \nearrow \infty$
- Have $\{\varrho_1, ..., \varrho_m\}$ CELLS of type R and $\{\sigma_1, ..., \sigma_n\}$ CELLS of type C.
- Cell ϱ_i has r_i TOKENS, cell σ_j has c_j TOKENS
- All TOKENS are LABELED
- *Random pairing*: Pair each R-type tokens with each C-type tokens uniformly from the space of (unlabeled) pairings ($d!$ of them)

Constructing Good State-dependent Importance Samplers

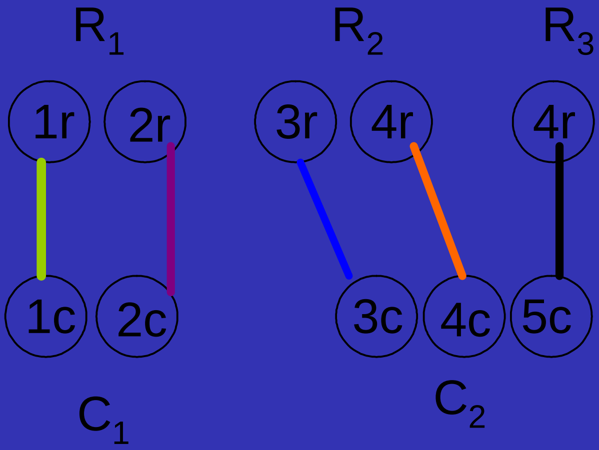
- Approximate # *tables* $(\mathbf{r}, \mathbf{c}) \sim M(\mathbf{r}, \mathbf{c})$ as $\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j = d \nearrow \infty$
- Have $\{\varrho_1, ..., \varrho_m\}$ CELLS of type R and $\{\sigma_1, ..., \sigma_n\}$ CELLS of type C.
- Cell ϱ_i has r_i TOKENS, cell σ_j has c_j TOKENS
- All TOKENS are LABELED
- *Random pairing*: Pair each R-type tokens with each C-type tokens uniformly from the space of (unlabeled) pairings ($d!$ of them)
- $N = \#$ of parallel pairs (pairs in the pairing that share the same cells).

Illustration of pairings with NO parallel pairs



A pairing with NO parallel pairs

Illustration of pairings with TWO parallel pairs



A pairing with 2 parallel pairs

Constructing Good State-dependent Importance Samplers

- Note that

$$\frac{\#\text{ of bipartite graphs}}{} = \frac{\#\text{ of pairings with no parallel pairs}}{c_1!...c_n!r_1!...r_m!}$$

Constructing Good State-dependent Importance Samplers

- Note that

$$= \frac{\frac{\#\text{ of bipartite graphs}}{\#\text{ of pairings with no parallel pairs}}}{c_1!...c_n!r_1!...r_m!}$$

- Approximate $N \approx Poisson(EN)$

$$\exp(-EN) \approx P(N=0)$$

Constructing Good State-dependent Importance Samplers

- Note that

$$\frac{\# \text{ of bipartite graphs}}{\# \text{ of pairings with no parallel pairs}} = \frac{}{c_1!...c_n!r_1!...r_m!}$$

- Approximate $N \approx Poisson(EN)$

$$\exp(-EN) \approx P(N = 0)$$

- Thus

$$\# \text{ of bipartite graphs} \approx \frac{d! \exp(-EN)}{c_1!...c_n!r_1!...r_m!}$$

Making Rigorous the Previous Approximation

- Using the inclusion-exclusion principle one obtains

Theorem (B. '07)

Suppose that $\sum c_j^2 = O(d) = \sum r_j^2$ *and* $\max c_j, r_i = o\left(d^{1/2}\right)$ *as* $d \nearrow \infty$. *Then,*

$$\#\text{ of bipartite graphs} \sim \frac{d! \exp(-EN)}{c_1!...c_n! r_1!...r_m!},$$

where $EN = 2\sum_{i,j} \binom{c_i}{2}\binom{r_j}{2} / (d(d-1))$.

Remark: Approx. shown by Mckay '84 under $\max c_j, r_i = o\left(d^{1/4}\right)$, see also Greenhill, McKay and Wang '06.

Back to Applying Importance Sampling

- AFTER sorting columns DECREASINGLY (we'll see why)

Back to Applying Importance Sampling

- AFTER sorting columns DECREASINGLY (we'll see why)
- Use a change-of-measure of the form

$$q_k\left(\mathbf{r}_k, \mathbf{r}_{k+1}\right) = \frac{p_k\left(\mathbf{r}_k, \mathbf{r}_{k+1}\right) v_{k+1}\left(\mathbf{r}_{k+1}, \mathbf{c}_{k+1}\right)}{w_k\left(\mathbf{r}_k, \mathbf{c}_k\right)}$$

Back to Applying Importance Sampling

- AFTER sorting columns DECREASINGLY (we'll see why)
- Use a change-of-measure of the form

$$q_k(\mathbf{r}_k, \mathbf{r}_{k+1}) = \frac{p_k(\mathbf{r}_k, \mathbf{r}_{k+1})\, v_{k+1}(\mathbf{r}_{k+1}, \mathbf{c}_{k+1})}{w_k(\mathbf{r}_k, \mathbf{c}_k)}$$

- Here $p_k(\mathbf{r}_k, \mathbf{r}_{k+1}) = \binom{m}{c_k}^{-1}$ & components of increment X_k are binary and add up to c_k.

Back to Applying Importance Sampling

- AFTER sorting columns DECREASINGLY (we'll see why)
- Use a change-of-measure of the form

$$q_k(\mathbf{r}_k, \mathbf{r}_{k+1}) = \frac{p_k(\mathbf{r}_k, \mathbf{r}_{k+1})\, v_{k+1}(\mathbf{r}_{k+1}, \mathbf{c}_{k+1})}{w_k(\mathbf{r}_k, \mathbf{c}_k)}$$

- Here $p_k(\mathbf{r}_k, \mathbf{r}_{k+1}) = \binom{m}{c_k}^{-1}$ & components of increment X_k are binary and add up to c_k.
- Previous theorem suggests put $d_k = \sum_{j=k}^{n} c_j$ AND $r_i' = \mathbf{r}_k(i)$

$$v_k(\mathbf{r}_k, \mathbf{c}_k) = \frac{d_k! \exp(-EN)}{c_k! \ldots c_n! r_1'! \ldots r_m'! \binom{m}{c_k} \ldots \binom{m}{c_n}}$$

The Form of the Increment Distribution

- Under the suggested change-of-measure turns out that X_k satisfies

$$
\begin{aligned}
& P^Q \left(X_k \left(1 \right) = x_1, ..., X_k \left(m \right) = x_m \right) \\
= \quad & P \left(Z_1 = x_1, ..., Z_m = x_m \middle| Z_1 + ... + Z_m = c_k \right)
\end{aligned}
$$

for Z_j's which are independent and Z_j is Bernoulli(p_j)

The Form of the Increment Distribution

- Under the suggested change-of-measure turns out that X_k satisfies

$$
\begin{aligned}
& P^Q\left(X_k(1)=x_1,\ldots,X_k(m)=x_m\right) \\
= \;& P\left(Z_1=x_1,\ldots,Z_m=x_m \,\middle|\, Z_1+\ldots+Z_m=c_k\right)
\end{aligned}
$$

for Z_j's which are independent and Z_j is Bernoulli(p_j)

- In turn

$$
\begin{aligned}
& P\left(Z_1=x_1,\ldots,Z_m=x_m \,\middle|\, Z_1+\ldots+Z_m=c_k\right) \\
\propto \;& \left(\frac{p_1}{1-p_1}\right)^{x_1}\times\ldots\times\left(\frac{p_m}{1-p_m}\right)^{x_m} I\left(x_1+\ldots+x_m=c_k\right)
\end{aligned}
$$

The Form of the Increment Distribution

- Under the suggested change-of-measure turns out that X_k satisfies

$$\begin{aligned} & P^Q\left(X_k(1)=x_1,\ldots,X_k(m)=x_m\right) \\ = \ & P\left(Z_1=x_1,\ldots,Z_m=x_m \mid Z_1+\ldots+Z_m=c_k\right) \end{aligned}$$

  for Z_j's which are independent and Z_j is Bernoulli(p_j)

- In turn

$$\begin{aligned} & P\left(Z_1=x_1,\ldots,Z_m=x_m \mid Z_1+\ldots+Z_m=c_k\right) \\ \propto \ & \left(\frac{p_1}{1-p_1}\right)^{x_1}\times\ldots\times\left(\frac{p_m}{1-p_m}\right)^{x_m} I\left(x_1+\ldots+x_m=c_k\right) \end{aligned}$$

- Our suggested change-of-measure induces

$$\frac{p_j}{1-p_j}\approx \mathbf{r}_k(j)+\frac{\theta_{k+1}}{d_{k+1}}\mathbf{r}_k(j)^2$$

  where θ_k depends on $c_k,\ldots,c_n$ and $\theta_k=O(1/d_k)$.

Agenda

1. General Overview: Approximate Counting and Rare-event Simulation

Agenda

1. General Overview: Approximate Counting and Rare-event Simulation
2. Bipartite Graphs: Motivation, Approximations and Algorithm Design

Agenda

1. General Overview: Approximate Counting and Rare-event Simulation
2. Bipartite Graphs: Motivation, Approximations and Algorithm Design
3. **Bipartite Graphs: Complexity Analysis and Exact Sampling**

Agenda

1. General Overview: Approximate Counting and Rare-event Simulation
2. Bipartite Graphs: Motivation, Approximations and Algorithm Design
3. **Bipartite Graphs: Complexity Analysis and Exact Sampling**
4. Conclusions

How Expensive is to Implement the Sampler?

1. Sampling from $Law(Z_1, ..., Z_m | Z_1 + ... + Z_m = c)$ takes $O(cm^2)$ operations (including the evaluation of normalizing constant, see Chen and Liu '97).

How Expensive is to Implement the Sampler?

1. Sampling from $Law(Z_1, ..., Z_m | Z_1 + ... + Z_m = c)$ takes $O(cm^2)$ operations (including the evaluation of normalizing constant, see Chen and Liu '97).
2. So, total complexity *per sample* is $O(dm^2)$ operations.

How to Estimate the Variance?

- Need to estimate

$$
\begin{aligned}
s_0(\mathbf{r}_0,\mathbf{c}_0) &= E^q_{\mathbf{r}_0,\mathbf{c}_0}\left(\prod_{k=0}^{n-1}\frac{w_k(\mathbf{r}_k,\mathbf{c}_k)^2}{v_{k+1}(\mathbf{r}_{k+1},\mathbf{c}_{k+1})^2} I(\mathbf{r}_n=0)\right) \\
&= v_0(\mathbf{r}_0,\mathbf{c}_0)^2\, E^Q_{\mathbf{r}_0,\mathbf{c}_0}\left(\prod_{k=0}^{n-1}\frac{w_k(\mathbf{r}_k,\mathbf{c}_k)^2}{v_k(\mathbf{r}_k,\mathbf{c}_k)^2}\frac{I(S_n=0)}{v_n(S_n,\mathbf{c}_n)^2}\right)
\end{aligned}
$$

How to Estimate the Variance?

- Need to estimate

$$s_0(\mathbf{r}_0,\mathbf{c}_0) = E^q_{\mathbf{r}_0,\mathbf{c}_0}\left(\prod_{k=0}^{n-1}\frac{w_k(\mathbf{r}_k,\mathbf{c}_k)^2}{v_{k+1}(\mathbf{r}_{k+1},\mathbf{c}_{k+1})^2}I(\mathbf{r}_n=0)\right)$$

$$= v_0(\mathbf{r}_0,\mathbf{c}_0)^2 E^Q_{\mathbf{r}_0,\mathbf{c}_0}\left(\prod_{k=0}^{n-1}\frac{w_k(\mathbf{r}_k,\mathbf{c}_k)^2}{v_k(\mathbf{r}_k,\mathbf{c}_k)^2}\frac{I(S_n=0)}{v_n(S_n,\mathbf{c}_n)^2}\right)$$

- **Lyapunov Bound (B & Glynn '07):** It suffices to find a positive Lyapunov function $h_k(\cdot)\geq 1$ such that

$$h_k(\mathbf{r}_k,\mathbf{c}_k)\geq\frac{w_k(\mathbf{r}_k,\mathbf{c}_k)^2}{v_k(\mathbf{r}_k,\mathbf{c}_k)^2}E^Q_{\mathbf{r}_k,k}(h_{k+1}(\mathbf{r}_k-X_{k+1},\mathbf{c}_{k+1})),$$

and as long as $\sum_{j=1}^m \mathbf{r}_k(j)\geq a$ (for a fixed) to obtain

$$s_0(\mathbf{r}_0,\mathbf{c}_0)\leq v_0(\mathbf{r}_0,\mathbf{c}_0)^2\cdot h_0(\mathbf{r}_0,\mathbf{c}_0)$$

Choosing the Lyapunov Function...

- Choosing the Lyapunov function requires creative thinking

Choosing the Lyapunov Function...

- Choosing the Lyapunov function requires creative thinking
- We always need a somewhat sharp estimate of the form

$$\frac{w_k(\mathbf{r}_k, \mathbf{c}_k)}{v_k(\mathbf{r}_k, \mathbf{c}_k)} = 1 + o(1)$$

as $d \nearrow \infty$. The more information in $o(1)$ the easier to get the Lyapunov bound and the better the performance of the algorithm...

For the Counting Problem...

- Turns out that (if $c_k = \max_{j \geq k} c_j$)

$$\frac{w_k(\mathbf{r}_k, \mathbf{c}_k)}{v_k(\mathbf{r}_k, \mathbf{c}_k)} \leq \exp\left(\lambda c_k^4 / d\right),$$

where λ is explicit and behaves like $\sum_{i=1}^{m} r_i^2 / d$.

For the Counting Problem...

- Turns out that (if $c_k = \max_{j \geq k} c_j$)

$$\frac{w_k(\mathbf{r}_k, \mathbf{c}_k)}{v_k(\mathbf{r}_k, \mathbf{c}_k)} \leq \exp\left(\lambda c_k^4 / d\right),$$

where λ is explicit and behaves like $\sum_{i=1}^m r_i^2 / d$.

- Consequently, one obtains

$$\begin{aligned} & \prod_{k=0}^{n-1} \frac{w_k(\mathbf{r}_k, \mathbf{c}_k)^2}{v_{k+1}(\mathbf{r}_{k+1}, \mathbf{c}_{k+1})^2} I(\mathbf{r}_n = 0) \\ \leq \quad & w_0(\mathbf{r}_0, \mathbf{s}_0)^2 \exp\left(\lambda \sum_{k=1}^n c_k^4 / d\right) \end{aligned}$$

Verifying Strong Efficiency

- **In particular Strong Efficiency (B '07):**

$$\frac{Var_q\left(\prod_{k=0}^{n-1}\frac{w_k(\mathbf{r}_k,\mathbf{c}_k)^2}{v(\mathbf{r}_{k+1},\mathbf{c}_{k+1})^2}I\left(\mathbf{r}_n=0\right)\right)}{u_0\left(\mathbf{r}_0,\mathbf{c}_0\right)^2}$$

$$\leq \exp\left(\lambda\sum_{k=1}^{n}c_k^4/d\right)\frac{v_0\left(\mathbf{r}_0,\mathbf{s}_0\right)^2}{u_0\left(\mathbf{r}_0,\mathbf{c}_0\right)^2}.$$

Verifying Strong Efficiency

- **In particular Strong Efficiency (B '07):**

$$\frac{Var_q\left(\prod_{k=0}^{n-1}\frac{w_k(\mathbf{r}_k,\mathbf{c}_k)^2}{v(\mathbf{r}_{k+1},\mathbf{c}_{k+1})^2}I\left(\mathbf{r}_n=0\right)\right)}{u_0\left(\mathbf{r}_0,\mathbf{c}_0\right)^2} \leq \exp\left(\lambda\sum_{k=1}^{n}c_k^4/d\right)\frac{v_0\left(\mathbf{r}_0,\mathbf{s}_0\right)^2}{u_0\left(\mathbf{r}_0,\mathbf{c}_0\right)^2}.$$

- Follows from counting estimates $\max\{c_k\}=o\left(d^{1/4}\right)$ as $d\nearrow\infty$ and if c_k's ORDERED decreasingly

$$\sum_{k=1}^{n}c_k^4=O\left(d\right).$$

Verifying Strong Efficiency

- **In particular Strong Efficiency (B '07):**

$$\frac{Var_q\left(\prod_{k=0}^{n-1}\frac{w_k(\mathbf{r}_k,\mathbf{c}_k)^2}{v(\mathbf{r}_{k+1},\mathbf{c}_{k+1})^2}I\left(\mathbf{r}_n=0\right)\right)}{u_0\left(\mathbf{r}_0,\mathbf{c}_0\right)^2}$$
$$\leq \exp\left(\lambda\sum_{k=1}^{n}c_k^4/d\right)\frac{v_0\left(\mathbf{r}_0,\mathbf{s}_0\right)^2}{u_0\left(\mathbf{r}_0,\mathbf{c}_0\right)^2}.$$

- Follows from counting estimates $\max\{c_k\}=o\left(d^{1/4}\right)$ as $d\nearrow\infty$ and if c_k's ORDERED decreasingly

$$\sum_{k=1}^{n}c_k^4=O\left(d\right).$$

- What would happen if we had chosen a different value of θ_k?

Fast Exact Sampling of Contingency Tables

- **Acceptance / Rejection**

$$\frac{dP}{dQ} I\left(\mathbf{r}_n = 0\right) \leq w\left(\mathbf{r}_0, \mathbf{c}_0\right)\beta,$$

$$\beta = \exp\left((3/4)\frac{1}{d}\sum_{k=1}^{n} c_k^2\right)\exp\left(\kappa\frac{1}{d}\sum_{k=1}^{n} c_k\right).$$

- **Algorithm**

1. Simulate according to $Q(\cdot)$ and let L be the likelihood ratio
2. Simulate $\widetilde{U}$ uniformly on $\left[0, w\left(\mathbf{r}_0, \mathbf{c}_0\right)\right]$ (independent of the table)
3. If $\widetilde{U} \leq L \cdot I\left(S_n = 0\right)/\beta$ accept the table, otherwise repeat Steps 1 and 2 until acceptance

Complexity of Exact Sampling of Contingency Tables

- **Acceptance probability**

$$P^{Q}_{\mathbf{r}_0,\mathbf{c}_0}\left(\widetilde{U} \leq L \cdot I\left(\mathbf{r}_n = 0\right)/\beta\right)$$

$$= \frac{E^{Q}_{\mathbf{r}_0,\mathbf{c}_0} L \cdot I\left(\mathbf{r}_n = 0\right)}{w\left(\mathbf{r}_0, \mathbf{c}_0\right)\beta} = \frac{P_{\mathbf{r}_0,\mathbf{c}_0}\left(\mathbf{r}_n = 0\right)}{w\left(\mathbf{r}_0, \mathbf{c}_0\right)\beta} \sim \frac{1}{\beta} \text{ as } d \nearrow \infty$$

- **Probability of getting a feasible table under Q ?**

$$\begin{aligned} P^{Q}_{\mathbf{r}_0,\mathbf{c}_0}\left(\mathbf{r}_n = 0\right) &= E \prod_{k=0}^{n-1} \frac{v\left(\mathbf{r}_{k+1}, \mathbf{c}_{k+1}\right)}{w\left(\mathbf{r}_k, \mathbf{c}_k\right)} I\left(\mathbf{r}_n = 0\right) \\ &= \frac{1}{w\left(\mathbf{r}_0, \mathbf{c}_0\right)} E \prod_{k=1}^{n-1} \frac{v\left(\mathbf{r}_k, \mathbf{c}_k\right)}{w\left(\mathbf{r}_k, \mathbf{c}_k\right)} I\left(\mathbf{r}_n = 0\right) \\ &\geq \frac{P\left(\mathbf{r}_n = 0\right)}{w\left(\mathbf{r}_0, \mathbf{c}_0\right)} \frac{1}{\beta} \sim \frac{1}{\beta} \text{ as } d \nearrow \infty \end{aligned}$$

Complexity of Exact Sampling of Contingency Tables

- **Conclusion:** *If* $\max\{c_k, r_k\} = o\left(d^{1/4}\right)$ *as* $d \nearrow \infty$ *and*

$$\max\left(\sum_{k=1}^{n} c_k^2, \sum_{j=1}^{m} r_j^2\right) = O\left(d\right).$$

Then, exact simulation of a binary table with given column and row sums can be done in $O\left(d^2 G\right)$ *operations, where* G *is Geometrically distributed with mean 2.*

Agenda

1. General Overview: Importance Sampling (IS) and Efficiency

Agenda

1. General Overview: Importance Sampling (IS) and Efficiency
2. Example: Binary Tables

Agenda

1. General Overview: Importance Sampling (IS) and Efficiency
2. Example: Binary Tables
3. Lyapunov Bounds, Efficiency and Exact Simulation

Agenda

Conclusions

- Approximate counting gives rise to interesting rare-event simulation problems

Conclusions

- Approximate counting gives rise to interesting rare-event simulation problems
- Rare-event simulation techniques provide insight into empirical performance of importance sampling for counting

Conclusions

- Approximate counting gives rise to interesting rare-event simulation problems
- Rare-event simulation techniques provide insight into empirical performance of importance sampling for counting
- Approximations –> design of statistical estimation / counting algorithms —> good complexity analysis

Conclusions

- Approximate counting gives rise to interesting rare-event simulation problems
- Rare-event simulation techniques provide insight into empirical performance of importance sampling for counting
- Approximations –> design of statistical estimation / counting algorithms —> good complexity analysis
- Efficient importance sampling —> efficient exact sampling from complex uniform distributions