Overview

• Introduction: Mixed graphical models

• SampleSearch: Sampling with Searching

• Exploiting structure in sampling: AND/OR
  Importance sampling
Overview

• Introduction: Mixed graphical models

• SampleSearch: Sampling with Searching

• Exploiting structure in sampling: AND/OR
  Importance sampling
Bayesian Networks: Representation

(Pearl, 1988)

Belief Updating:

\[ P(\text{lung cancer} = \text{yes} \mid \text{smoking} = \text{no, dyspnoea} = \text{yes}) = ? \]
Mixed Networks: Mixing Belief and Constraints

Belief or Bayesian Networks

Constraint Networks

Variables: A, B, C, D, E, F
Domains: \( D_A = D_B = D_C = D_D = D_E = D_F = \{0,1\} \)
CPTs: \( P(A), P(B | A), P(C | A), P(D | B, C) \)
\( P(E | A, B), P(F | A) \)

<table>
<thead>
<tr>
<th>B</th>
<th>C</th>
<th>D=0</th>
<th>D=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>.1</td>
<td>.9</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>.3</td>
<td>.7</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Constraints could be specified externally or may occur as zeros in the Belief network

Variables: A, B, C, D, E, F
Domains: \( D_A = D_B = D_C = D_D = D_E = D_F = \{0,1\} \)
Constraints: \( R_1(ABC), R_2(ACF), R_3(BCD), R_4(A, E) \)
Expresses the set of solutions: \( sol(R) \)

\( B = 0, C = 0, D = 0 \) is not allowed
\( B = 1, C = 1, D = 1 \) is not allowed
Mixed networks: Distribution and Queries

• The distribution represented by a mixed network $T=(B,R)$:

$$P_T(x) = \begin{cases} \frac{1}{M} P_B(x), & \text{if } x \in sol(R) \\ 0, & \text{otherwise} \end{cases}$$

• Queries:

  – **Weighted Counting** (Equivalent to $P(e)$, partition function, solution counting)

  $$M = \sum_{x \in sol(R)} P_B(x)$$

  – **Marginal distribution**: $P_T(X_i)$
Applications

• **Determinism**: More Ubiquitous than you may think!

• **Transportation Planning** (Liao et al. 2004, Gogate et al. 2005)
  – Predicting and Inferring Car Travel Activity of individuals

• **Genetic Linkage Analysis** (Fischelson and Geiger, 2002)
  – associate functionality of genes to their location on chromosomes.

• **Functional/Software Verification** (Bergeron, 2000)
  – Generating random test programs to check validity of hardware

• **First Order Probabilistic models** (Domingos et al. 2006, Milch et al. 2005)
  – Citation matching
Model for locus 1

Functions on Orange nodes are deterministic

Model for locus 2
Approximate Inference

• Approximations are hard with determinism
  • Randomized Polynomial $\varepsilon$-approximation possible when no zeros are present (Karp 1993, Cheng 2001)
  • $\varepsilon$-approximation NP-hard in the presence of zeros
  • Gibbs sampling is problematic when MCMC is not ergodic.

• Current remedies
  – Replace zeros with very small values (Laplace correction: Naive Bayes, NLP)
  – bad performance when zeros or determinism is real!
Overview

• Introduction: Mixed graphical models

• SampleSearch: Sampling with Searching
  – Rejection problem
  – Recovery, and analysis
  – Empirical evaluation

• Exploiting structure in sampling: AND/OR
  Importance sampling
**Importance Sampling: Overview**

\[
P(e) = \sum_{X \setminus E} P_B(X, E = e)
\]

\[
M = \sum_{z \in Z} f(z) = P(e) \text{ where } Z = X \setminus E
\]

- Given a proposal or importance distribution \(Q(z)\) such that \(f(z) > 0\) implies \(Q(z) > 0\), rewrite

\[
M = \sum_{z \in Z} f(z) = \sum_{z \in Z} f(z) \frac{Q(z)}{Q(z)} = E_Q \left[ \frac{f(z)}{Q(z)} \right]
\]

- Given i.i.d. samples \(z_1, \ldots, z_N\) from \(Q(z)\),

\[
\hat{M} = \frac{1}{N} \sum_{j=1}^{N} \frac{f(z_j)}{Q(z_j)} = \frac{1}{N} \sum_{j=1}^{N} w(z_j) \ E_Q[\hat{M}] = M = P(e)
\]
Generating i.i.d. samples from $Q$

$$Q(X) = Q(X_1) \times Q(X_2 \mid X_1) \times \ldots \times Q(X_n \mid X_1, \ldots, X_{n-1})$$

$$Q(A, B, C) = Q(A) \times Q(B \mid A) \times Q(C \mid A, B)$$

$$Q(A) = (0.8, 0.2), \; Q(B \mid A) = (0.4, 0.6, 0.2, 0.8), \; Q(C \mid A, B) = Q(C) = (0.2, 0.8)$$
Importance sampling: \( \hat{M} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{Q(x_i)} \)

\( f(x_i) = 0 \) if \( x_i \) is not a solution.

Toss a biased coin

\( P(H) = 0.8, P(T) = 0.2 \)

Say we get H
Rejection Problem

Importance sampling: $$\tilde{M} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{Q(x_i)}$$

$$f(x_i) = 0$$ if $$x_i$$ is not a solution.

Toss a biased coin
P(H)=0.4, P(T)=0.6
Say, We get a Head
Rejection Problem

Importance sampling: \( \tilde{M} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{Q(x_i)} \)

\( f(x_i) = 0 \) if \( x_i \) is not a solution.

Toss a biased coin
P(H)=0.4, P(T)=0.6
Say, We get a Head
Rejection Problem

Importance sampling:
\[
\tilde{M} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{Q(x_i)}
\]
for \( f(x_i) = 0 \) if \( x_i \) is not a solution.

Toss a biased coin
\( P(H)=0.4, \ P(T)=0.6 \)
Say, We get a Head

• A large number of assignments generated will be rejected, thrown away
Rejection Problem

Importance sampling: $\hat{M} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{Q(x_i)}$

$f(x_i) = 0$ if $x_i$ is not a solution.

Constraints: $A\neq B$, $A\neq C$

All Blue leaves correspond to solutions i.e. $f(x) > 0$
All Red leaves correspond to non-solutions i.e. $f(x) = 0$
Revising Q to backtrack-free distribution:

\[ Q^F(\text{branch}) = 0 \text{ if no solutions under it} \]
\[ Q^F(\text{branch}) = 1 \text{ if no solutions under the other branch} \]
\[ Q^F(\text{branch}) = Q(\text{branch}) \text{ otherwise} \]

Constraints: \( A \neq B \) \( A \neq C \)

All Blue leaves correspond to solutions i.e. \( f(x) > 0 \)
All Red leaves correspond to non-solutions i.e. \( f(x) = 0 \)
Generating samples from $Q^F$

Constraints: $A \neq B$, $A \neq C$

$Q^F(\text{branch}) = 0$ if no solutions under it
$Q^F(\text{branch}) = 1$ if no solutions under the other branch
$Q^F(\text{branch}) = Q(\text{branch})$ otherwise

- Invoke an oracle at each branch.
- Oracle returns True if there is a solution under a branch
- False, otherwise
Generating samples from $Q^F$

Constraints: $A \neq B, A \neq C$

- Oracles: In practice
  - Adaptive consistency as preprocessing step
  - A complete CSP solver
- Too costly
  - $O(\exp(\text{treewidth}))$
  - Invoked $O(nd)$ for each sample
Approximations of $Q^F$

- Use i-consistency instead of adaptive consistency
  - $O(n^i)$ time and space complexity
  - Identify some zeros so that they are never sampled
- Cons: Too weak when constraint portion is hard.

---

**Rejection Rate**

- No processing
- Polynomial consistency enforcement
- NP-hard exponential time and space

**Percentage of Zeros identified**
Algorithm SampleSearch

Constraints: \( A \neq B \) \( A \neq C \)

\[
\hat{M} = \frac{1}{N} \sum_{j=1}^{N} \frac{f(z_j)}{Q(z_j)}
\]
Algorithm SampleSearch

Constraints: A≠B A≠C

\[ \hat{M} = \frac{1}{N} \sum_{j=1}^{N} \frac{f(z_j)}{Q(z_j)} \]
Algorithm SampleSearch

Constraints: A≠B A≠C

\[
\hat{M} = \frac{1}{N} \sum_{j=1}^{N} \frac{f(z_j)}{Q(z_j)}
\]

Gogate and Dechter, AISTATS 2007, AAAI 2007
Algorithm SampleSearch

Constraints: $A \neq B$ $A \neq C$

Until $f(\text{sample}) > 0$
Generate more Samples

Constraints: A ≠ B A ≠ C

$$\hat{M} = \frac{1}{N} \sum_{j=1}^{N} \frac{f(z_j)}{Q(z_j)}$$
Generate more Samples

Constraints: A ≠ B, A ≠ C

\[ \hat{M} = \frac{1}{N} \sum_{j=1}^{N} \frac{f(z_j)}{Q(z_j)} \]
Traces of SampleSearch

Constraints: A≠B A≠C
SampleSearch: Sampling Distribution

- Problem: Due to Search, the samples are no longer i.i.d. from $Q$

\[
\hat{M} = \frac{1}{N} \sum_{j=1}^{N} \frac{f(z_j)}{Q(z_j)}, \quad E_Q[\hat{M}] \neq M
\]

- Theorem: SampleSearch generates i.i.d. samples from the backtrack-free distribution $Q^F$:

\[
\overline{M} = \frac{1}{N} \sum_{j=1}^{N} \frac{f(z_j)}{Q^F(z_j)} \quad E_{Q^F}[\overline{M}] = M
\]
The Sampling distribution $Q^F$ of SampleSearch

Constraints: $A \neq B \neq C$

What is probability of generating $A=0$?

$Q^F(A=0)=0.8$

Why? SampleSearch is systematic

What is probability of generating $(A=0,B=1)$?

$Q^F(B=1|A=0)=1$

Why? SampleSearch is systematic

What is probability of generating $(A=0,B=0)$?

Simple: $Q^F(B=0|A=0)=0$

All samples generated by SampleSearch are solutions

Backtrack-free distribution
Asymptotic approximations of $Q^F$

- **IF Hole THEN**
  - $U^F=Q$ (i.e. there is a solution at the other branch)
  - $L^F=0$ (i.e. no solution at the other branch)

- $Q^F(\text{branch})=0$ if no solutions under it
- $Q^F(\text{branch})=Q(\text{branch})$ otherwise
Approximations: Convergence in the limit

- Store all possible traces
Approximations:
Convergence in the limit

- From the combined sample tree, update $U$ and $L$.

\[
\text{IF Hole THEN } U_N^F = Q \text{ and } L_N^F = 0
\]

\[
M = \sum_{x \in X} f(x) = E \left[ \frac{f(x)}{Q^F(x)} \right]
\]

\[
\lim_{N \to \infty} E \left[ \frac{f(x)}{U_N^F(x)} \right] = \lim_{N \to \infty} E \left[ \frac{f(x)}{L_N^F(x)} \right] = M
\]

Asymptotically unbiased

Bounding \( U_N^F(x) \leq Q^F(x) \leq L_N^F(x) \)

Gogate and Dechter, AISTATS 2007, AAAI 2007
Improving Naive SampleSearch: The IJGP-wc-SS algorithm

- **Better Search Strategy**
  - Can use any state-of-the-art CSP/SAT solver e.g. minisat (Een and Sorrenson 2006)

- **Better Proposal distribution**
  - Use output of IJGP- a generalized belief propagation to compute the initial importance function

- **w-cutset importance sampling** (Bidyuk and Dechter, 2007)
  - Reduce variance by sampling from a subspace
Experiments

• **Tasks**
  – Weighted Counting
  – Marginals

• **Benchmarks**
  – Satisfiability problems (counting solutions)
  – Linkage networks
  – Relational instances (First order probabilistic networks)
  – Grid networks
  – Logistics planning instances

• **Algorithms**
  – IJGP-wc-SS/LB and IJGP-wc-SS/UB
  – IJGP-wc-IS (Vanilla algorithm that does not perform search)
  – SampleCount (Gomes et al. 2007, SAT)
  – ApproxCount (Wei and Selman, 2007, SAT)
  – EPIS (Changhe and Druzdzel, 2006)
  – RELSAT (Bayardo and Peshoushke, 2000, SAT)
  – Edge Deletion Belief Propagation (Choi and Darwiche, 2006)
  – Iterative Join Graph Propagation (Dechter et al., 2002)
  – Variable Elimination and Conditioning (VEC)
### Results: Probability of Evidence

**Linkage instances (UAI 2006 evaluation)**

<table>
<thead>
<tr>
<th>Problem</th>
<th>( \langle n, k, c, e, w \rangle )</th>
<th>Exact</th>
<th>IJGP-wc-SS/LB</th>
<th>IJGP-wc-SS/UB</th>
<th>VEC</th>
<th>EDBP</th>
<th>IJGP-wc-IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN_69</td>
<td>( \langle 777, 7, 228, 78, 47 \rangle )</td>
<td>( Z )</td>
<td>5.28E-054</td>
<td>3.00E-55</td>
<td>6.84E+5</td>
<td>3.00E-55</td>
<td>6.84E+5</td>
</tr>
<tr>
<td>BN_70</td>
<td>( \langle 2315, 5, 484, 159, 87 \rangle )</td>
<td>( Z )</td>
<td>2.00E-71</td>
<td>1.21E-73</td>
<td>1.92E+5</td>
<td>1.21E-73</td>
<td>1.92E+5</td>
</tr>
<tr>
<td>BN_71</td>
<td>( \langle 1740, 6, 663, 202, 70 \rangle )</td>
<td>( Z )</td>
<td>5.12E-111</td>
<td>1.28E-111</td>
<td>7.46E+4</td>
<td>1.28E-111</td>
<td>7.46E+4</td>
</tr>
<tr>
<td>BN_72</td>
<td>( \langle 2155, 6, 752, 252, 86 \rangle )</td>
<td>( Z )</td>
<td>4.21E-150</td>
<td>4.73E-150</td>
<td>1.53E+5</td>
<td>4.73E-150</td>
<td>1.53E+5</td>
</tr>
<tr>
<td>BN_73</td>
<td>( \langle 2140, 5, 651, 216, 101 \rangle )</td>
<td>( Z )</td>
<td>2.26E-113</td>
<td>2.00E-115</td>
<td>7.75E+4</td>
<td>2.00E-115</td>
<td>7.75E+4</td>
</tr>
<tr>
<td>BN_74</td>
<td>( \langle 749, 6, 223, 66, 45 \rangle )</td>
<td>( Z )</td>
<td>3.75E-45</td>
<td>2.13E-46</td>
<td>2.80E+5</td>
<td>2.13E-46</td>
<td>2.80E+5</td>
</tr>
<tr>
<td>BN_75</td>
<td>( \langle 1820, 5, 477, 155, 92 \rangle )</td>
<td>( Z )</td>
<td>5.88E-91</td>
<td>2.19E-91</td>
<td>7.72E+4</td>
<td>2.19E-91</td>
<td>7.72E+4</td>
</tr>
<tr>
<td>BN_76</td>
<td>( \langle 2155, 7, 605, 169, 64 \rangle )</td>
<td>( Z )</td>
<td>4.93E-110</td>
<td>1.95E-111</td>
<td>2.52E+4</td>
<td>1.95E-111</td>
<td>2.52E+4</td>
</tr>
</tbody>
</table>

**Time Bound**: 3 hrs  
**M**: number of samples generated in 10 hrs  
**Z**: Probability of Evidence
Results: Probability of Evidence
Relational instances (UAI 2008 evaluation)

<table>
<thead>
<tr>
<th>Problem</th>
<th>(&lt;n, k, c, e, w&gt;)</th>
<th>Exact</th>
<th>IJGP-wc -SS/LB</th>
<th>IJGP-wc -SS/UB</th>
<th>VEC</th>
<th>EDBP</th>
<th>IJGP-wc -IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friends and Smokers</td>
<td>(Z)</td>
<td>(M)</td>
<td>(1.52E-05)</td>
<td>(8.11E-06)</td>
<td>(8.11E-06)</td>
<td>(1.53E-05^{(1)})</td>
<td>X</td>
</tr>
<tr>
<td>(fs-04)</td>
<td>((262, 2, 74, 226, 12))</td>
<td>(1.00E+6)</td>
<td>(1.00E+6)</td>
<td>(1.00E+6)</td>
<td>(2.17E+8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(fs-07)</td>
<td>((1225, 2, 371, 1120, 35))</td>
<td>(9.80E-17)</td>
<td>(2.23E-16)</td>
<td>(2.23E-16)</td>
<td>(1.78E-15^{(7)})</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(fs-10)</td>
<td>((3385, 2, 1055, 3175, 71))</td>
<td>(7.88E-31)</td>
<td>(2.49E-32)</td>
<td>(2.49E-32)</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>(fs-13)</td>
<td>((7228, 2, 2288, 6877, 119))</td>
<td>(1.33E-51)</td>
<td>(3.26E-55)</td>
<td>(3.26E-55)</td>
<td>(1.33E-51)</td>
<td>4.67E+7</td>
<td></td>
</tr>
<tr>
<td>(fs-16)</td>
<td>((13240, 2, 4232, 12712, 171))</td>
<td>(8.63E-78)</td>
<td>(6.04E-79)</td>
<td>(6.04E-79)</td>
<td>(8.63E-78)</td>
<td>1.37E+7</td>
<td></td>
</tr>
<tr>
<td>(fs-19)</td>
<td>((21907, 2, 7049, 21166, 243))</td>
<td>(2.12E-109)</td>
<td>(1.62E-114)</td>
<td>(1.62E-114)</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>(fs-22)</td>
<td>((33715, 2, 10901, 32725, 335))</td>
<td>(2.00E-146)</td>
<td>(4.88E-147)</td>
<td>(4.88E-147)</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>(fs-25)</td>
<td>((49150, 2, 15950, 47875, 431))</td>
<td>(7.18E-189)</td>
<td>(2.67E-189)</td>
<td>(2.67E-189)</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>(fs-28)</td>
<td>((68698, 2, 22358, 67102, 527))</td>
<td>(9.82E-237)</td>
<td>(4.53E-237)</td>
<td>(4.53E-237)</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>(fs-29)</td>
<td>((76212, 2, 24824, 74501, 559))</td>
<td>(6.81E-254)</td>
<td>(9.44E-255)</td>
<td>(9.44E-255)</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
</tbody>
</table>

Time Bound: 10 hrs
M: number of samples generated in 10 hrs
Z: Probability of Evidence
Approximation Error vs Time for BN_76, num-vars=2155

Probability of Evidence

Time in seconds

Exact  EDBP  IJGP-wc-SS/UB  VEC  IJGP-wc-SS/LB
## Results: Solution Counts

### Latin Square instances (size 8 to 16)

<table>
<thead>
<tr>
<th>Problem</th>
<th>(\langle n, k, c, w \rangle)</th>
<th>Exact</th>
<th>Sample Count</th>
<th>Approx Count</th>
<th>REL SAT</th>
<th>IJGP-wc-SS/LB</th>
<th>IJGP-wc-SS/UB</th>
<th>IJGP-wc-IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ls8-norm</td>
<td>(\langle 512, 2, 5584, 255 \rangle)</td>
<td>(Z)</td>
<td>5.40E11</td>
<td>5.15E+11</td>
<td>3.52E+11</td>
<td>2.44E+08</td>
<td>5.91E+11</td>
<td>5.91E+11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(M)</td>
<td></td>
<td>16514</td>
<td>17740</td>
<td></td>
<td>236510</td>
<td>236510</td>
</tr>
<tr>
<td>ls9-norm</td>
<td>(\langle 729, 2, 9009, 363 \rangle)</td>
<td>(Z)</td>
<td>3.80E17</td>
<td>4.49E+17</td>
<td>1.26E+17</td>
<td>1.78E+08</td>
<td>3.44E+17</td>
<td>3.44E+17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(M)</td>
<td></td>
<td>7762</td>
<td>8475</td>
<td></td>
<td>138572</td>
<td>138572</td>
</tr>
<tr>
<td>ls10-norm</td>
<td>(\langle 1000, 2, 13820, 676 \rangle)</td>
<td>(Z)</td>
<td>7.60E24</td>
<td>7.28E+24</td>
<td>1.17E+24</td>
<td>1.36E+08</td>
<td>6.74E+24</td>
<td>6.74E+24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(M)</td>
<td></td>
<td>3854</td>
<td>4313</td>
<td></td>
<td>95567</td>
<td>95567</td>
</tr>
<tr>
<td>ls11-norm</td>
<td>(\langle 1331, 2, 20350, 956 \rangle)</td>
<td>(Z)</td>
<td>5.40E33</td>
<td>2.08E+34</td>
<td>4.91E+31</td>
<td>1.09E+08</td>
<td>3.87E+33</td>
<td>3.87E+33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(M)</td>
<td></td>
<td>2002</td>
<td>2289</td>
<td></td>
<td>66795</td>
<td>66795</td>
</tr>
</tbody>
</table>

**Time Bound:** 10 hrs

**M:** number of samples generated in 10 hrs

**Z:** Solution Counts
### Results: Solution Counts

#### Langford instances

| Problem   | \(|n, k, c, w|\) | Exact | Sample Count | Approx Count | REL SAT | IJGP-wc-SS/LB | IJGP-wc-SS/UB | IJGP-wc-UB |
|-----------|-----------------|-------|--------------|--------------|---------|----------------|----------------|------------|
| lang12    | \(576, 2, 13584, 383\) | Z     | 2.16E+5      | 1.93E+05     | 2.95E+04 | 2.16E+05*(297s) | 2.16E+05      | 999991     |
| lang16    | \(1024, 2, 32320, 639\) | Z     | 6.53E+08     | 5.97E+08     | 8.22E+06 | 6.28E+06        | 6.51E+08      | 14971      |
| lang19    | \(1444, 2, 54226, 927\) | Z     | 5.13E+11     | 9.73E+10     | 6.87E+08 | 8.52E+05        | 6.38E+11      | 3431       |
| lang20    | \(1600, 2, 63280, 1023\) | Z     | 5.27E+12     | 1.13E+11     | 3.99E+09 | 8.55E+04        | 2.83E+12      | 2961       |
| lang23    | \(2116, 2, 96370, 1407\) | Z     | 7.60E+15     | 7.53E+14     | 3.70E+12 | X               | 4.17E+15      | 1111       |
| lang24    | \(2304, 2, 109536, 1535\) | Z     | 9.37E+16     | 1.17E+13     | 4.15E+11 | X               | 8.74E+15      | 271        |

**Time Bound: 10 hrs**

**M:** number of samples generated in 10 hrs

**Z:** Solution Counts
Results on Marginals

• Evaluation Criteria

\[ \text{Exact : } P(x_i) \quad \text{Approximate : } A(x_i) \]

\[
\text{Hellinger distance} = \frac{1}{n} \sum_{i=1}^{n} \sum_{x_i \in D_i} \left( \sqrt{P(x_i)} - \sqrt{A(x_i)} \right)^2
\]

• Always bounded between 0 and 1
• Lower Bounds the KL distance
• When probabilities close to zero are present KL distance may tend to infinity.
## Results: Posterior Marginals

**Linkage instances (UAI 2006 evaluation)**

<table>
<thead>
<tr>
<th>Problem</th>
<th>$\langle n, k, c, e, w \rangle$</th>
<th>$\Delta$</th>
<th>$M$</th>
<th>IJGP-wc-SS</th>
<th>IJGP</th>
<th>EPIS</th>
<th>EDBP</th>
<th>IJGP-wc-IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN_69</td>
<td>$\langle 777, 7, 228, 78, 47 \rangle$</td>
<td>9.4E-04</td>
<td>6.84E+5</td>
<td>3.2E-02</td>
<td>8.0E-02</td>
<td>1</td>
<td>8.0E-02</td>
<td>1</td>
</tr>
<tr>
<td>BN_70</td>
<td>$\langle 2315, 5, 484, 159, 87 \rangle$</td>
<td>2.6E-03</td>
<td>1.92E+5</td>
<td>3.3E-02</td>
<td>9.6E-02</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>BN_71</td>
<td>$\langle 1740, 6, 663, 202, 70 \rangle$</td>
<td>5.6E-03</td>
<td>7.46E+4</td>
<td>1.9E-02</td>
<td>2.5E-02</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>BN_72</td>
<td>$\langle 2155, 6, 752, 252, 86 \rangle$</td>
<td>3.6E-03</td>
<td>1.53E+5</td>
<td>7.2E-03</td>
<td>1.3E-02</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>BN_73</td>
<td>$\langle 2140, 5, 651, 216, 101 \rangle$</td>
<td>2.1E-02</td>
<td>7.75E+4</td>
<td>2.8E-02</td>
<td>6.1E-02</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>BN_74</td>
<td>$\langle 749, 6, 223, 66, 45 \rangle$</td>
<td>6.9E-04</td>
<td>2.80E+5</td>
<td>4.3E-06</td>
<td>1</td>
<td>4.3E-02</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>BN_75</td>
<td>$\langle 1820, 5, 477, 155, 92 \rangle$</td>
<td>8.0E-03</td>
<td>7.72E+4</td>
<td>6.2E-02</td>
<td>1</td>
<td>9.3E-02</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>BN_76</td>
<td>$\langle 2155, 7, 605, 169, 64 \rangle$</td>
<td>1.8E-02</td>
<td>2.52E+4</td>
<td>2.6E-02</td>
<td>2.7E-02</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

**Time Bound:** 3 hrs

Table shows the Hellinger distance ($\Delta$) and Number of samples: $M$
Summary: SampleSearch

• Manages rejection problem while sampling

• Sampling Distribution is the backtrack-free distribution $Q^F$

• Approximation of $Q^F$ by storing all traces yielding an asymptotically unbiased estimator
  – Linear time and space overhead
  – Bound the weighted counts from above and below

• Empirically, when a substantial number of zero probabilities are present, SampleSearch dominate.
Overview

• Introduction: Mixed graphical models

• SampleSearch: Sampling with Searching

• Exploiting structure in sampling: AND/OR Importance sampling
Motivation

\[ M = \sum_{z} f(z) = \sum_{x_1, x_2, x_3, x_4} f_A(x_1, x_2, x_3) f_B(x_2, x_3, x_4) f_C(x_4) \]

\[ M = \sum_{z} f(z) = \sum_{x_1, x_2, x_3, x_4} f_A(x_1) f_B(x_2) f_C(x_3, x_4) \]

Given \( Q(z) \), Importance sampling totally disregards the structure of \( f(z) \) while approximating it.

\[ M = \sum_{z \in Z} f(z) = \sum_{z \in Z} f(z) \frac{Q(z)}{Q(z)} = E_Q \left[ \frac{f(z)}{Q(z)} \right] \]
OR Search Tree

Constraint Satisfaction – Counting Solutions

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>R_{ABC}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>C</th>
<th>D</th>
<th>R_{BCD}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>E</th>
<th>R_{ABE}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>E</th>
<th>F</th>
<th>R_{AEF}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Diagram of the OR Search Tree showing the 2^6 = 64 solutions for the given constraint set.
AND/OR Search Tree

Constraint Satisfaction – Counting Solutions

pseudo tree
AND/OR Tree

**OR**  –  casing upon variable values

**AND**  –  decomposition into independent subproblems
## Complexity of AND/OR Tree Search

<table>
<thead>
<tr>
<th></th>
<th>AND/OR tree</th>
<th>OR tree</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space</strong></td>
<td>$O(n)$</td>
<td>$O(n)$</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>$O(n \ k^m)$</td>
<td>$O(k^n)$</td>
</tr>
<tr>
<td></td>
<td>$O(n \ k^{w^* \log n})$</td>
<td></td>
</tr>
</tbody>
</table>

[Freuder & Quinn85], [Collin, Dechter & Katz91], [Bayardo & Miranker95], [Darwiche01]

$k = \text{domain size}$  
$m = \text{depth of pseudo-tree}$  
$n = \text{number of variables}$  
$w^* = \text{treewidth}$
Background: AND/OR search space

Problem

Chain Pseudo-tree

Pseudo-tree

AND/OR Tree

OR Tree
Example Bayesian network

\[ M = P(a, b) = \sum_{X,Y,Z} f(X|Z)f(Y|Z) f(Z) \]
Recap:

Conventional Importance Sampling

\[ M = \sum_{XYZ} f(XZ)f(YZ)f(Z) \]

\[ Q(XYZ) = Q(Z)Q(X \mid Z)Q(Y \mid Z) \]

\[ M = \sum_{XYZ} \frac{f(XZ)f(YZ)f(Z)}{Q(Z)Q(X \mid Z)Q(Y \mid Z)} \cdot Q(Z)Q(X \mid Z)Q(Y \mid Z) \]

\[ = E_Q \left[ \frac{f(XZ)f(YZ)f(Z)}{Q(Z)Q(X \mid Z)Q(Y \mid Z)} \right] \]

AND/OR idea!
Decompose this expectation
AND/OR Importance Sampling
(General Idea)

• Decompose Expectation

\[
M = \sum_{XYZ} \frac{f(XZ)f(YZ)f(Z)}{Q(Z)Q(X \mid Z)Q(Y \mid Z)} Q(Z)Q(X \mid Z)Q(Y \mid Z)
\]

\[
M = \left( \sum_{Z} \frac{f(Z)}{Q(Z)} Q(Z) \left( \sum_{X} \frac{f(XZ)}{Q(X \mid Z)} Q(X \mid Z) \right) \left( \sum_{Y} \frac{f(YZ)}{Q(Y \mid Z)} Q(Y \mid Z) \right) \right)
\]

\[
M = E_Q \left[ \frac{f(Z)}{Q(Z)} \right] E_Q \left[ \frac{f(XZ)}{Q(X \mid Z)} \mid Z \right] E_Q \left[ \frac{f(YZ)}{Q(Y \mid Z)} \mid Z \right]
\]
AND/OR Importance Sampling (General Idea)

\[ M = E_Q \left[ \frac{f(Z)}{Q(Z)} \right] E_Q \left[ \frac{f(XZ)}{Q(X | Z)} | Z \right] E_Q \left[ \frac{f(YZ)}{Q(Y | Z)} | Z \right] \]

- Estimate all conditional expectations separately
- How?
  - Record all samples
  - For each sample that has Z=j
    - Estimate the conditional expectations X|Z and Y|Z using samples corresponding to X|Z=j and Y|Z=j respectively.
    - Combine the results
AND/OR Importance Sampling

\[ M = E_Q \left[ \frac{f(Z)}{Q(Z)} \right] E_Q \left[ \frac{f(XZ)}{Q(X | Z)} \right] | Z \] \[ E_Q \left[ \frac{f(YZ)}{Q(Y | Z)} \right] | Z \]

Estimate of \[ E \left[ \frac{f(X, Z = 0)}{Q(X | Z = 0)} \right] | Z = 0 \]

= Average Weight of samples on X having Z = 0

= \( w(x = 1, z = 0) + w(x = 2, z = 0) / 2 \)

Sample # | Z | X | Y
---|---|---|---
1 | 0 | 1 | 0
2 | 0 | 2 | 1
3 | 1 | 1 | 1
4 | 1 | 2 | 0

Pseudo-tree
AND/OR Importance Sampling

All AND nodes: Separate Components. Take Product

**Operator: Product**

All OR nodes: Conditional Expectations given the assignment above it

**Operator: Average**
Algorithm AND/OR Importance Sampling

1. Generate samples $x_1, \ldots, x_N$ from $Q$ along $O$.
2. Build a AND/OR sample tree for the samples $x_1, \ldots, x_N$ along the ordering $O$.
3. FOR all leaf nodes $i$ of AND-OR tree do
   1. IF AND-node $v(i) = 1$ ELSE $v(i) = 0$
4. FOR every node $n$ from leaves to the root do
   1. IF AND-node $v(n) =$ product of children
   2. IF OR-node $v(n) =$ Average of children
5. Return $v(\text{root-node})$
Properties of AND/OR Importance Sampling

• Unbiased estimate of weighted counts.
• AND/OR estimate has lower Variance than conventional importance sampling estimate.

• Variance Reduction
  – Easy to Prove for case of complete independence (Goodman, 1960)
  – Complicated to prove for general conditional independence case (Gogate thesis, papers!)
AND/OR w-cutset (Rao-Blackwellised) sampling

- Rao-Blackwellisation (Rao, 1963)
  - Partition $X$ into $K$ and $R$, such that we can compute $P(R|k)$ efficiently.
  - Sample from $K$ and sum out $R$
  - Estimate:

$$\hat{M}_{RB} = \frac{1}{N} \sum_{i=1}^{N} \sum_{R} P(R|k_i) Q(k_i)$$

- w-cutset sampling (Bidyuk and Dechter, 2003): Select $K$ such that the treewidth of $R$ after removing $K$ is bounded by “$w$”. 

Weighted Counts conditioned on $K=k_i$
**AND/OR w-cutset sampling**

- Perform AND/OR tree or graph sampling on $K$
- Exact inference on $R$
- Orthogonal approaches:
  - Theorem: Combining AND/OR sampling and w-cutset sampling yields further variance reduction.

Graphical model

Full pseudo tree

Start pseudo tree on the cutset variables

OR pseudo tree on the cutset variables
From Search Trees to Search Graphs

- Any two nodes that root identical subtrees (subgraphs) can be merged
From Search Trees to Search Graphs

- Any two nodes that root identical subtrees (subgraphs) can be merged
Merging Based on **Context**

- One way of recognizing nodes that can be merged:

\[
\text{context } (X) = \text{ancestors of } X \text{ in pseudo tree that are connected to } X, \text{ or to descendants of } X
\]
**AND/OR Graphs**

- **AND/OR Tree**
- **AND/OR Graph**
AND/OR graph sampling

\[
\begin{align*}
\text{AND/OR Sample}
\end{align*}
\]

\[
\begin{align*}
\text{Tree} & \quad \text{Graph}
\end{align*}
\]
Variance Hierarchy and Complexity

- **IS**
  - \(O(1)\)
  - \(O(nN)\)
  - AND/OR Tree IS
    - \(O(h)\)
    - \(O(n^{N_w*})\)
  - \(O(cN + (n-c)Nexp(w))\)
  - \(O(n^{N_w*} + (n-c)Nexp(w))\)

- **AND/OR w-cutset IS**
  - \(O(1)\)
  - \(O(cN + (n-c)Nexp(w))\)
  - \(O(h + (n-c)exp(w))\)

- **AND/OR w-cutset Tree IS**
  - \(O(1)\)
  - \(O(nN)\)
  - \(O(cN + (n-c)exp(w))\)

- **AND/OR Graph IS**
  - \(O(1)\)
  - \(O(nN)\)
  - \(O(cNw* + (n-c)Nexp(w))\)
  - \(O(cN + (n-c)exp(w))\)
Experiments

- Benchmarks
  - Linkage analysis
  - Graph coloring
  - Grids
- Algorithms
  - OR tree sampling
  - AND/OR tree sampling
  - AND/OR graph sampling
  - w-cutset versions of the three schemes above
## Results: Probability of Evidence

### Linkage instances (UAI 2006 evaluation)

<table>
<thead>
<tr>
<th>Problem</th>
<th>( (n, k, E, t^*, c) )</th>
<th>Exact</th>
<th>or-tree-IS ( \Delta )</th>
<th>ao-tree-IS ( \Delta )</th>
<th>ao-graph-IS ( \Delta )</th>
<th>or-wc-tree-IS ( \Delta )</th>
<th>ao-wc-tree-IS ( \Delta )</th>
<th>ao-wc-graph-IS ( \Delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN_69.uai</td>
<td>( (777, 7, 78, 47, 59) )</td>
<td>5.28E-54</td>
<td>2.26E-02</td>
<td>2.46E-02</td>
<td>2.43E-02</td>
<td>2.42E-02</td>
<td>2.34E-02</td>
<td>4.22E-03</td>
</tr>
<tr>
<td>BN_70.uai</td>
<td>( (2315, 5, 159, 87, 98) )</td>
<td>2.00E-71</td>
<td>6.32E-02</td>
<td>7.25E-02</td>
<td>5.12E-02</td>
<td>8.18E-02</td>
<td>5.36E-02</td>
<td>2.62E-02</td>
</tr>
<tr>
<td>BN_71.uai</td>
<td>( (1740, 6, 202, 70, 139) )</td>
<td>5.12E-111</td>
<td>6.74E-02</td>
<td>5.51E-02</td>
<td>2.35E-02</td>
<td>8.58E-02</td>
<td>9.46E-03</td>
<td>1.21E-02</td>
</tr>
<tr>
<td>BN_72.uai</td>
<td>( (2155, 6, 252, 86, 88) )</td>
<td>4.21E-150</td>
<td>3.19E-02</td>
<td>4.61E-02</td>
<td>2.46E-03</td>
<td>6.12E-02</td>
<td>1.41E-03</td>
<td>2.63E-03</td>
</tr>
<tr>
<td>BN_73.uai</td>
<td>( (2140, 5, 216, 101, 149) )</td>
<td>2.26E-113</td>
<td>1.18E-01</td>
<td>1.12E-01</td>
<td>4.55E-02</td>
<td>1.58E-01</td>
<td>3.54E-02</td>
<td>3.95E-02</td>
</tr>
<tr>
<td>BN_74.uai</td>
<td>( (749, 6, 66, 45, 72) )</td>
<td>3.75E-45</td>
<td>5.34E-02</td>
<td>4.31E-02</td>
<td>2.87E-02</td>
<td>8.08E-02</td>
<td>2.83E-02</td>
<td>2.76E-02</td>
</tr>
<tr>
<td>BN_75.uai</td>
<td>( (1820, 5, 155, 92, 131) )</td>
<td>5.88E-91</td>
<td>4.47E-02</td>
<td>8.15E-02</td>
<td>4.73E-02</td>
<td>7.28E-02</td>
<td>4.20E-02</td>
<td>7.60E-03</td>
</tr>
<tr>
<td>BN_76.uai</td>
<td>( (2155, 7, 169, 64, 239) )</td>
<td>4.93E-110</td>
<td>1.07E-01</td>
<td>1.39E-01</td>
<td>6.95E-02</td>
<td>1.13E-01</td>
<td>5.03E-02</td>
<td>2.26E-02</td>
</tr>
<tr>
<td>BN_77.uai</td>
<td>( (1020, 9, 135, 22, 97) )</td>
<td>6.88E-79</td>
<td>1.06E-01</td>
<td>9.38E-02</td>
<td>8.26E-02</td>
<td>1.24E-01</td>
<td>6.75E-02</td>
<td>3.27E-02</td>
</tr>
</tbody>
</table>

**Time Bound: 1hr**
Summary: AND/OR Importance sampling

• AND/OR sampling: A general scheme to exploit conditional independence in sampling

• Theoretical guarantees: lower sampling error than conventional sampling

• Variance reduction orthogonal to Rao-Blackwellised sampling.

• Better empirical performance than conventional sampling.