Overview

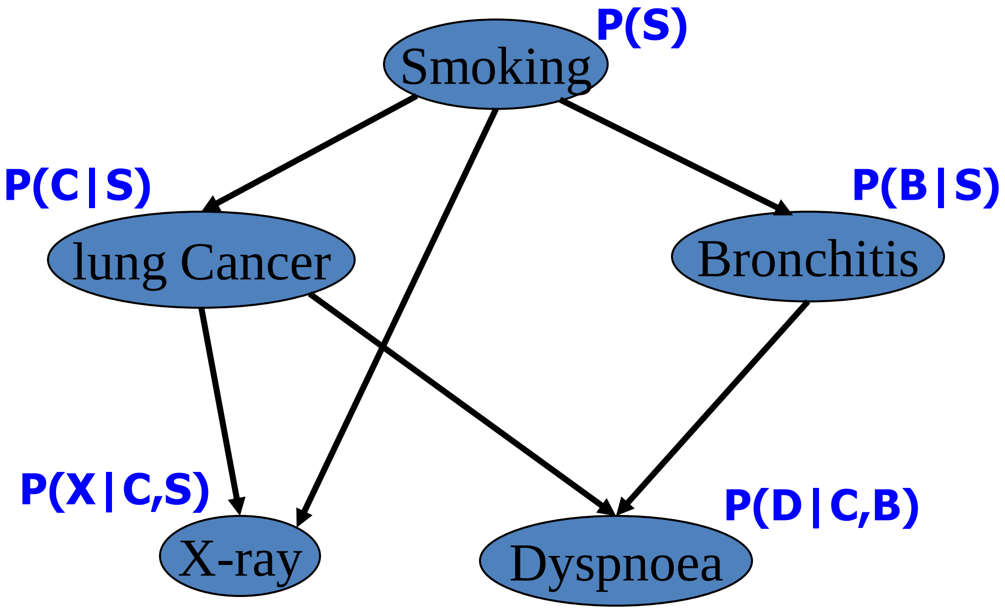
- Introduction: Mixed graphical models
- SampleSearch: Sampling with Searching
- Exploiting structure in samplinig: AND/OR Importance sampling

Overview

- Introduction: Mixed graphical models
- SampleSearch: Sampling with Searching
- Exploiting structure in samplinig: AND/OR Importance sampling

Bayesian Networks: Representation

(Pearl, 1988)



CPTs: $P(X_i \mid pa(X_i))$

$$P(X) = \prod_{i=1}^{n} P(X_i \mid pa(X_i))$$

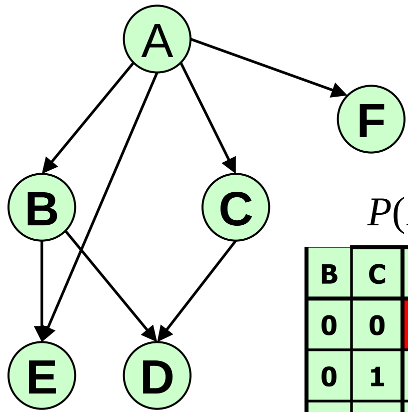
$P(S, C, B, X, D) = P(S)\, P(C|S)\, P(B|S)\, P(X|C,S)\, P(D|C,B)$

Belief Updating:

P (lung cancer=yes | smoking=no, dyspnoea=yes) = ?

Mixed Networks: Mixing Belief and Constraints

Belief or Bayesian Networks



$P(D \mid B, C)$

B	C	D=0	D=1
0	0	0	1
0	1	.1	.9
1	0	.3	.7
1	1	1	0

B=

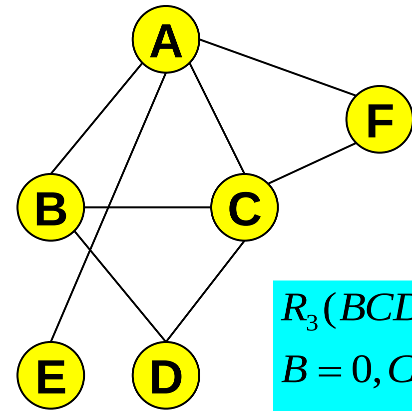
Variables : A, B, C, D, E, F

Domains : $D_A = D_B = D_C = D_D = D_E = D_F = \{0,1\}$

CPTS: $P(A), P(B \mid A), P(C \mid A), P(D \mid B, C)$

$P(E \mid A, B), P(F \mid A)$

Constraint Networks



$R_3(BCD)$

$B = 0, C = 0, D = 0$ is not allowed

$B = 1, C = 1, D = 1$ is not allowed

R=

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 : $sol(R)$

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

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)

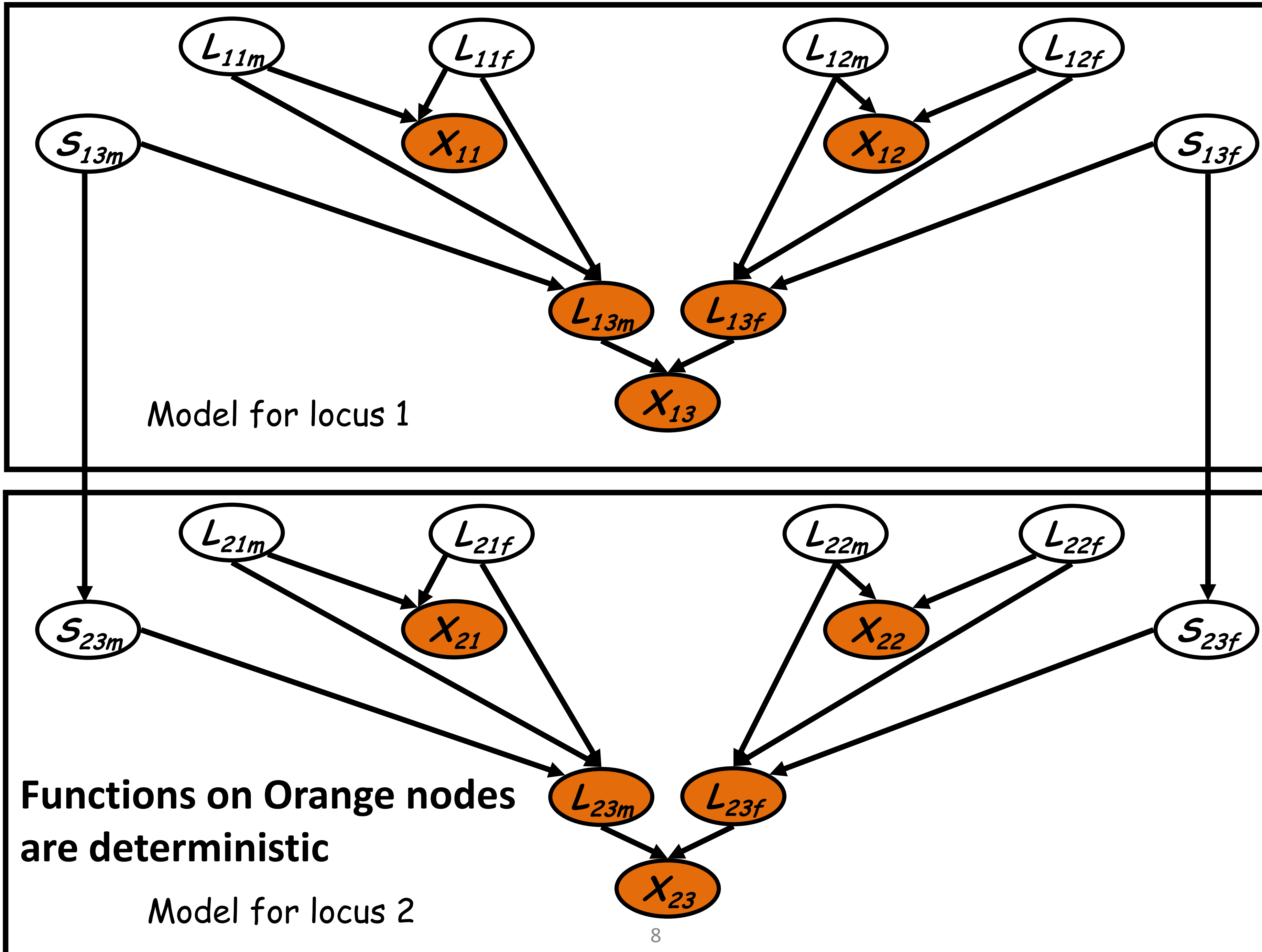
$$M = \sum_{x \in sol(R)} P_B(x)$$

  - 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



Functions on Orange nodes are deterministic

Approximate Inference

- Approximations are hard with determinism
  - Randomized Polynomial ε-approximation possible when no zeros are present (Karp 1993, Cheng 2001)
  - ε-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 samplinig: AND/OR Importance sampling

Importance Sampling: Overview

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

$$M=\sum_{z\in Z}f(z)=\mathrm{P(e)}\ where\ Z=X\setminus E$$

- 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 $z_1,..,z_N$ from Q(z),

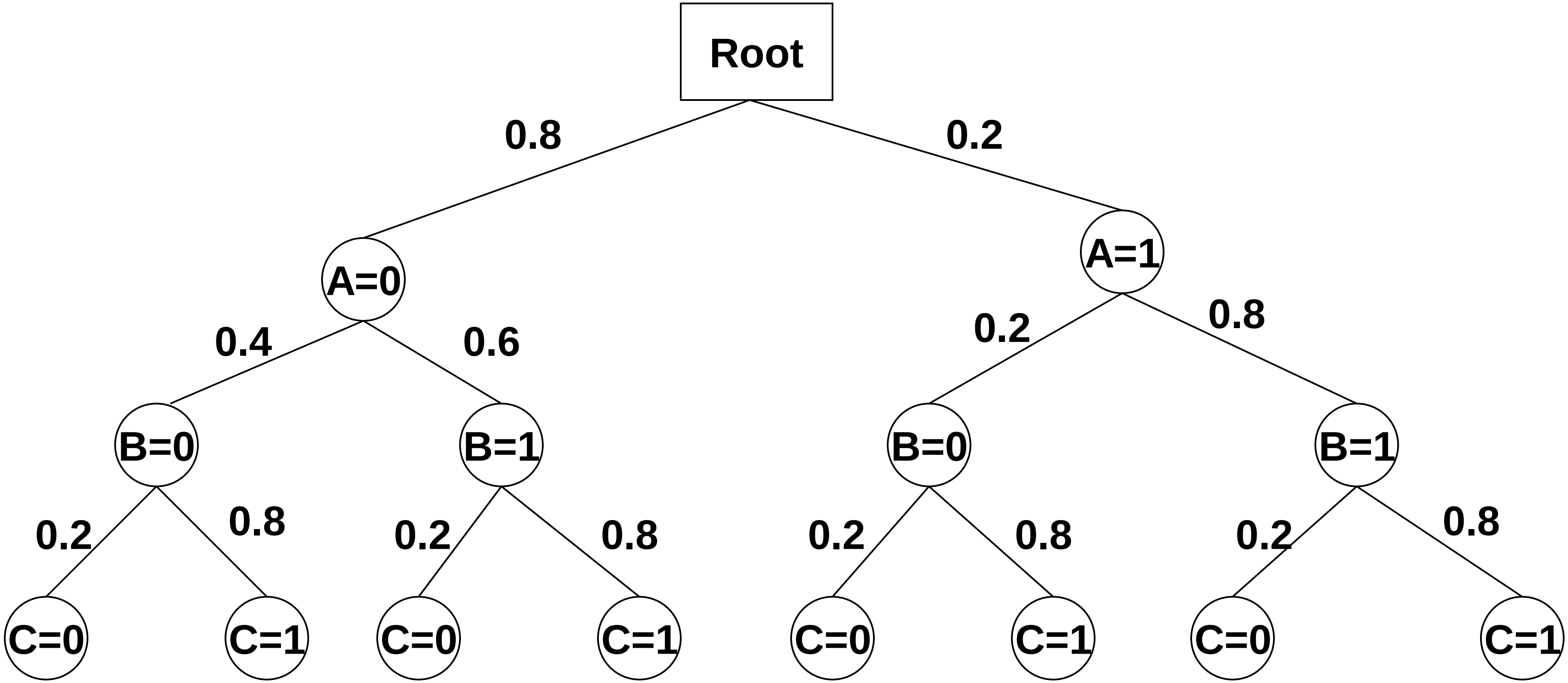
$$\hat{M}=\frac{1}{N}\sum_{j=1}^{N}\frac{f(z_j)}{Q(z_j)}=\frac{1}{N}\sum_{j=1}^{N}w(z_j)\quad \mathrm{E_Q}[\hat{\mathrm{M}}]=\mathrm{M}=\mathrm{P(e)}$$

Generating i.i.d. samples from Q

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

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

$$Q(A) = (0.8, 0.2), Q(B \mid A) = (0.4, 0.6, 0.2, 0.8), Q(C \mid A,B) = Q(C) = (0.2.0.8)$$



Rejection Problem

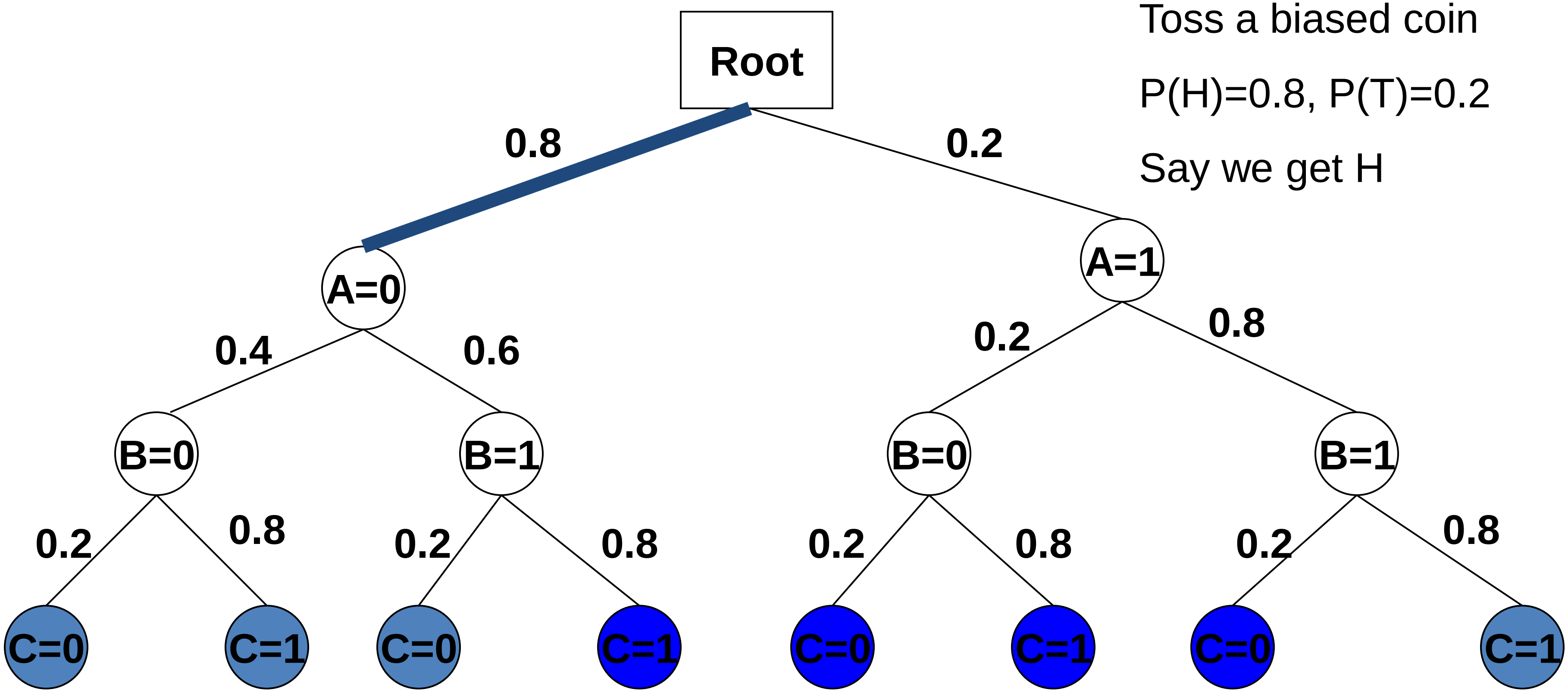
Importance sampling: $\widetilde{M} = \frac{1}{N}\sum_{i=1}^{N}\frac{f(xi)}{Q(xi)}$

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

Toss a biased coin

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

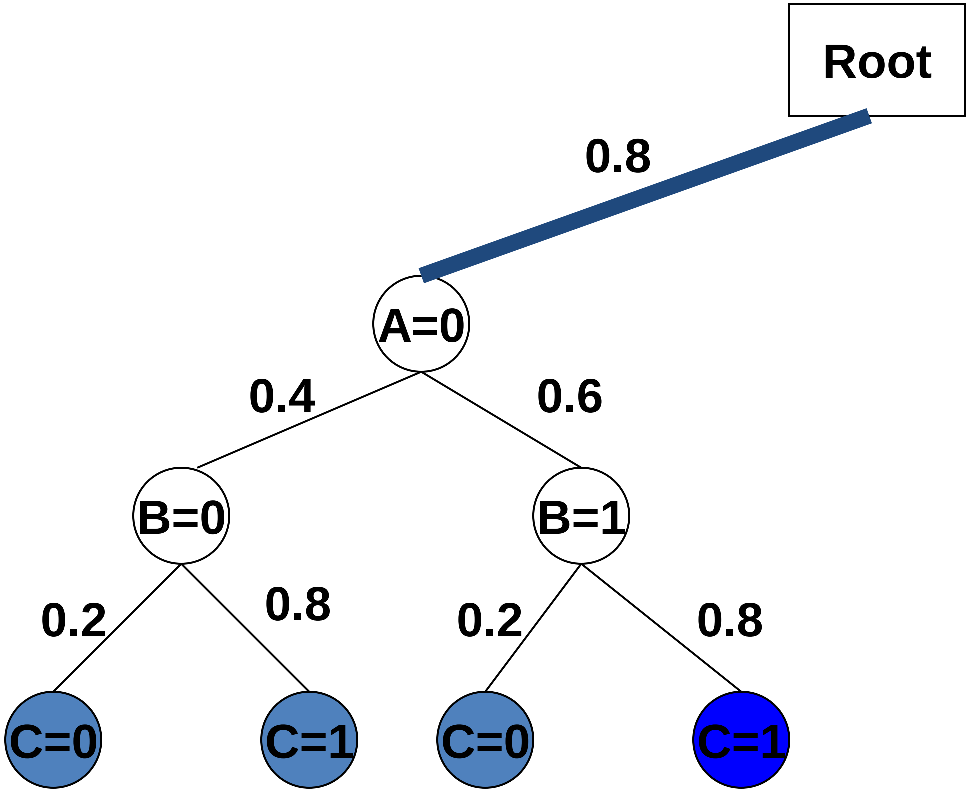
Say we get H



Rejection Problem

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

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



Toss a biased coin

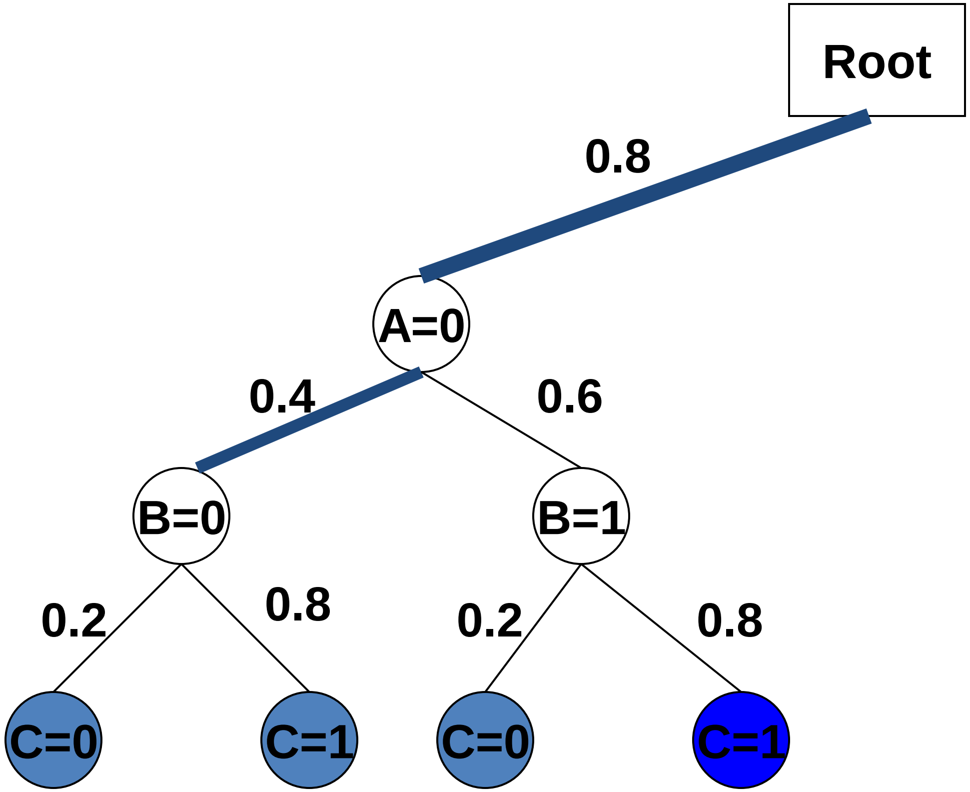
P(H)=0.4, P(T)=0.6

Say, We get a Head

Rejection Problem

$$\textbf{Importance sampling}: \widetilde{M} = \frac{1}{N}\sum_{i=1}^{N}\frac{f(xi)}{Q(xi)}$$

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



Toss a biased coin

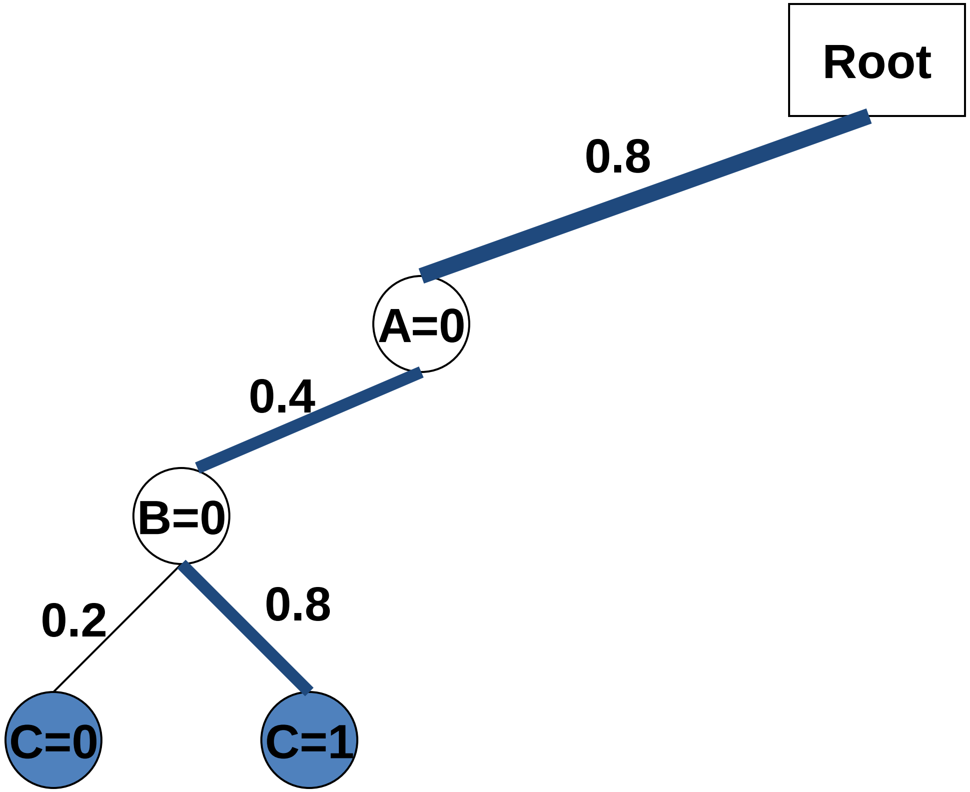
P(H)=0.4, P(T)=0.6

Say, We get a Head

Rejection Problem

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

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



Toss a biased coin

P(H)=0.4, P(T)=0.6

Say, We get a Head

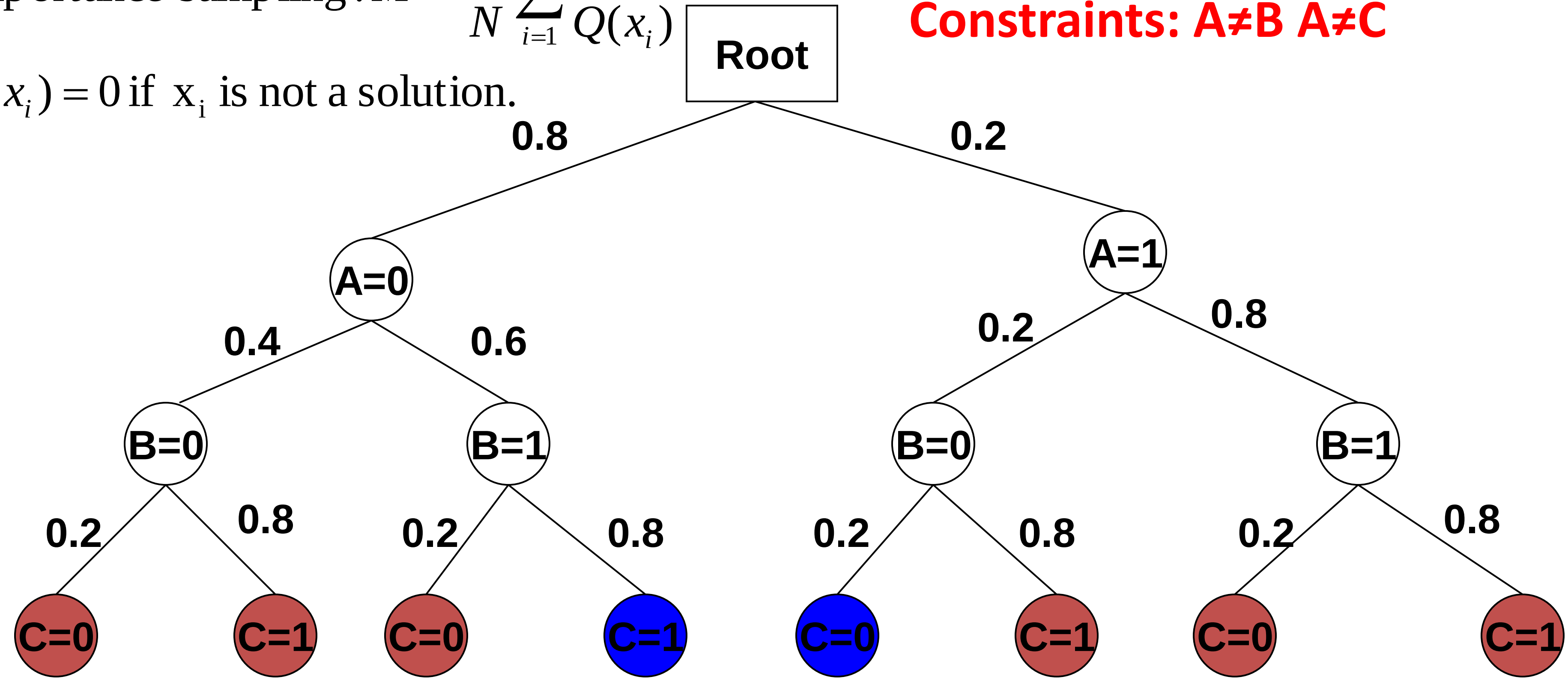
- A large number of assignments generated will be rejected, thrown away

Rejection Problem

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

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

Constraints: A≠B A≠C



All Blue leaves correspond to solutions i.e. f(x) >0

All Red leaves correspond to non-solutions i.e. f(x)=0

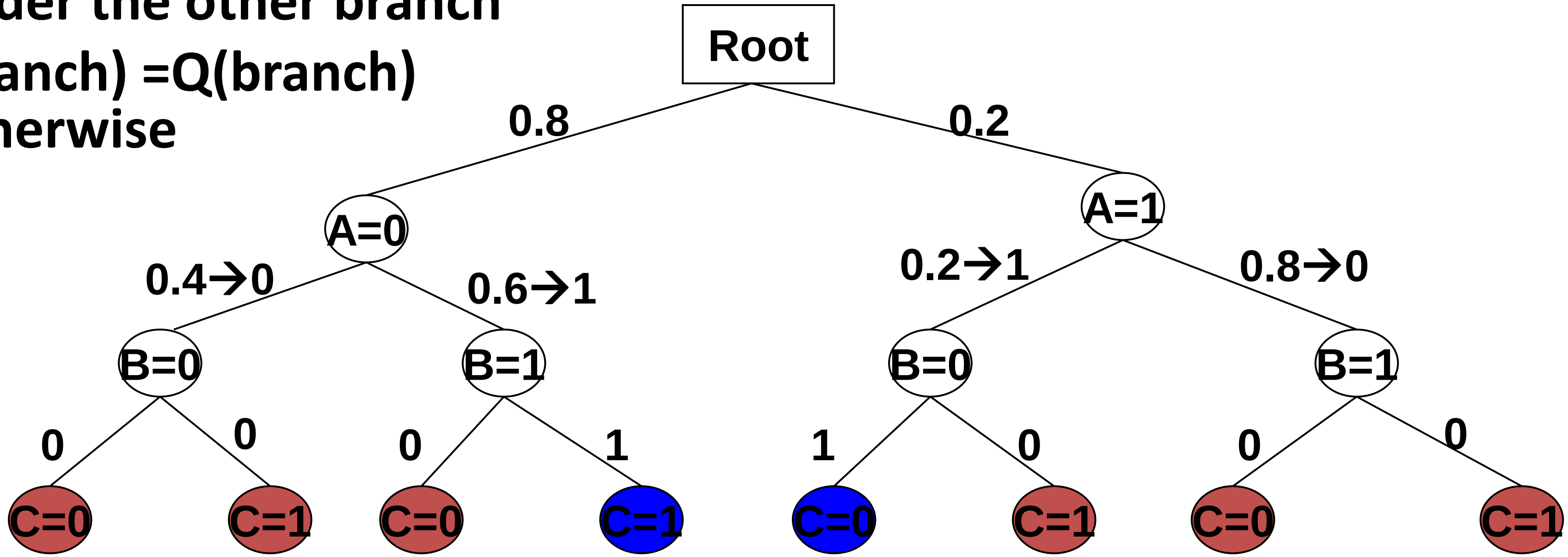
Revising Q to backtrack-free distribution:

Q^F(branch)=0 if no solutions under it

Q^F(branch)=1 if no solutions under the other branch

Q^F(branch) =Q(branch) otherwise

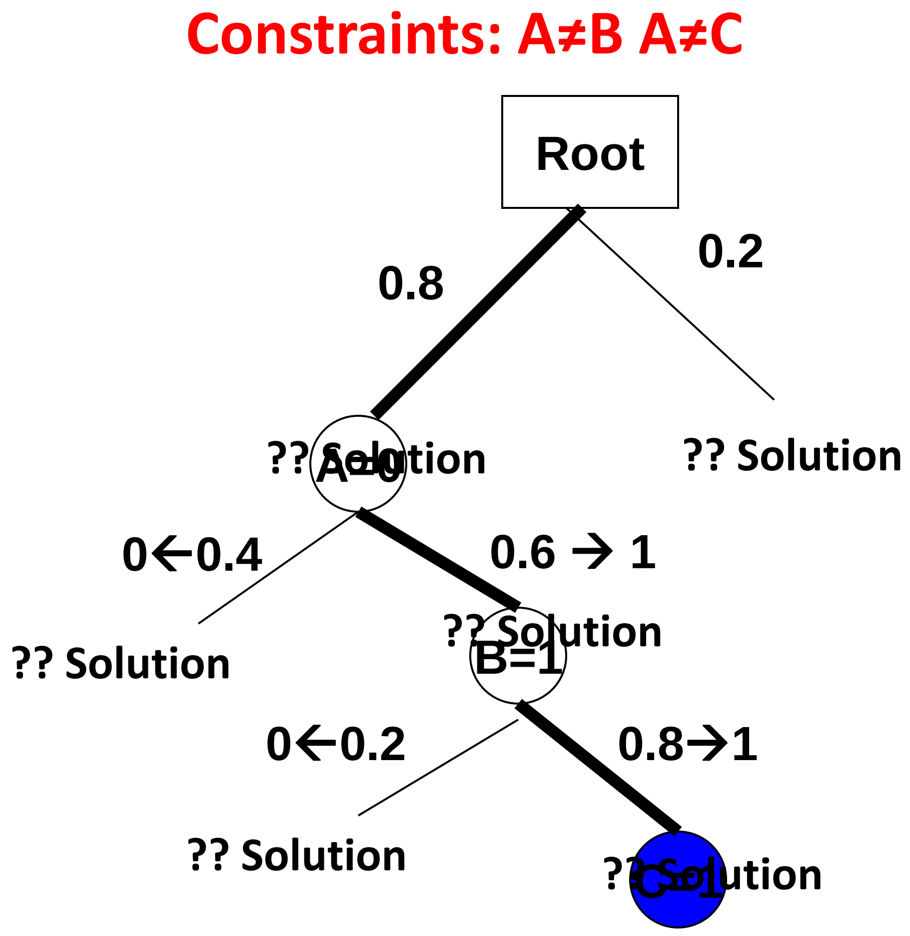
Constraints: A≠B A≠C



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



Q^F(branch)=0 if no solutions under it

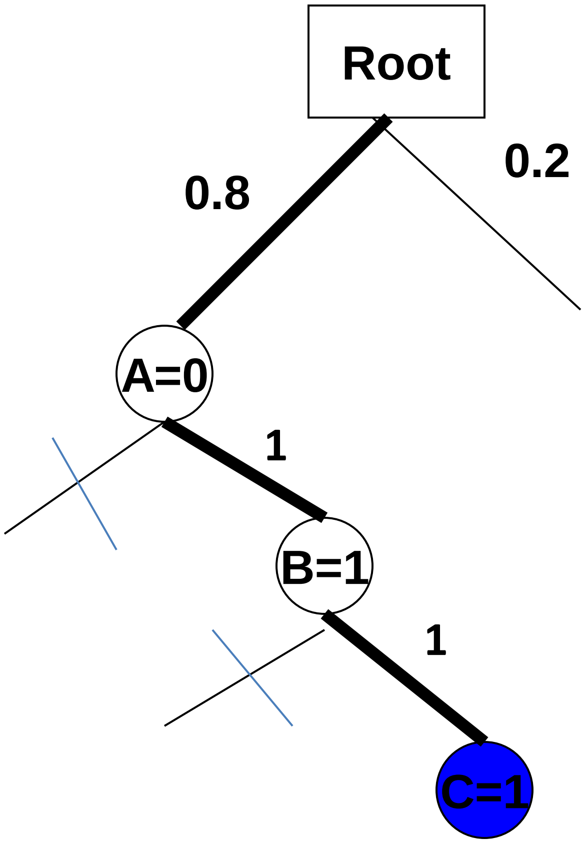
Q^F(branch)=1 if no solutions under the other branch

Q^F(branch) =Q(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

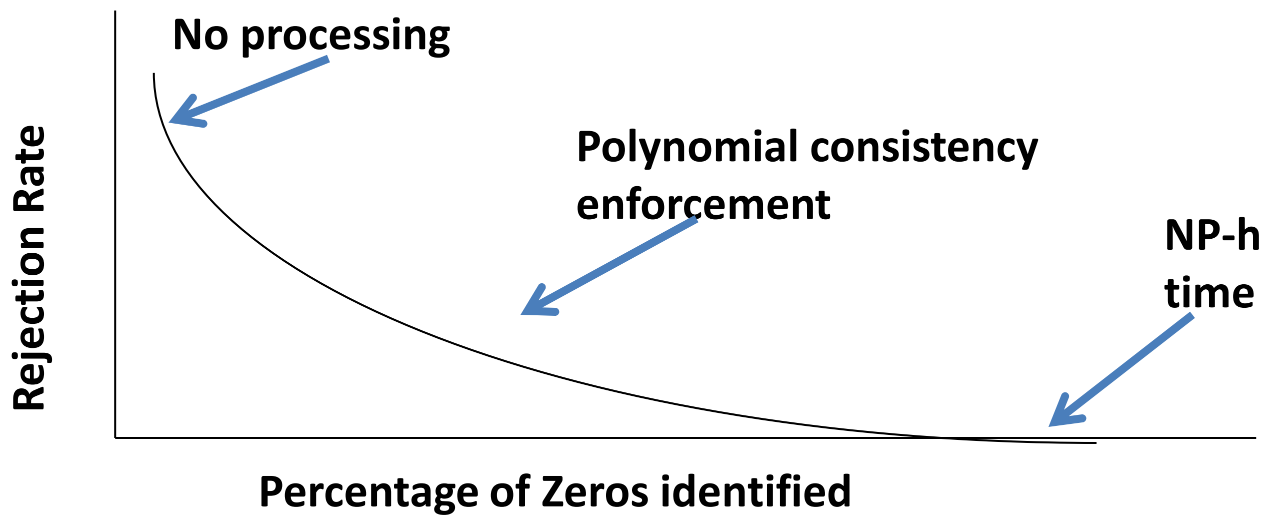
Constraints: A≠B A≠C



- Oracles: In practice
  - Adaptive consistency as pre-processing step
  - A complete CSP solver
- Too costly
  - O(exp(treewidth))
  - Invoked O(nd) for each sample

Approximations of Q^F

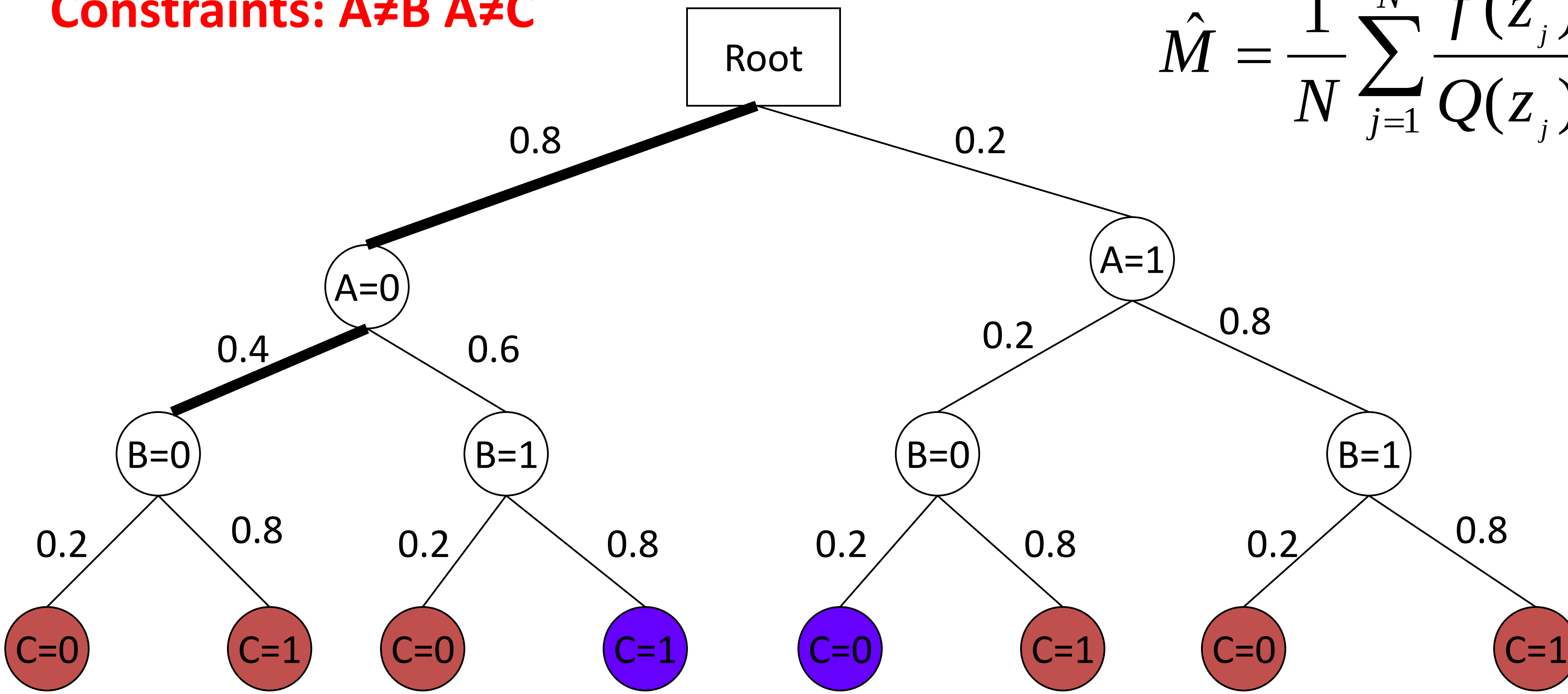
- Use i-consistency instead of adaptive consistency
  - $O(n^i)$ time and space complexity
  - identify some zeros so that they are never sampled
- Cons: Too weak when constraint portion is hard.



Algorithm SampleSearch

Constraints: A≠B A≠C

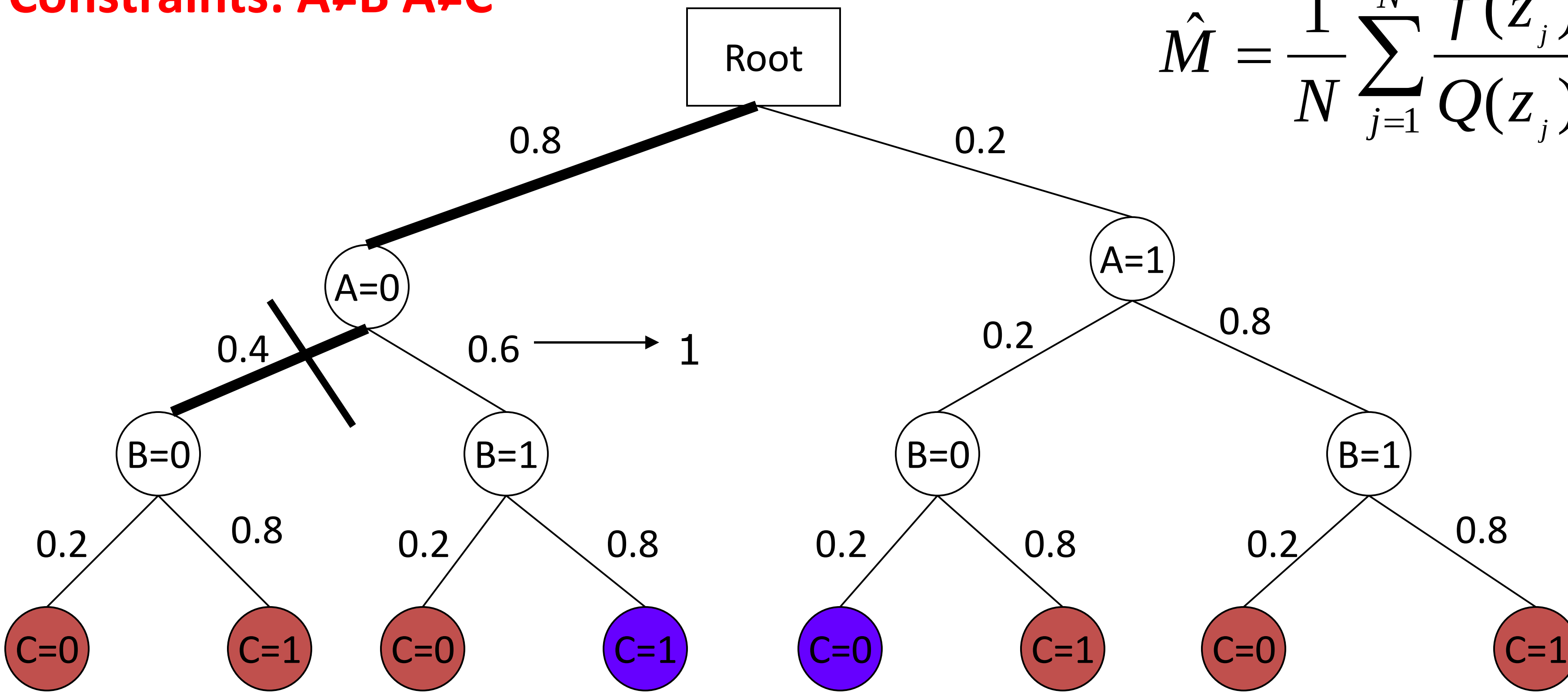
$$\hat{M} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(z_j)}{Q(z_j)}$$



Algorithm SampleSearch

Constraints: A≠B A≠C

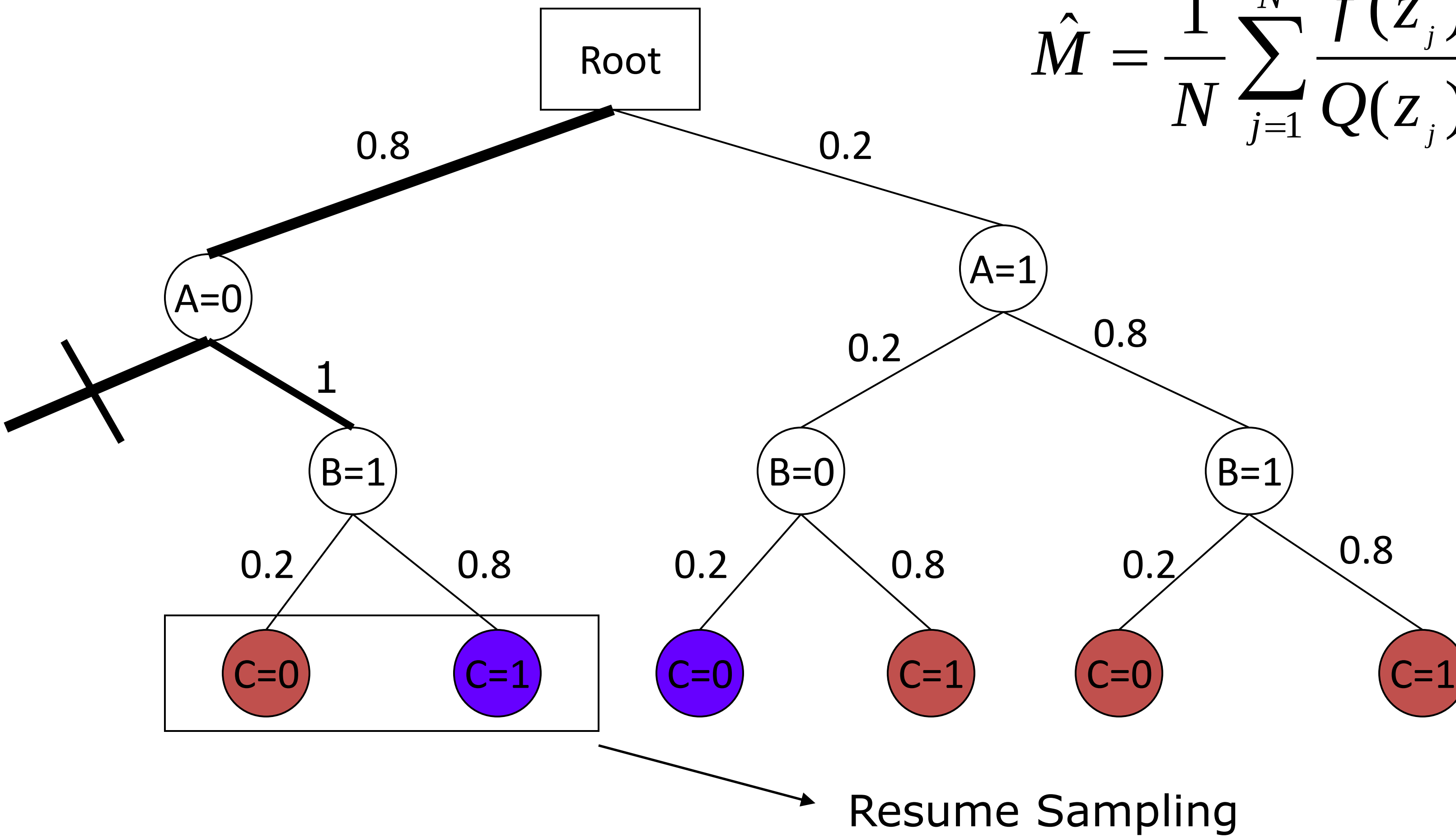
$$\hat{M} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(z_j)}{Q(z_j)}$$



Algorithm SampleSearch

Constraints: A≠B A≠C

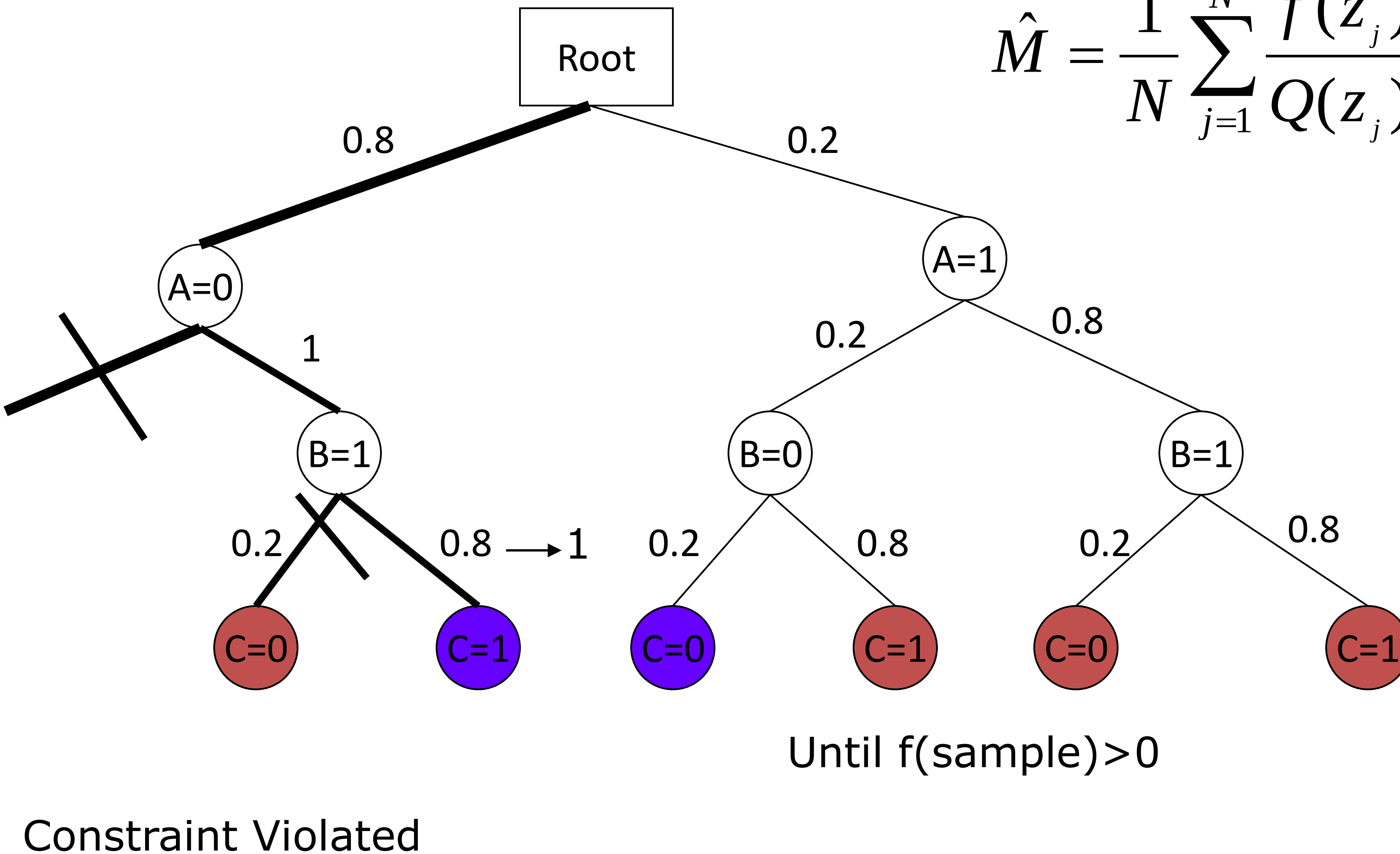
$$\hat{M} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(z_j)}{Q(z_j)}$$



Algorithm SampleSearch

Constraints: A≠B A≠C

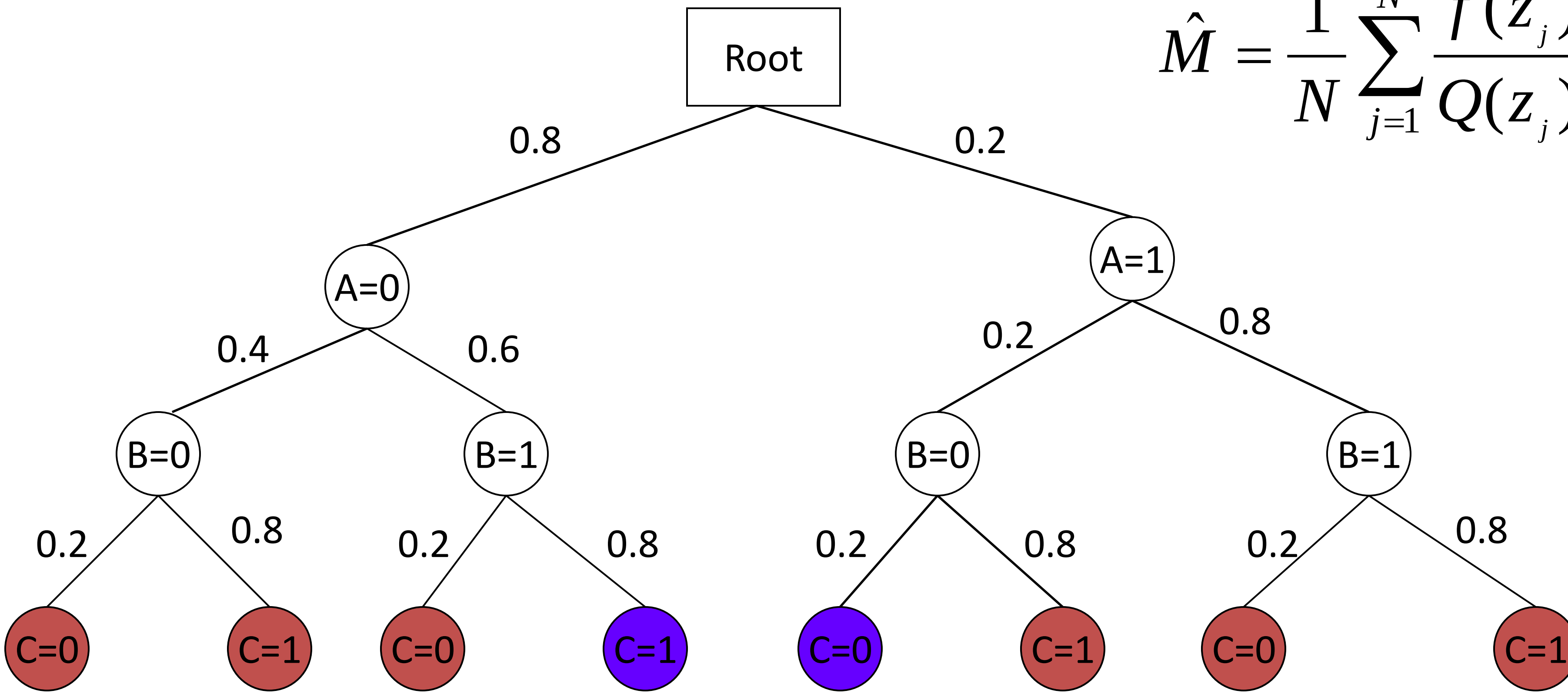
$$\hat{M} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(z_j)}{Q(z_j)}$$



Generate more Samples

Constraints: A≠B A≠C

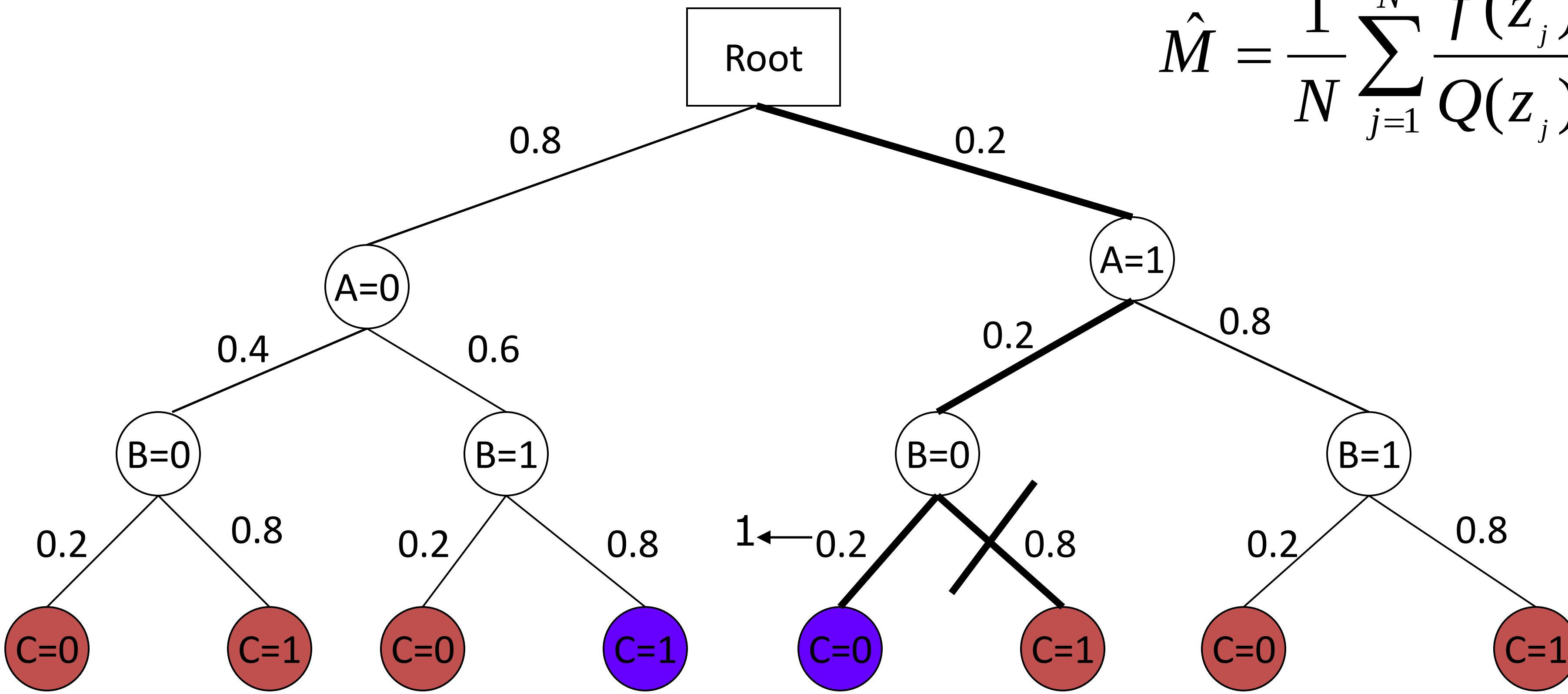
$$\hat{M} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(z_j)}{Q(z_j)}$$



Generate more Samples

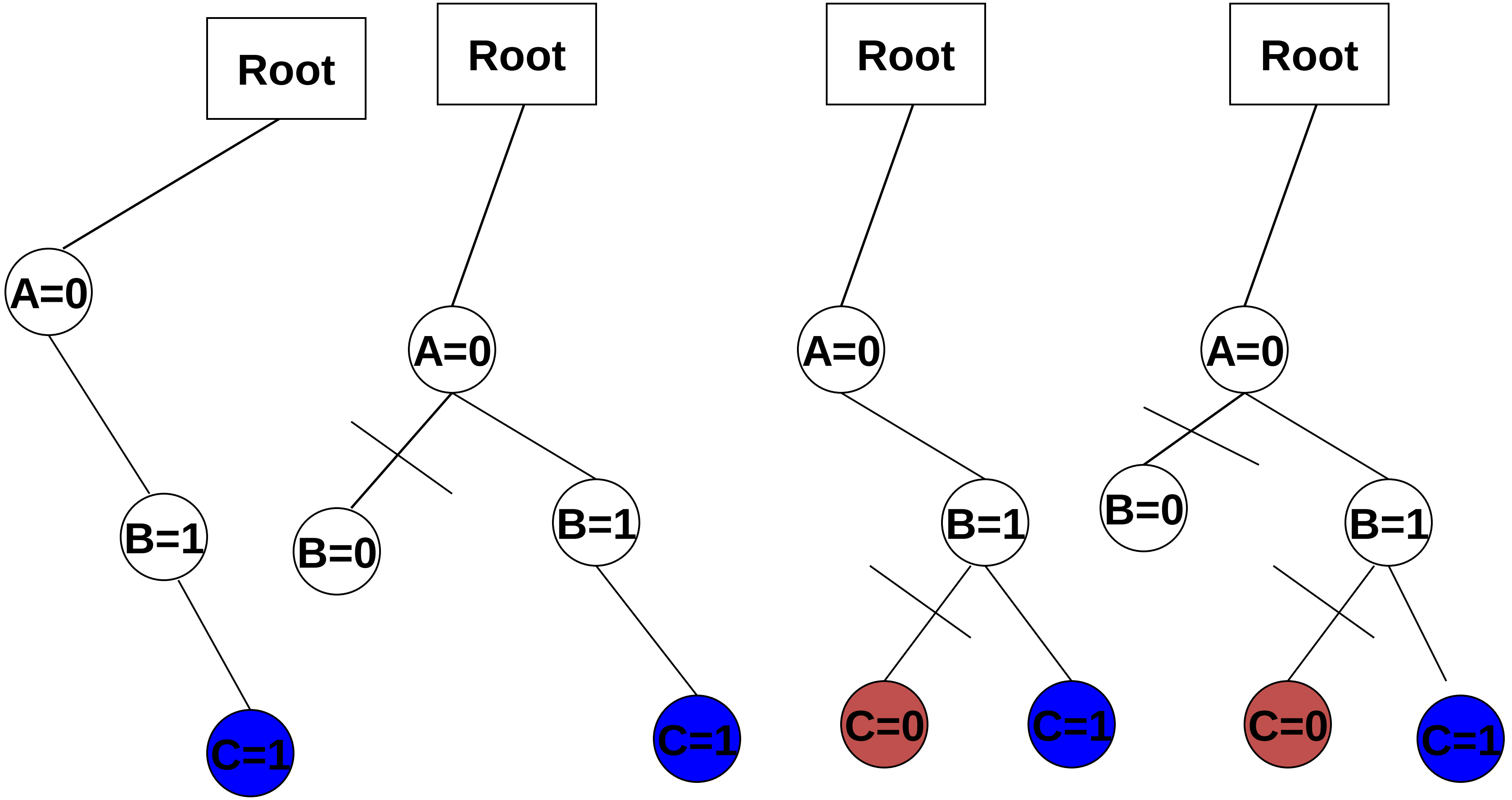
Constraints: A≠B A≠C

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



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

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

- **Theorem**: SampleSearch generates i.i.d. samples from the **backtrack-free distribution**

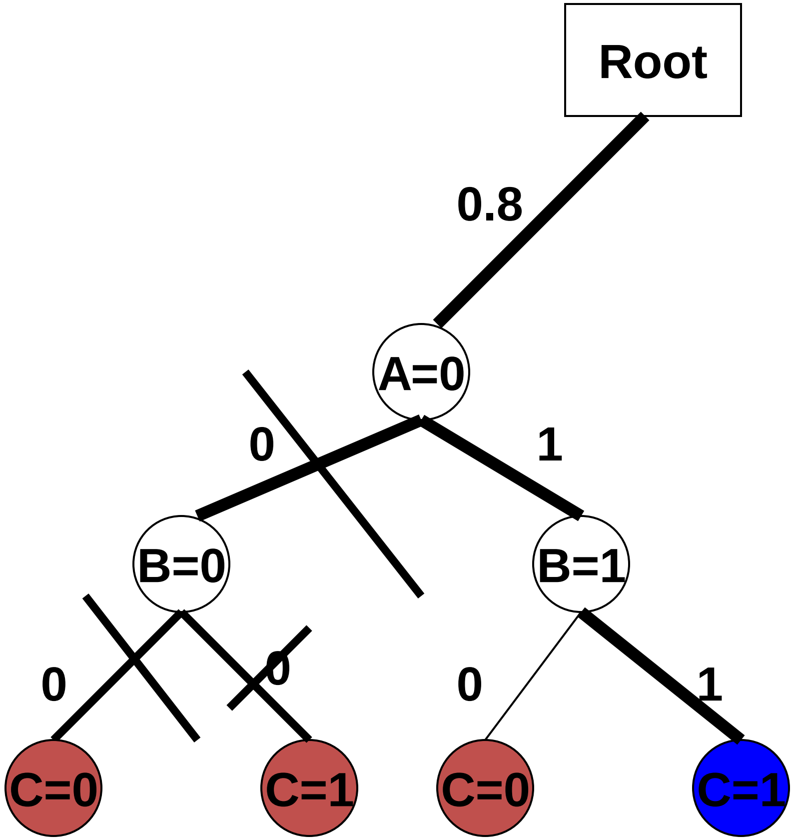
$$Q^F : \text{Backtrack - free distribution}$$

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

The Sampling distribution Q^F of SampleSearch

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

Constraints: A≠B A≠C



Backtrack-free distribution

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)?

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

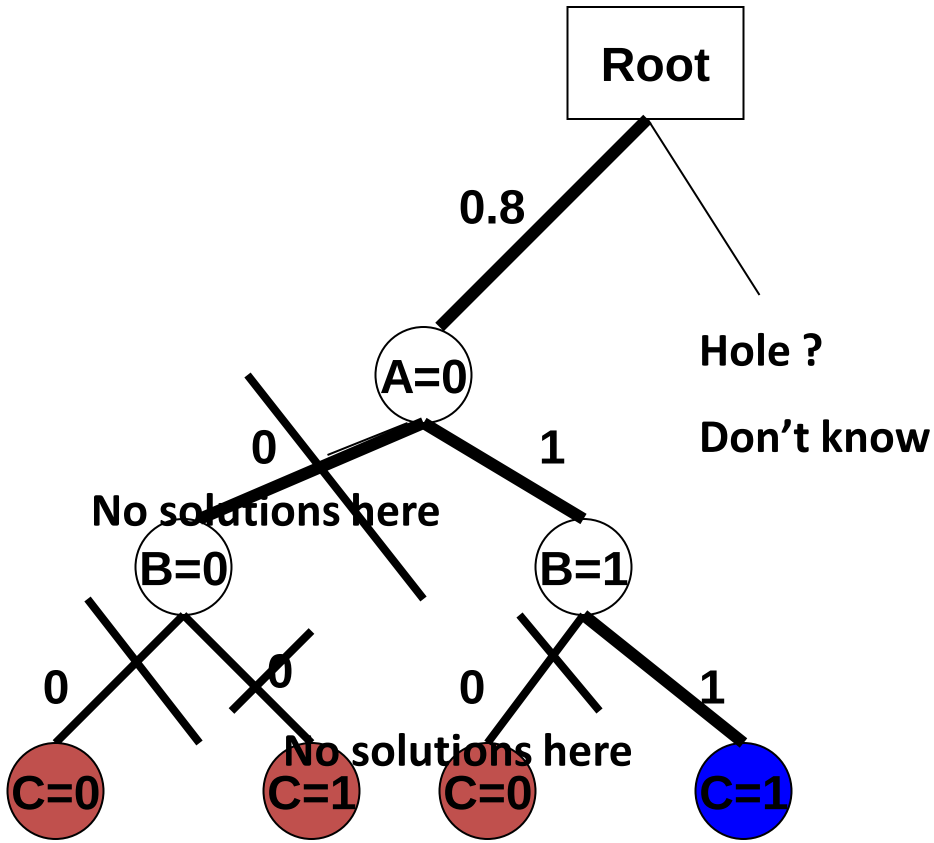
Why? SampleSearch is systematic

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

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

All samples generated by SampleSearch are solutions

Asymptotic approximations of Q^F



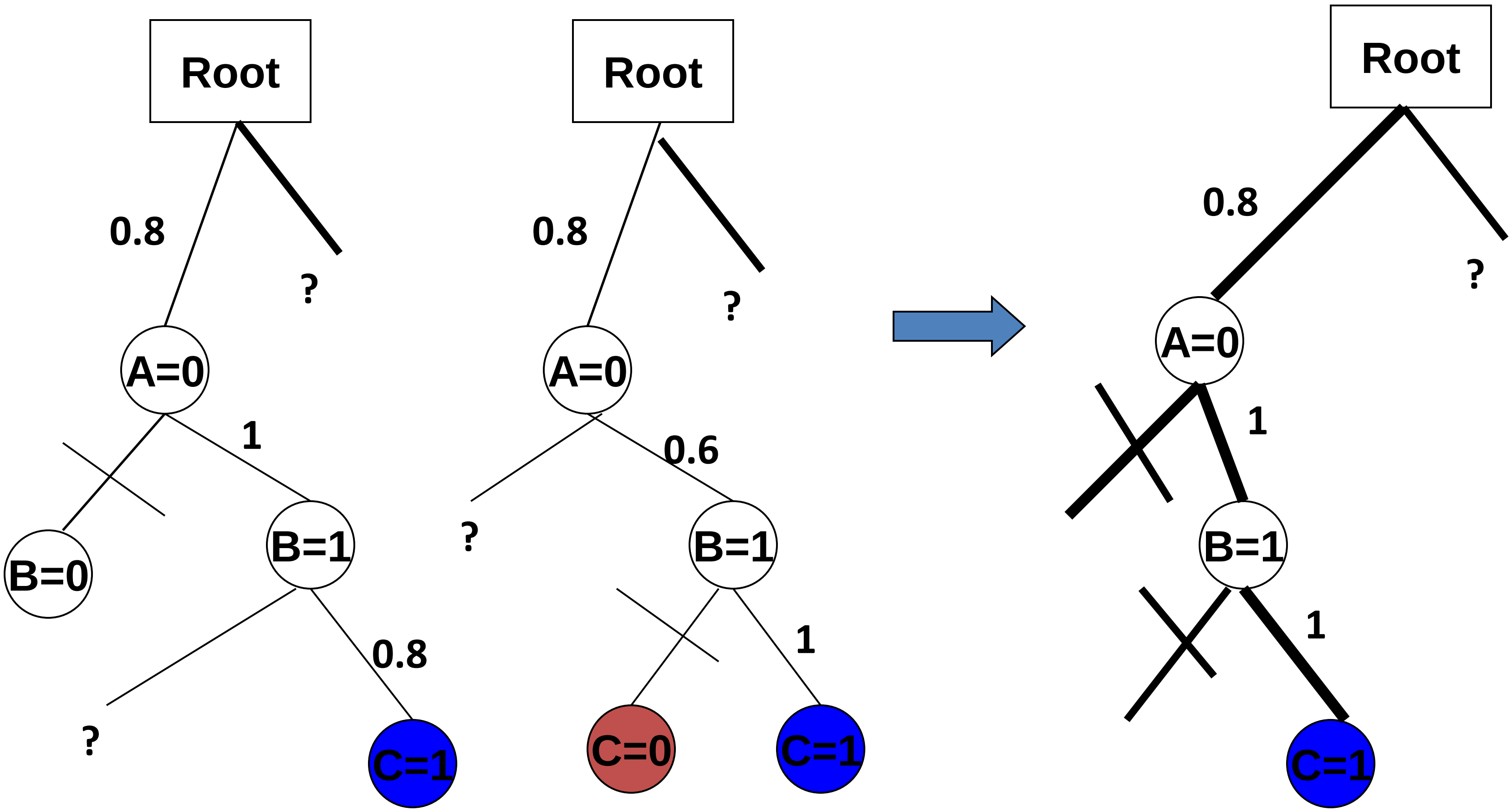
- IF Hole THEN
  - U^F=Q (i.e. there is a solution at the other branch)
  - L^F=0 (i.e. no solution at the other branch)

Q^F(branch)=0 if no solutions under it

Q^F(branch) =Q(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.**

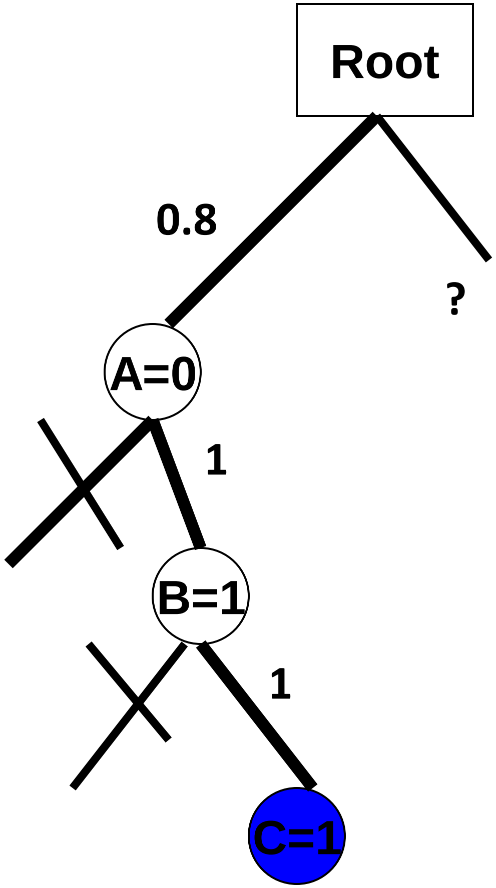
IF Hole THEN U^F_N=Q and L^F_N=0

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

$$\lim_{N \to \infty} E\left[\frac{f(x)}{U^F_N(x)}\right] = \lim_{N \to \infty} E\left[\frac{f(x)}{L^F_N(x)}\right] = M$$

Asymptotically unbiased

$$\text{Bounding}: U^F_N(x) \le Q^F(x) \le L^F_N(x)$$



Improving Naive SampleSeach: 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 Peshoueshk, 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)

Problem	$\langle n, k, c, e, w \rangle$		Exact	IJGP-wc -SS/LB	IJGP-wc -SS/UB	VEC	EDBP	IJGP-wc -IS
BN_69	$\langle 777, 7, 228, 78, 47 \rangle$	Z	5.28E-054	**3.00E-55**	**3.00E-55**	1.93E-61	2.39E-57	X
		M		6.84E+5	6.84E+5			0
BN_70	$\langle 2315, 5, 484, 159, 87 \rangle$	Z	2.00E-71	**1.21E-73**	**1.21E-73**	7.99E-82	6.00E-79	X
		M		1.92E+5	1.92E+5			0
BN_71	$\langle 1740, 6, 663, 202, 70 \rangle$	Z	5.12E-111	**1.28E-111**	**1.28E-111**	7.05E-115	1.01E-114	X
		M		7.46E+4	7.46E+4			0
BN_72	$\langle 2155, 6, 752, 252, 86 \rangle$	Z	4.21E-150	**4.73E-150**	**4.73E-150**	1.32E-153	9.21E-155	X
		M		1.53E+5	1.53E+5			0
BN_73	$\langle 2140, 5, 651, 216, 101 \rangle$	Z	2.26E-113	**2.00E-115**	**2.00E-115**	6.00E-127	2.24E-118	X
		M		7.75E+4	7.75E+4			0
BN_74	$\langle 749, 6, 223, 66, 45 \rangle$	Z	3.75E-45	**2.13E-46**	**2.13E-46**	3.30E-48	5.84E-48	X
		M		2.80E+5	2.80E+5			0
BN_75	$\langle 1820, 5, 477, 155, 92 \rangle$	Z	5.88E-91	**2.19E-91**	**2.19E-91**	5.83E-97	3.10E-96	X
		M		7.72E+4	7.72E+4			0
BN_76	$\langle 2155, 7, 605, 169, 64 \rangle$	Z	4.93E-110	**1.95E-111**	**1.95E-111**	1.00E-126	3.86E-114	X
		M		2.52E+4	2.52E+4			0

Time Bound: 3 hrs

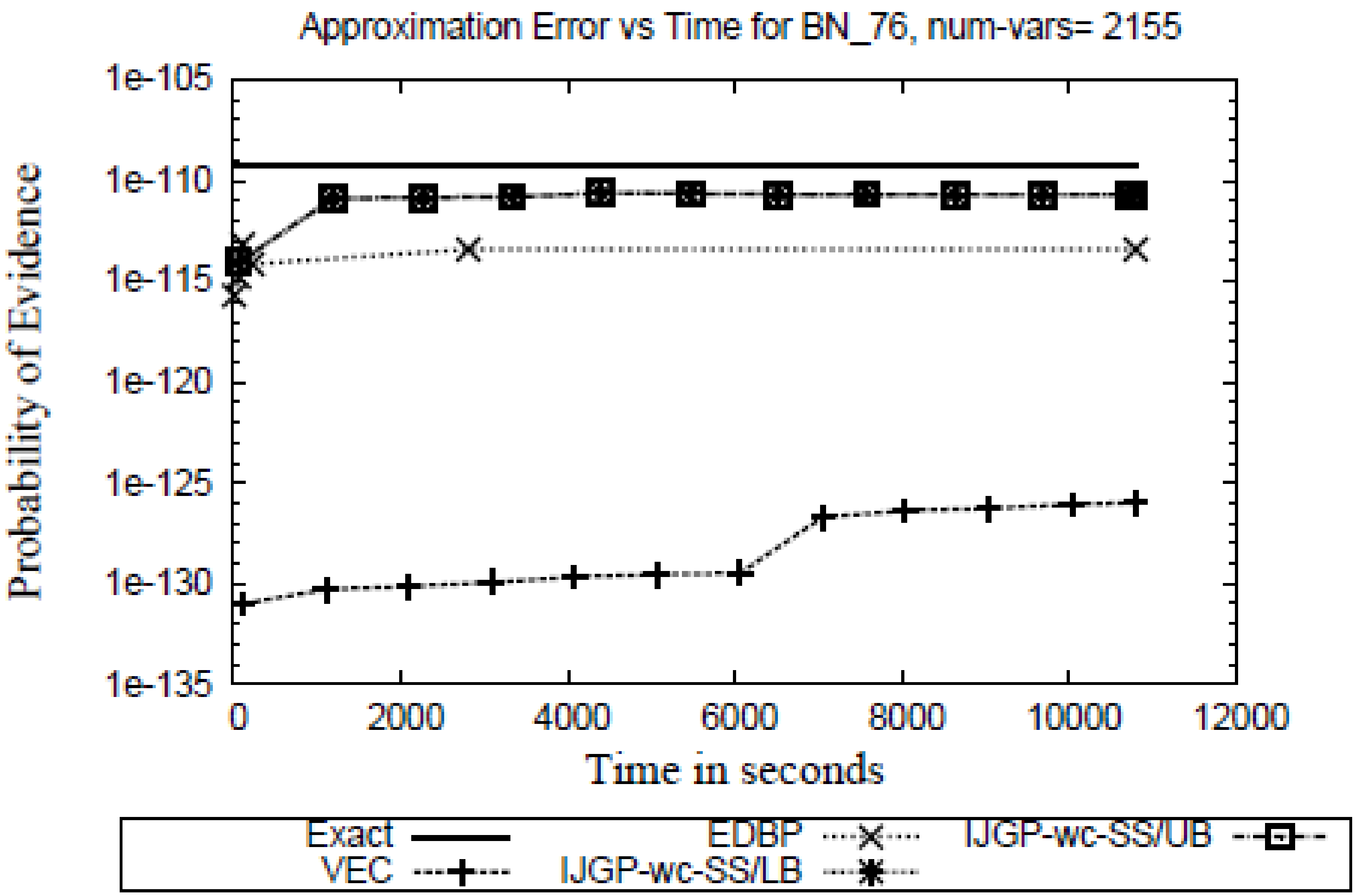
M: number of samples generated in 10 hrs

Z: Probability of Evidence

Results: Probability of Evidence Relational instances (UAI 2008 evaluation)

Problem	$\langle n, k, c, e, w \rangle$		Exact	IJGP-wc -SS/LB	IJGP-wc -SS/UB	VEC	EDBP	IJGP-wc -IS
Friends and Smokers								
fs-04	$\langle 262, 2, 74, 226, 12 \rangle$	Z	1.52E-05	8.11E-06	8.11E-06	**1.53E-05**$^{*(1s)}$	X	1.52E-05
		M		1.00E+6	1.00E+6			2.17E+8
fs-07	$\langle 1225, 2, 371, 1120, 35 \rangle$	Z	9.80E-17	2.23E-16	2.23E-16	**1.78E-15**$^{*(708s)}$	X	X
		M		1.00E+6	1.00E+6			0
fs-10	$\langle 3385, 2, 1055, 3175, 71 \rangle$	Z	7.88E-31	**2.49E-32**	**2.49E-32**	X	X	X
		M		8.51E+5	8.51E+5			0
fs-13	$\langle 7228, 2, 2288, 6877, 119 \rangle$	Z	1.33E-51	3.26E-55	3.26E-55	X	X	**1.33E-51**
		M		5.41E+5	5.41E+5			4.67E+7
fs-16	$\langle 13240, 2, 4232, 12712, 171 \rangle$	Z	8.63E-78	6.04E-79	6.04E-79	X	X	**8.63E-78**
		M		1.79E+5	1.79E+5			1.37E+7
fs-19	$\langle 21907, 2, 7049, 21166, 243 \rangle$	Z	2.12E-109	**1.62E-114**	**1.62E-114**	X	X	X
		M		1.90E+5	1.90E+5			0
fs-22	$\langle 33715, 2, 10901, 32725, 335 \rangle$	Z	2.00E-146	**4.88E-147**	**4.88E-147**	X	X	X
		M		1.18E+5	1.18E+5			0
fs-25	$\langle 49150, 2, 15950, 47875, 431 \rangle$	Z	7.18E-189	**2.67E-189**	**2.67E-189**	X	X	X
		M		9.23E+4	9.23E+4			0
fs-28	$\langle 68698, 2, 22358, 67102, 527 \rangle$	Z	9.82E-237	**4.53E-237**	4.53E-237	X	X	X
		M		9.35E+4	9.35E+4			0
fs-29	$\langle 76212, 2, 24824, 74501, 559 \rangle$	Z	6.81E-254	**9.44E-255**	9.44E-255	X	X	X
		M		2.62E+4	2.62E+4			0

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
1e-105
1e-110
1e-115
1e-120
1e-125
1e-130
1e-135
0
2000
4000
6000
8000
10000
12000
Time in seconds
Exact
EDBP
IJGP-wc-SS/UB
VEC
IJGP-wc-SS/LB

Results: Solution Counts
Latin Square instances (size 8 to 16)

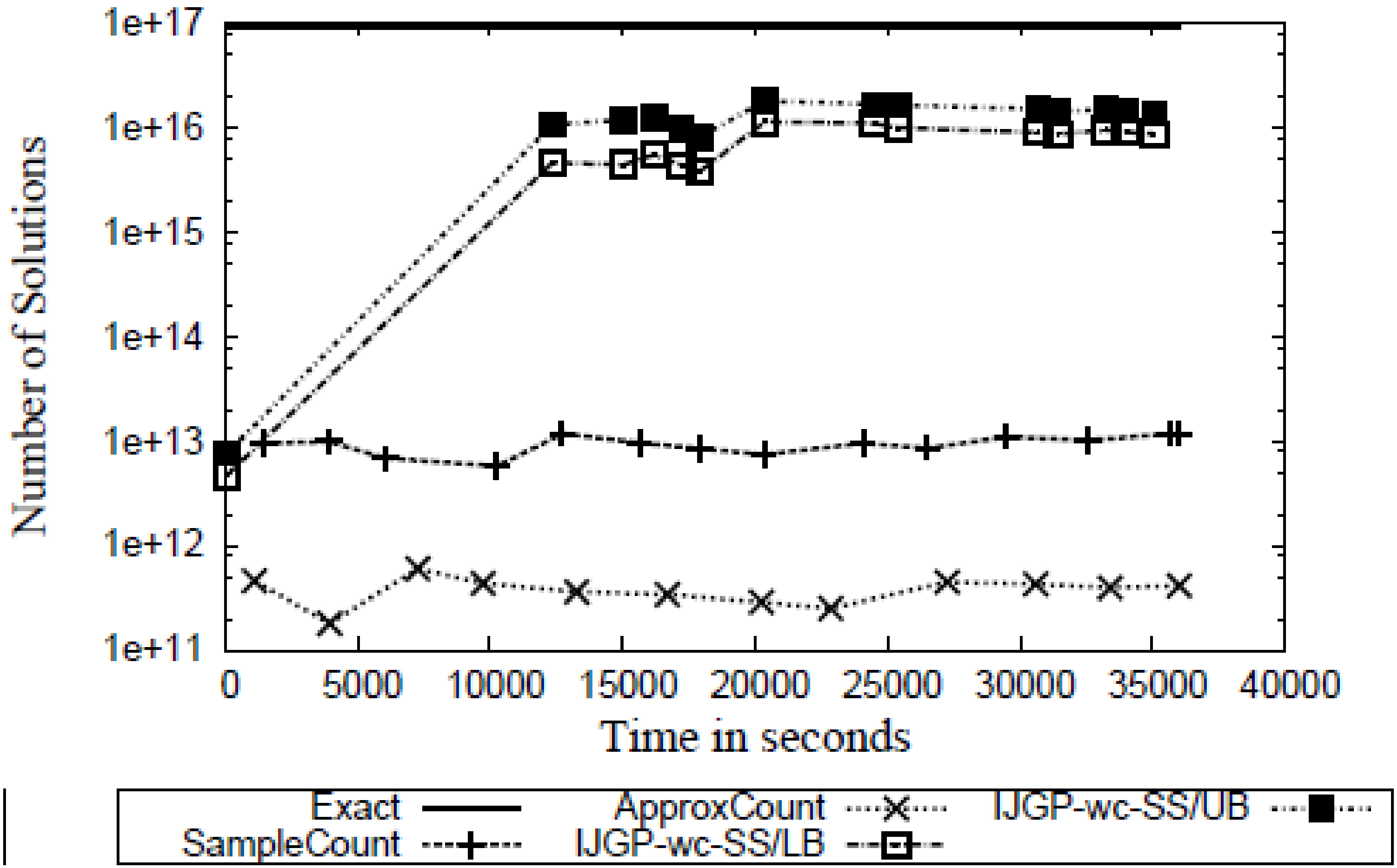
Problem	$\langle n,k,c,w \rangle$		Exact	Sample Count	Approx Count	REL SAT	IJGP-wc-SS/LB	IJGP-wc-SS/UB	IJGP-wc-IS
ls8-norm	$\langle 512, 2, 5584, 255 \rangle$	Z	5.40E11	**5.15E+11**	3.52E+11	2.44E+08	5.91E+11	5.91E+11	X
		M		16514	17740		236510	236510	0
ls9-norm	$\langle 729, 2, 9009, 363 \rangle$	Z	3.80E17	4.49E+17	1.26E+17	1.78E+08	**3.44E+17**	3.44E+17	X
		M		7762	8475		138572	138572	0
ls10-norm	$\langle 1000, 2, 13820, 676 \rangle$	Z	7.60E24	**7.28E+24**	1.17E+24	1.36E+08	6.74E+24	6.74E+24	X
		M		3854	4313		95567	95567	0
ls11-norm	$\langle 1331, 2, 20350, 956 \rangle$	Z	5.40E33	2.08E+34	4.91E+31	1.09E+08	**3.87E+33**	3.87E+33	X
		M		2002	2289		66795	66795	0

Time Bound: 10 hrs
M: number of samples generated in 10 hrs
Z: Solution Counts

Results: Solution Counts Langford instances

Problem	$\langle n, k, c, w \rangle$		Ex-act	Sample Count	Approx Count	REL SAT	IJGP-wc-SS/LB	IJGP-wc-SS/UB	IJGP-wc-IS
lang12	$\langle 576, 2, 13584, 383 \rangle$	Z	2.16E+5	1.93E+05	2.95E+04	**2.16E+05**$^{*(297s)}$	2.16E+05	2.16E+05	X
		M		2720	4668		999991	999991	0
lang16	$\langle 1024, 2, 32320, 639 \rangle$	Z	6.53E+08	5.97E+08	8.22E+06	6.28E+06	**6.51E+08**	6.99E+08	X
		M		328	641		14971	14971	0
lang19	$\langle 1444, 2, 54226, 927 \rangle$	Z	5.13E+11	9.73E+10	6.87E+08	8.52E+05	**6.38E+11**	7.31E+11	X
		M		146	232		3431	3431	0
lang20	$\langle 1600, 2, 63280, 1023 \rangle$	Z	5.27E+12	1.13E+11	3.99E+09	8.55E+04	2.83E+12	**3.45E+12**	X
		M		120	180		2961	2961	0
lang23	$\langle 2116, 2, 96370, 1407 \rangle$	Z	7.60E+15	7.53E+14	3.70E+12	X	4.17E+15	**4.19E+15**	X
		M		38	54		1111	1111	0
lang24	$\langle 2304, 2, 109536, 1535 \rangle$	Z	9.37E+16	1.17E+13	4.15E+11	X	8.74E+15	**1.40E+16**	X
		M		25	33		271	271	0

Time Bound: 10 hrs
M: number of samples generated in 10 hrs
Z: Solution Counts



Number of Solutions
1e+17
1e+16
1e+15
1e+14
1e+13
1e+12
1e+11
0
5000
10000
15000
20000
25000
30000
35000
40000
Time in seconds
Exact
ApproxCount
IJGP-wc-SS/UB
SampleCount
IJGP-wc-SS/LB

Results on Marginals

- Evaluation Criteria

$$Exact : P(x_i) \quad Approximate : A(x_i)$$

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

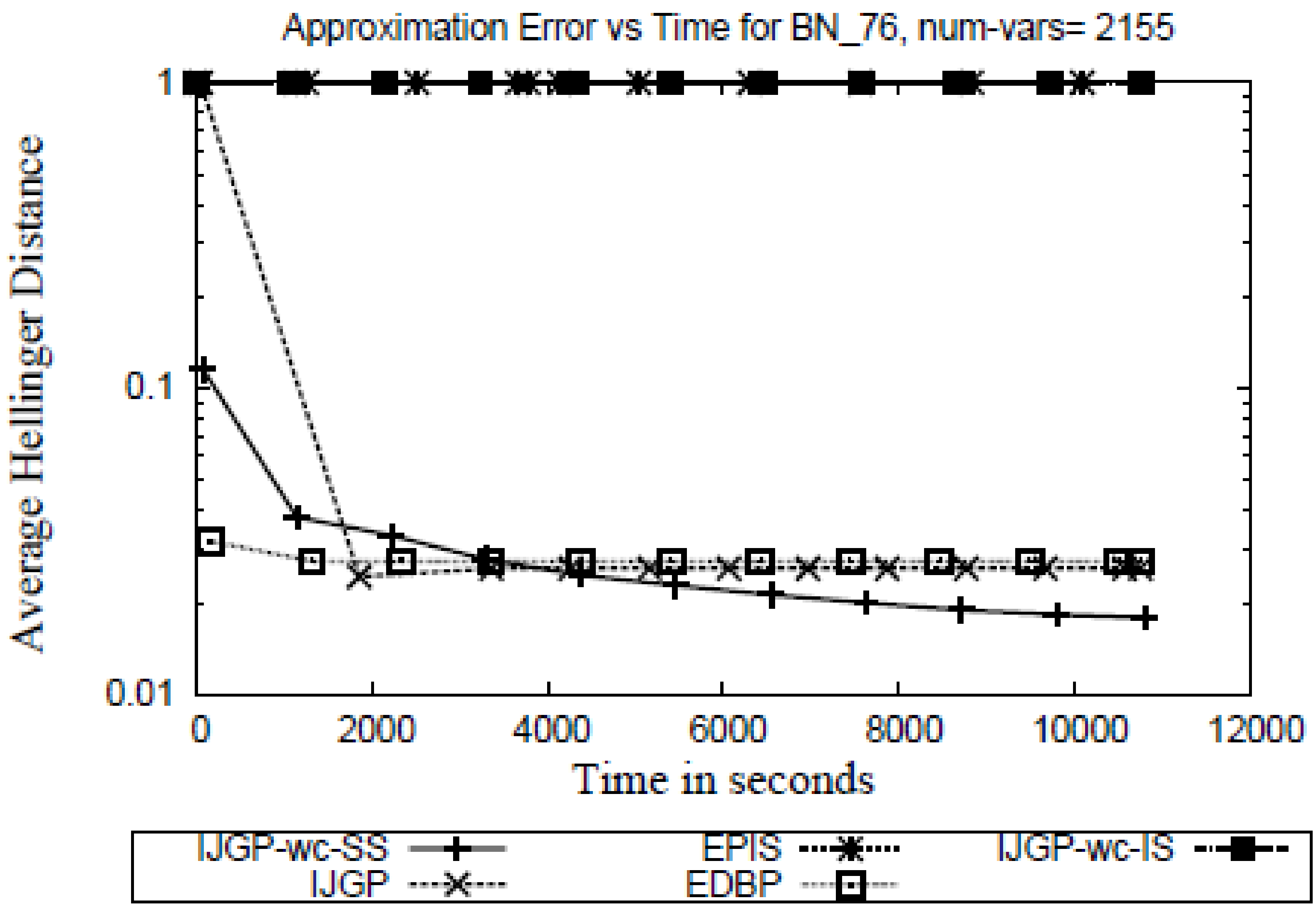
- Always bounded between 0 and 1
- Lower Bounds the KL distance
- When probabilities close to zero are present KL distance may tend to infinity.

Results: Posterior Marginals Linkage instances (UAI 2006 evaluation)

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

Time Bound: 3 hrs

Table shows the Hellinger distance (Δ) and Number of samples: M


Approximation Error vs Time for BN_76, num-vars= 2155
Average Hellinger Distance
1
0.1
0.01
0
2000
4000
6000
8000
10000
12000
Time in seconds
IJGP-wc-SS
EPIS
IJGP-wc-IS
IJGP
EDBP

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

$$M = \sum_{z} f(z) = \sum_{x1,x2,x3,x4} f_A(x_1, x_2, x_3) f_B(x_2, x_3, x_4) f_C(x_4)$$

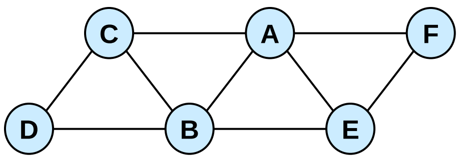
$$M = \sum_{z} f(z) = \sum_{x1,x2,x3,x4} f_A(x_1) f_B(x_2) f_C(x_3, x_4)$$

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

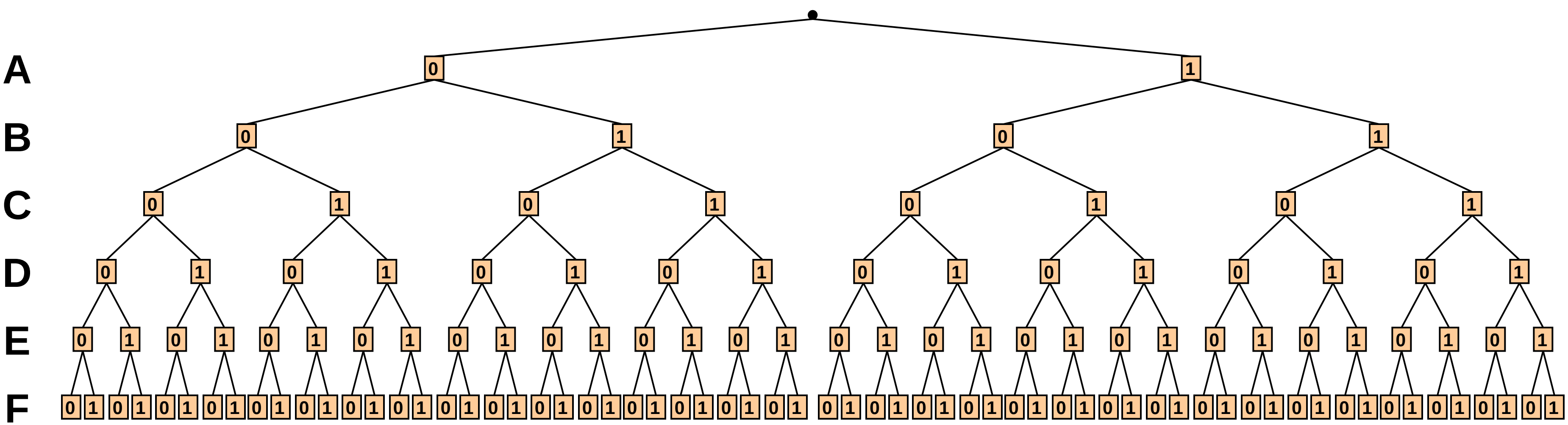


A	B	C	R_{ABC}
0	0	0	1
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	0

B	C	D	R_{BCD}
0	0	0	1
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	1

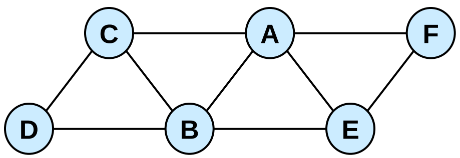
A	B	E	R_{ABE}
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	0

A	E	F	R_{AEF}
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	0



AND/OR Search Tree

Constraint Satisfaction – Counting Solutions

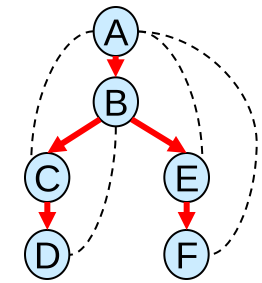


A	B	C	R_{ABC}
0	0	0	**1**
0	0	1	**1**
0	1	0	**0**
0	1	1	**1**
1	0	0	**1**
1	0	1	**1**
1	1	0	**1**
1	1	1	**0**

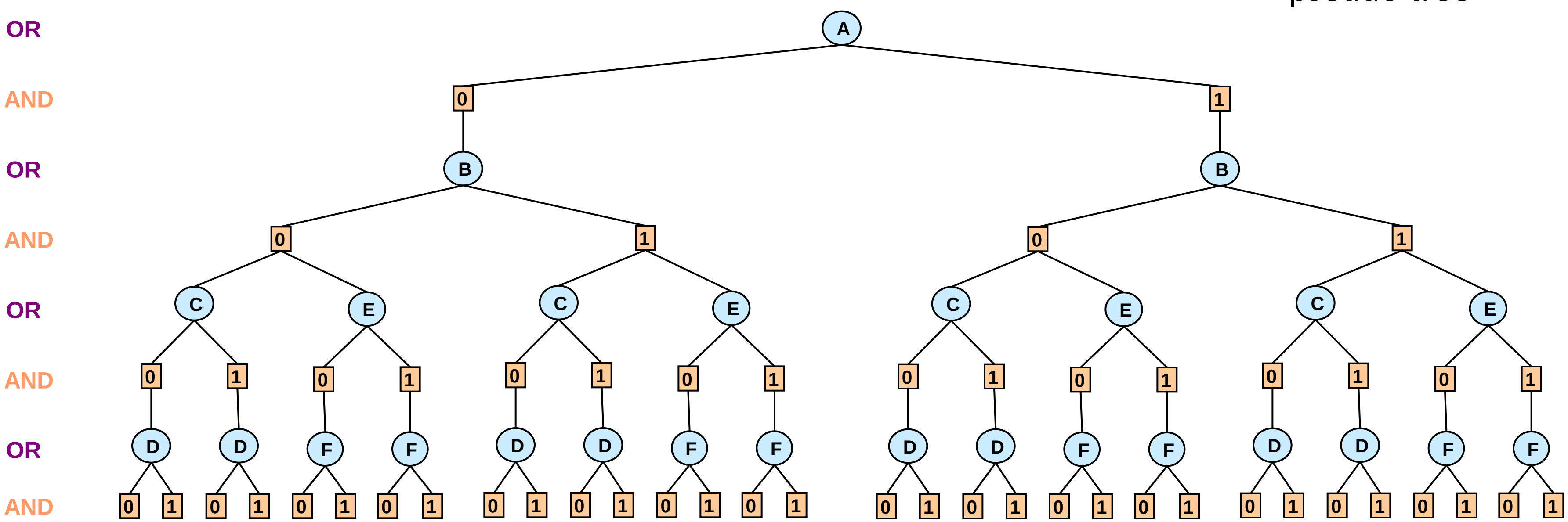
B	C	D	R_{BCD}
0	0	0	**1**
0	0	1	**1**
0	1	0	**1**
0	1	1	**0**
1	0	0	**1**
1	0	1	**0**
1	1	0	**1**
1	1	1	**1**

A	B	E	R_{ABE}
0	0	0	**1**
0	0	1	**0**
0	1	0	**1**
0	1	1	**1**
1	0	0	**0**
1	0	1	**1**
1	1	0	**1**
1	1	1	**0**

A	E	F	R_{AEF}
0	0	0	**0**
0	0	1	**1**
0	1	0	**1**
0	1	1	**1**
1	0	0	**1**
1	0	1	**1**
1	1	0	**1**
1	1	1	**0**



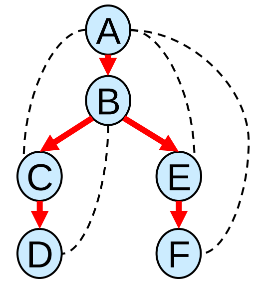
pseudo tree



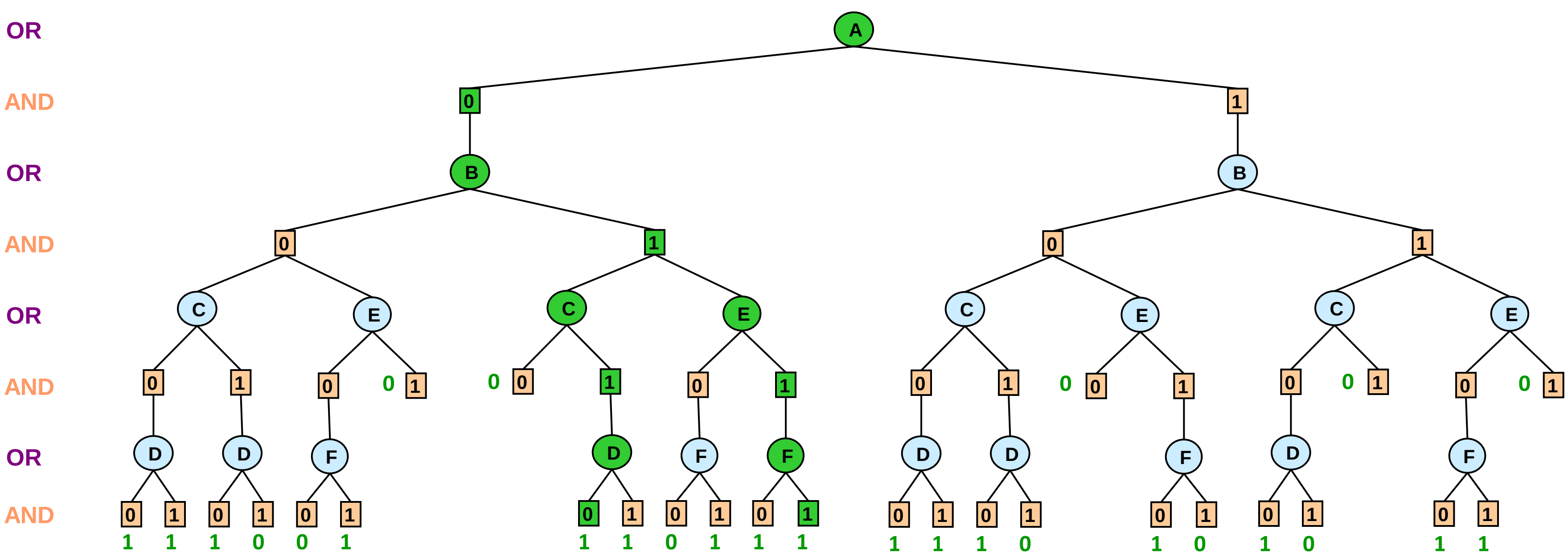
AND/OR Tree

OR – casing upon variable values

AND – decomposition into independent subproblems



pseudo tree

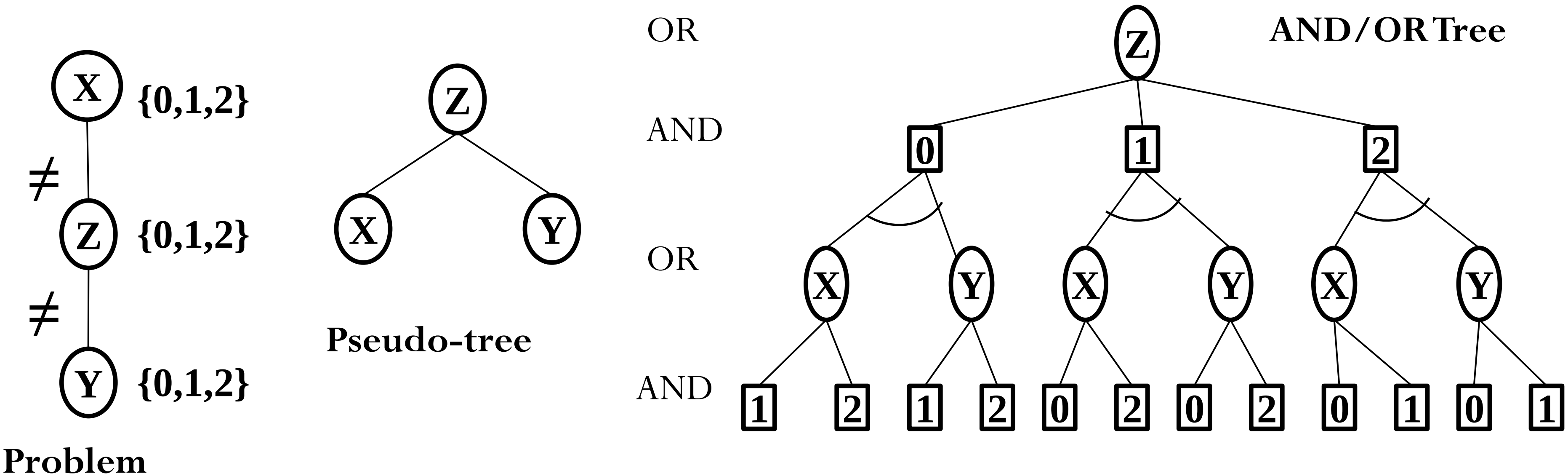


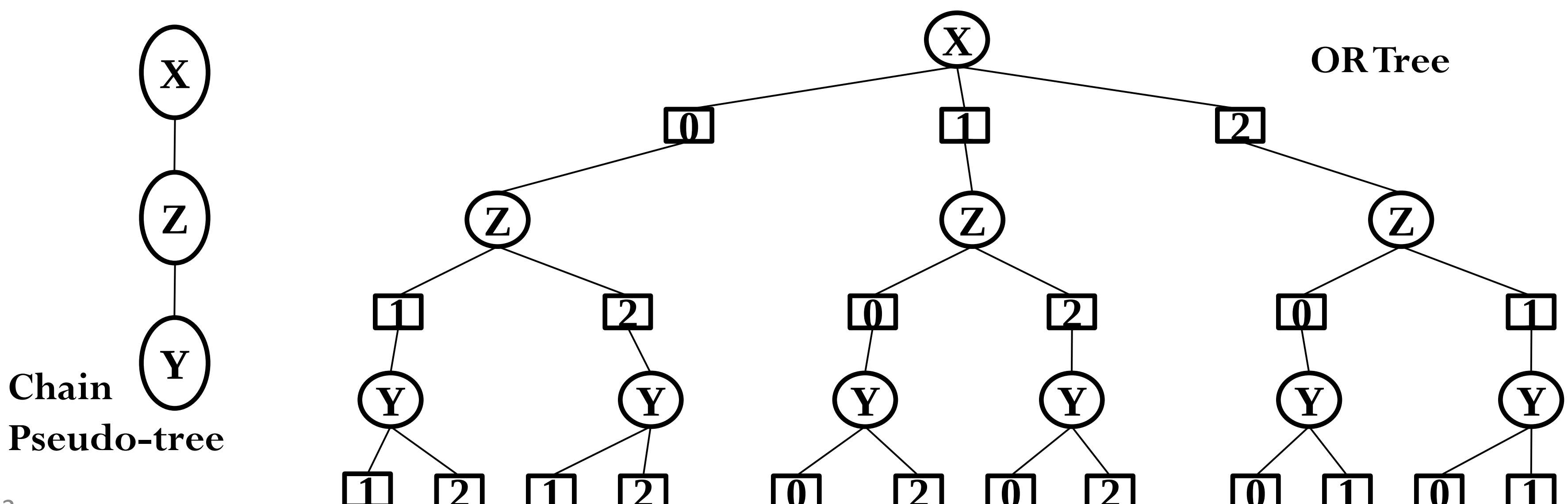
Complexity of AND/OR Tree Search

	AND/OR tree	OR tree
Space	$O(n)$	$O(n)$
Time	$O(n\ k^m)$ $O(n\ k^{w^* \log n})$ [Freuder & Quinn85], [Collin, Dechter & Katz91], [Bayardo & Miranker95], [Darwiche01]	$O(k^n)$

k = domain size
m = depth of pseudo-tree
n = number of variables
w*= treewidth

Background: AND/OR search space



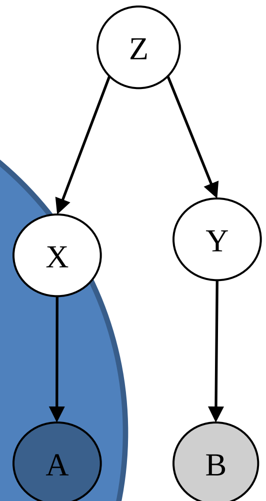


Example Bayesian network

P(Z)	Z=0	Z=1
	0.8	0.2

f(Z)

P(X|Z)	X=0	X=1	X=2
Z=0	0.3	0.4	0.3
Z=1	0.2	0.7	0.1



P(Y|Z)	Y=0	Y=1	Y=2
Z=0	0.5	0.1	0.4
Z=1	0.2	0.6	0.2

f(XZ)

f(YZ)

P(X)	
X=0	0.1
X=1	0.2
X=2	0.6

Evidence A=0, B=0

P(Y)	
Y=0	0.2
Y=1	0.7
Y=2	0.1

$$M = P(a,b) = \sum_{XYZ} f(XZ)\, f(YZ)\, f(Z)$$

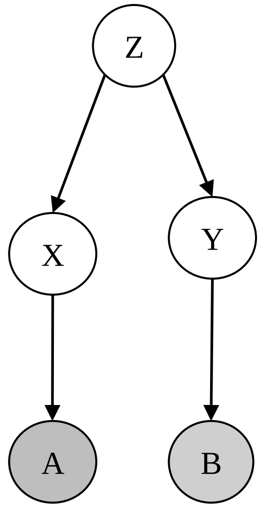
Recap:
Conventional Importance Sampling

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

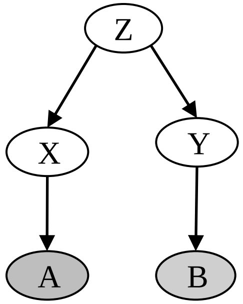
$$Q(XYZ) = Q(Z)Q(X \mid Z)Q(Y \mid Z)$$

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

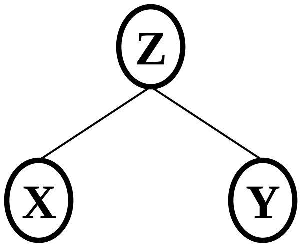
$$= E_Q\left[\frac{f(XZ) f(YZ) f(Z)}{Q(Z)Q(X \mid Z)Q(Y \mid Z)}\right]$$



AND/OR idea!
Decompose this expectation



AND/OR Importance Sampling (General Idea)



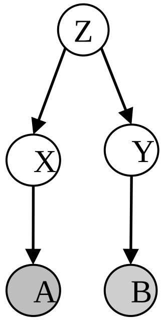
Pseudo-tree

- Decompose Expectation

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

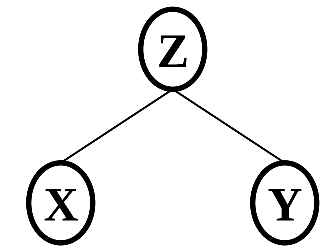
$$M = \left(\sum_{Z} \frac{f(Z)}{Q(Z)} Q(Z) \left(\sum_{X} \frac{f(XZ)}{Q(X\mid Z)} Q(X\mid Z) \right) \left(\sum_{Y} \frac{f(YZ)}{Q(Y\mid Z)} Q(Y\mid Z) \right) \right)$$

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



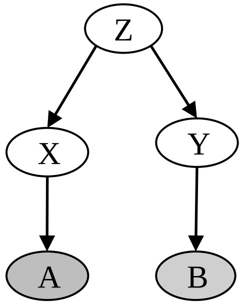
AND/OR Importance Sampling (General Idea)

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



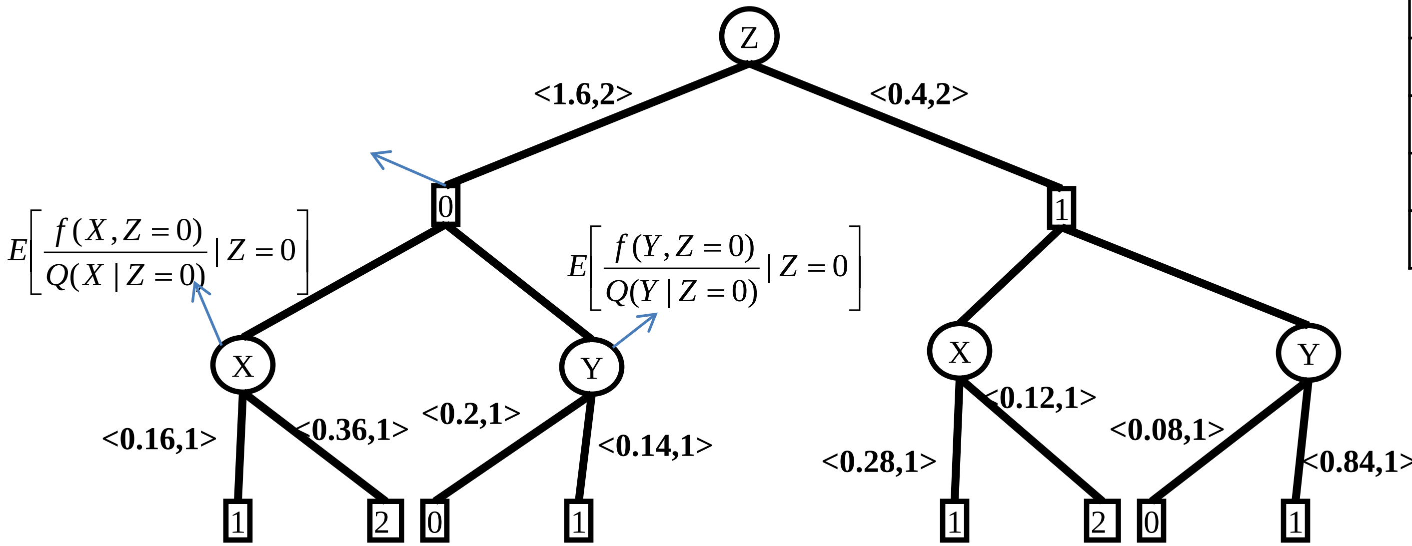
Pseudo-tree

- 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

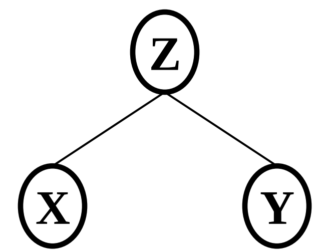


AND/OR Importance Sampling

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



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



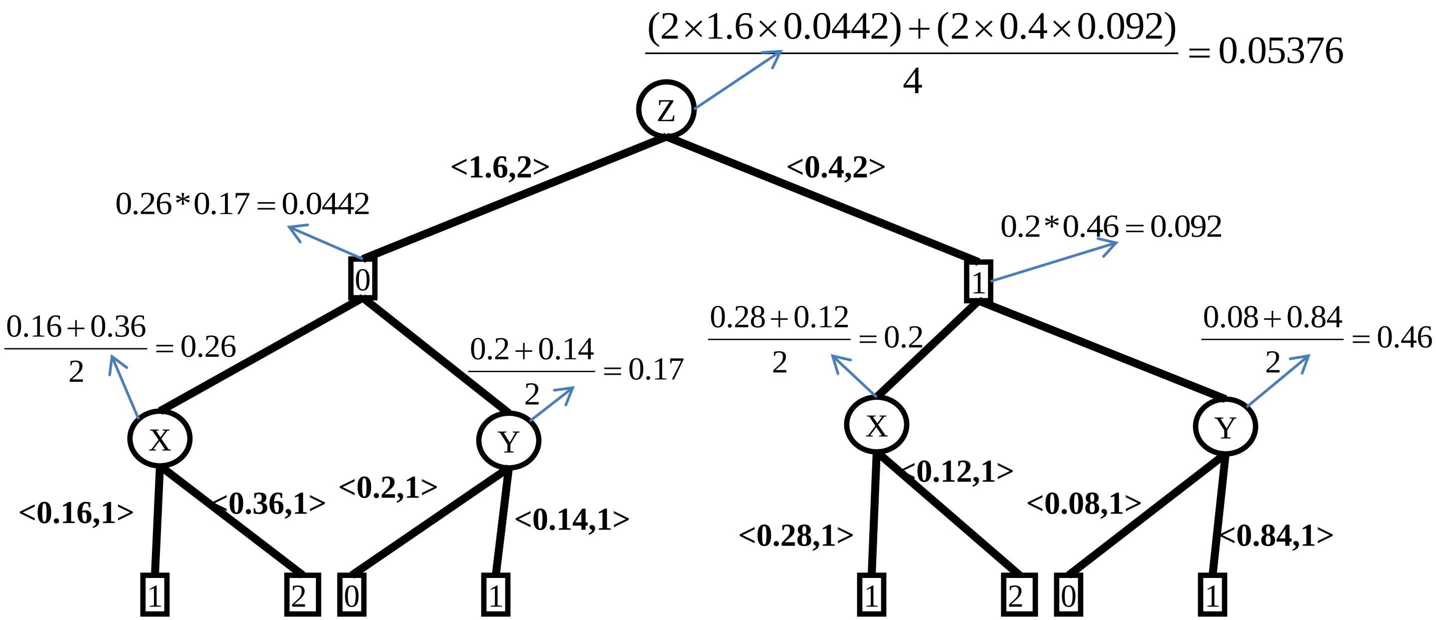
Pseudo-tree

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

$$= \text{Average Weight of samples on X having } Z=0$$

$$= \text{w}(x=1, z=0) + \text{w}(x=2, z=0)/2$$

AND/OR Importance Sampling



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

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 $\mathbf{x}_1, \ldots, \mathbf{x}_N$ *from Q along O.*
2. Build a AND/OR sample tree f*or 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(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:

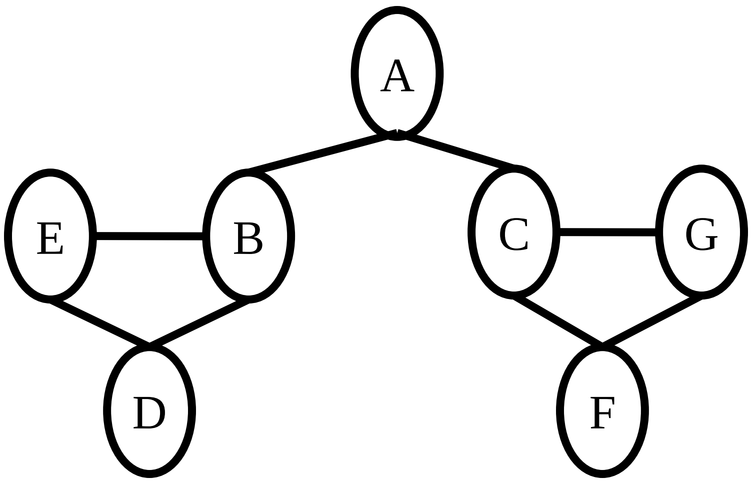
$$\hat{M}_{RB} = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{R} P(R|k_i)}{Q(k_i)}$$

Weighted Counts conditioned on K=k_i

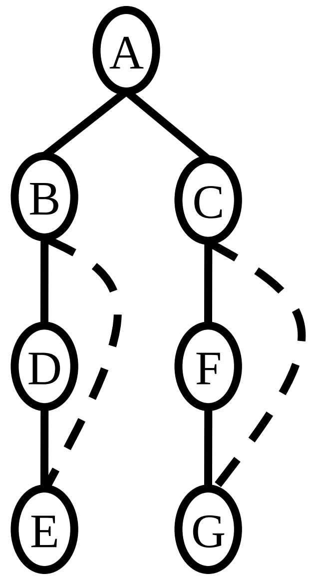
  - **w-cutset sampling (Bidyuk and Dechter, 2003)**: Select K such that the treewidth of R after removing K is bounded by "w".

AND/OR w-cutset sampling

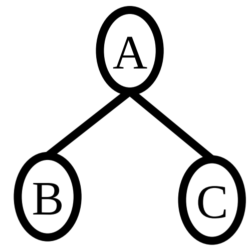
- Perform AND/OR tree or graph sampling on K
- Exact inference on R
- Orthogonal approaches:
  - Theorem: Combining AND/OR sampling and w-cutset sampling yields further variance reduction.



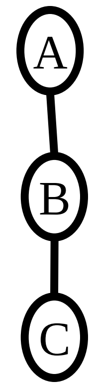
Graphical model



Full pseudo tree



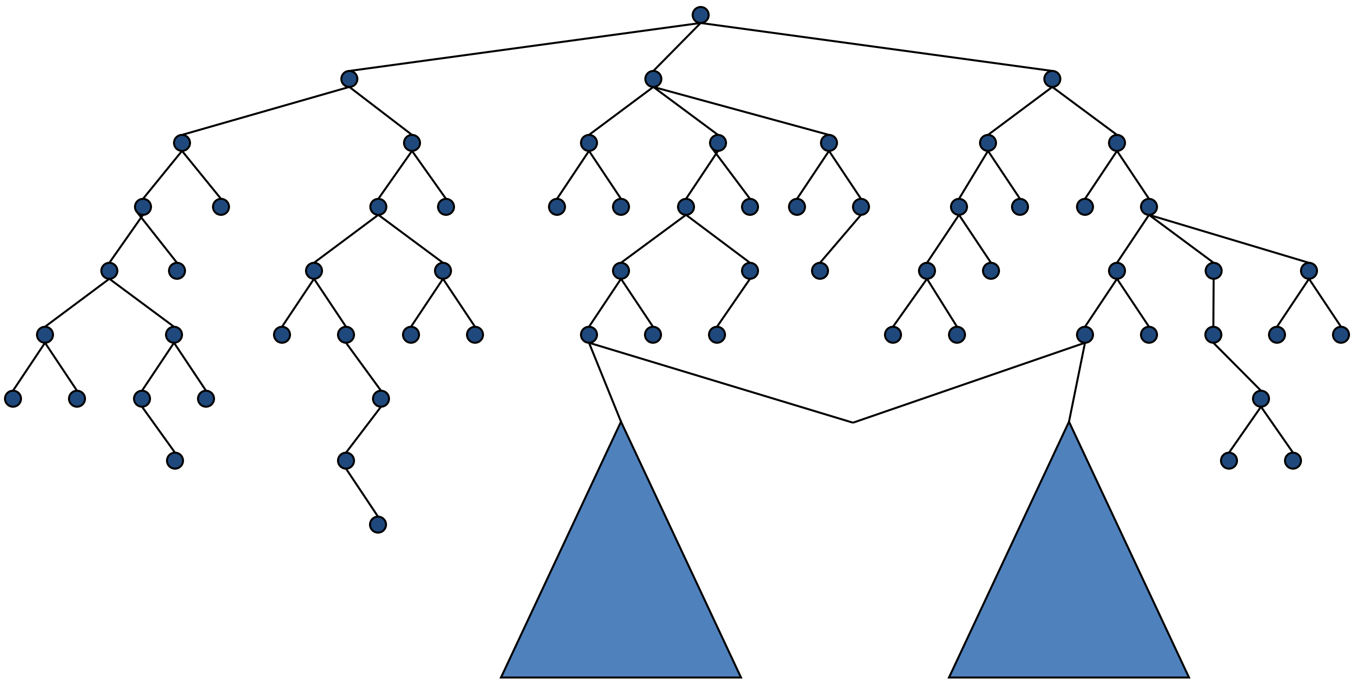
Start pseudo tree on the cutset variables



OR pseudo tree on the cutset variables

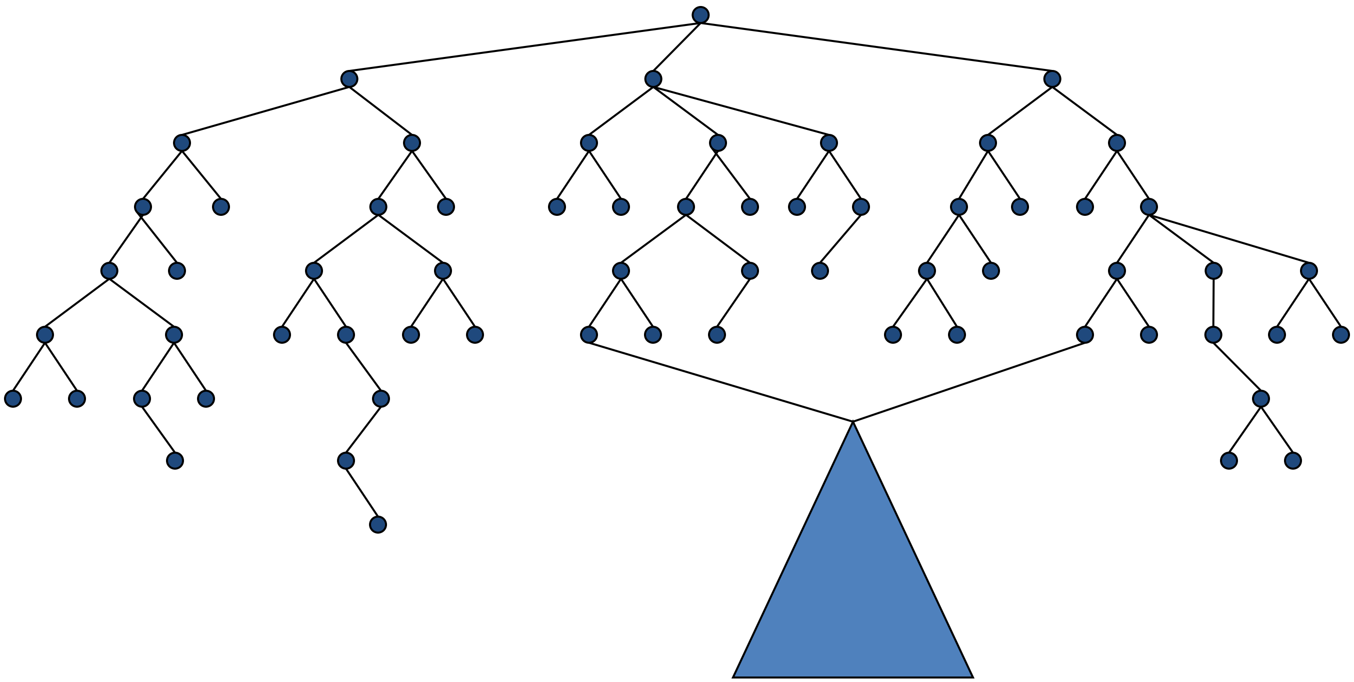
From Search Trees to Search Graphs

- Any two nodes that root identical subtrees (subgraphs) can be merged



From Search Trees to Search Graphs

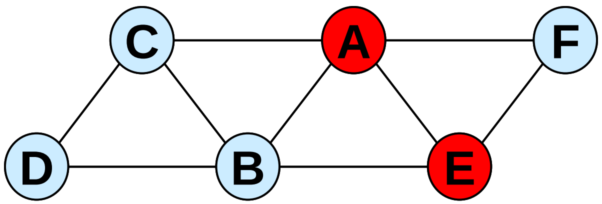
- Any two nodes that root identical subtrees (subgraphs) can be merged

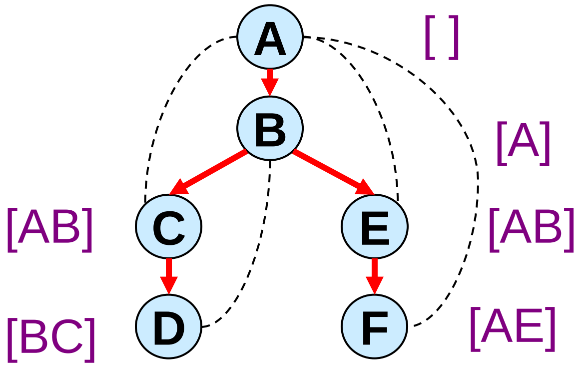


Merging Based on Context

- One way of recognizing nodes that can be merged:

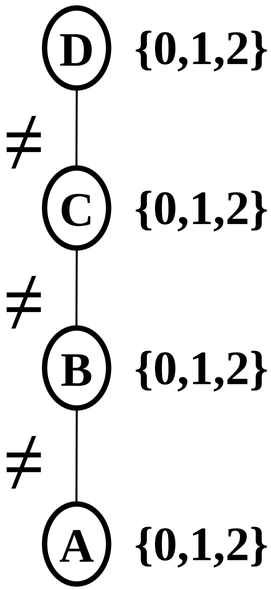
context (X) = ancestors of X in pseudo tree that are connected to X, or to descendants of X

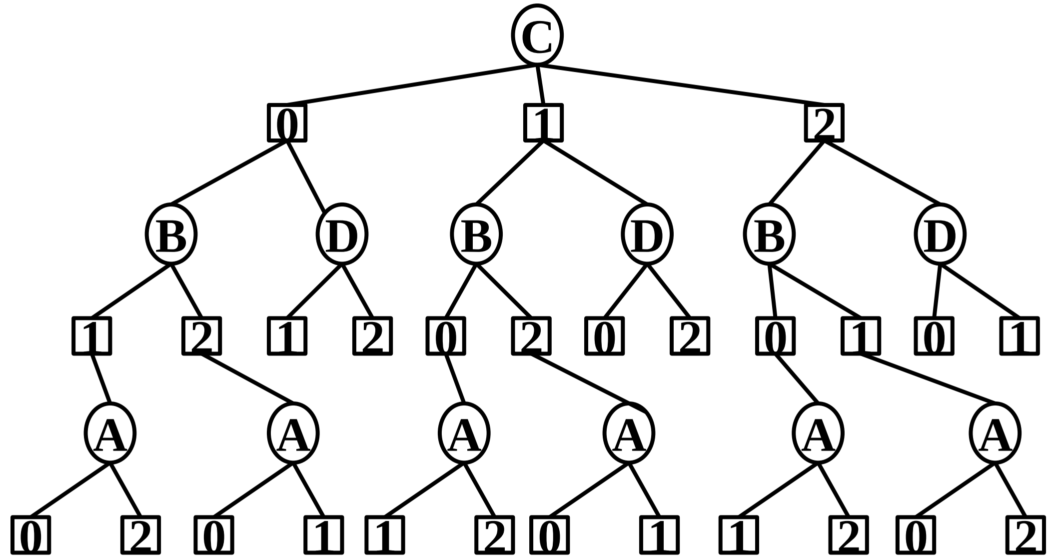




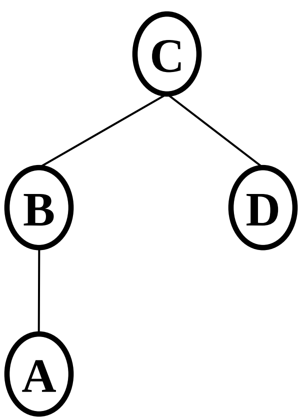
pseudo tree

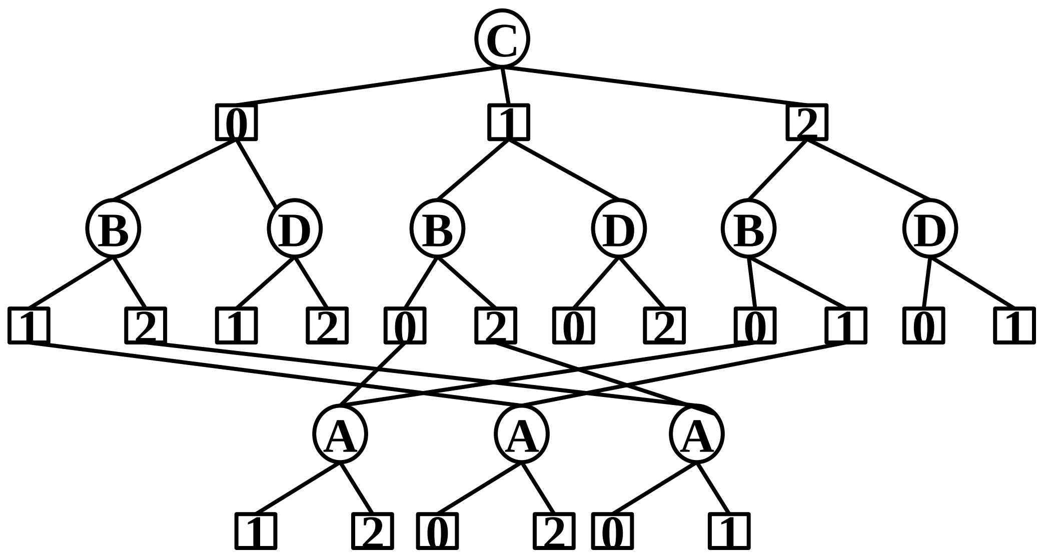
AND/OR Graphs





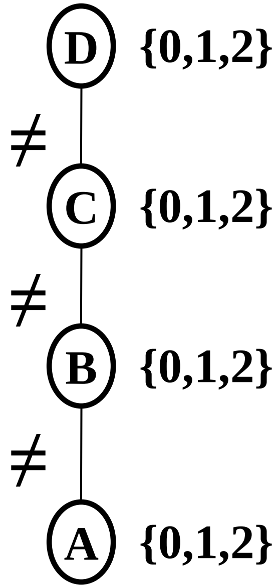
AND/OR tree

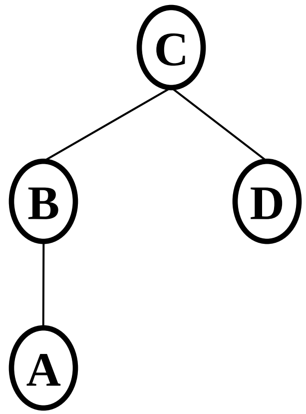




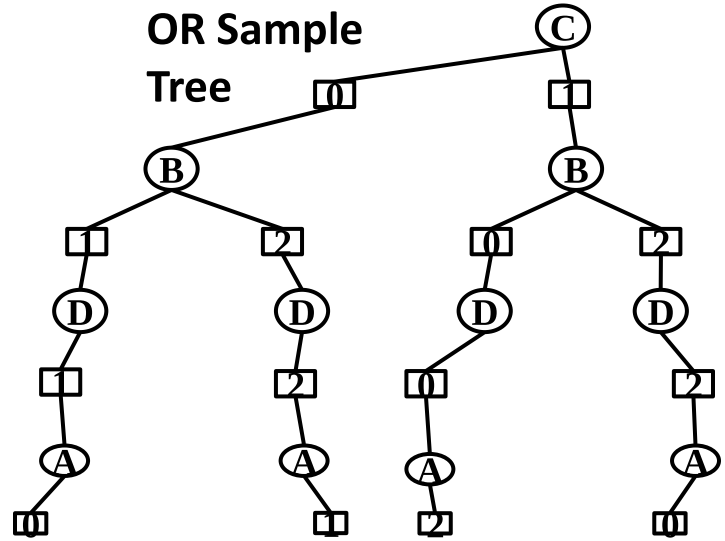
AND/OR graph

AND/OR graph sampling

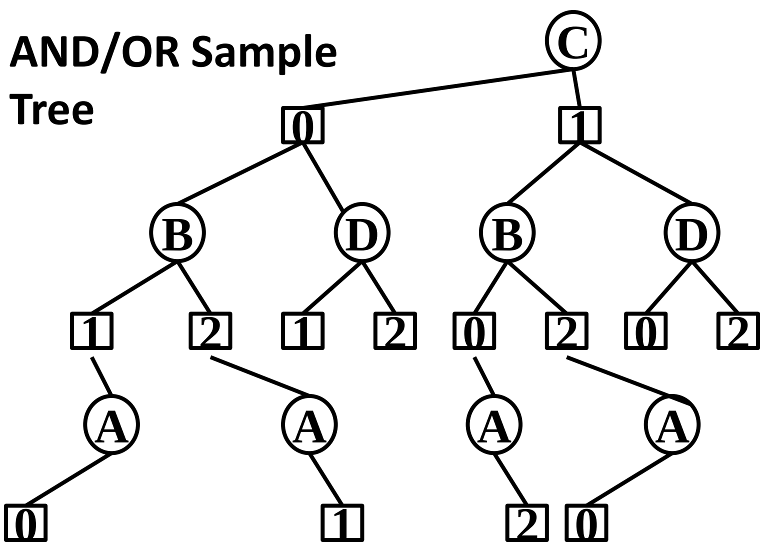




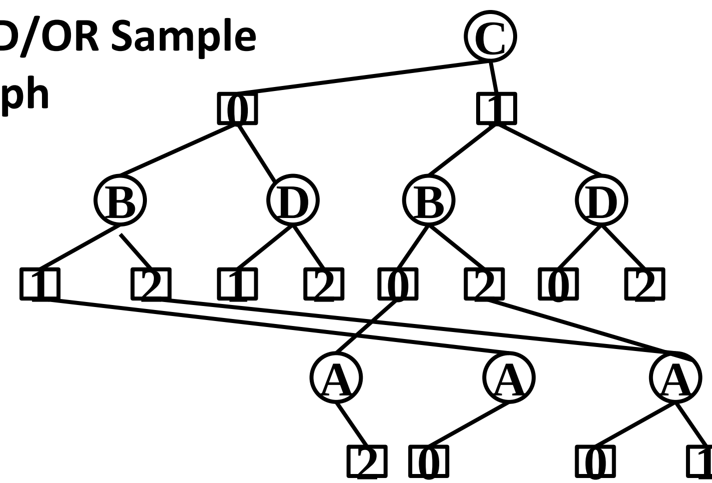
OR Sample Tree



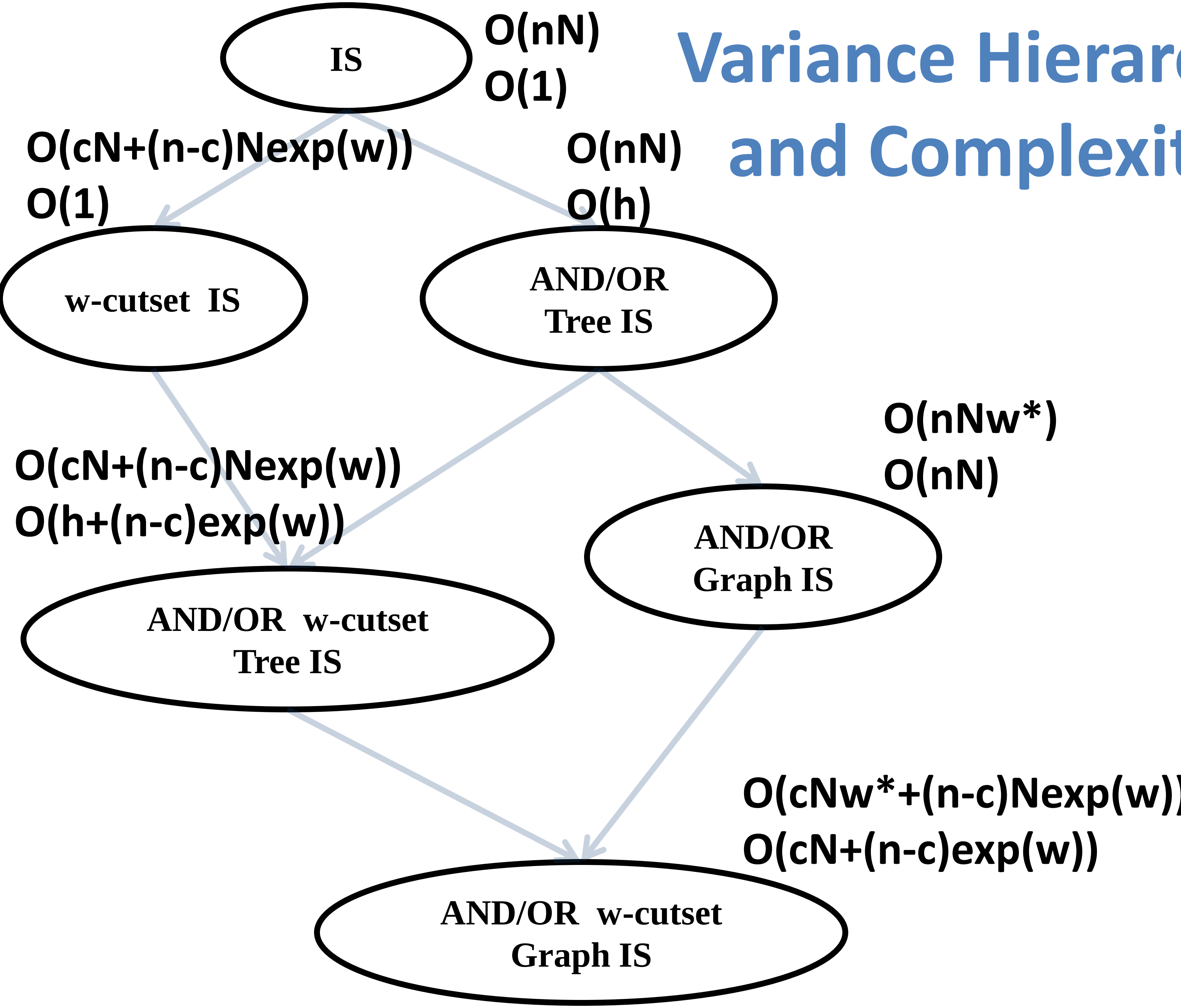
AND/OR Sample Tree



AND/OR Sample Graph



Variance Hierarchy and Complexity



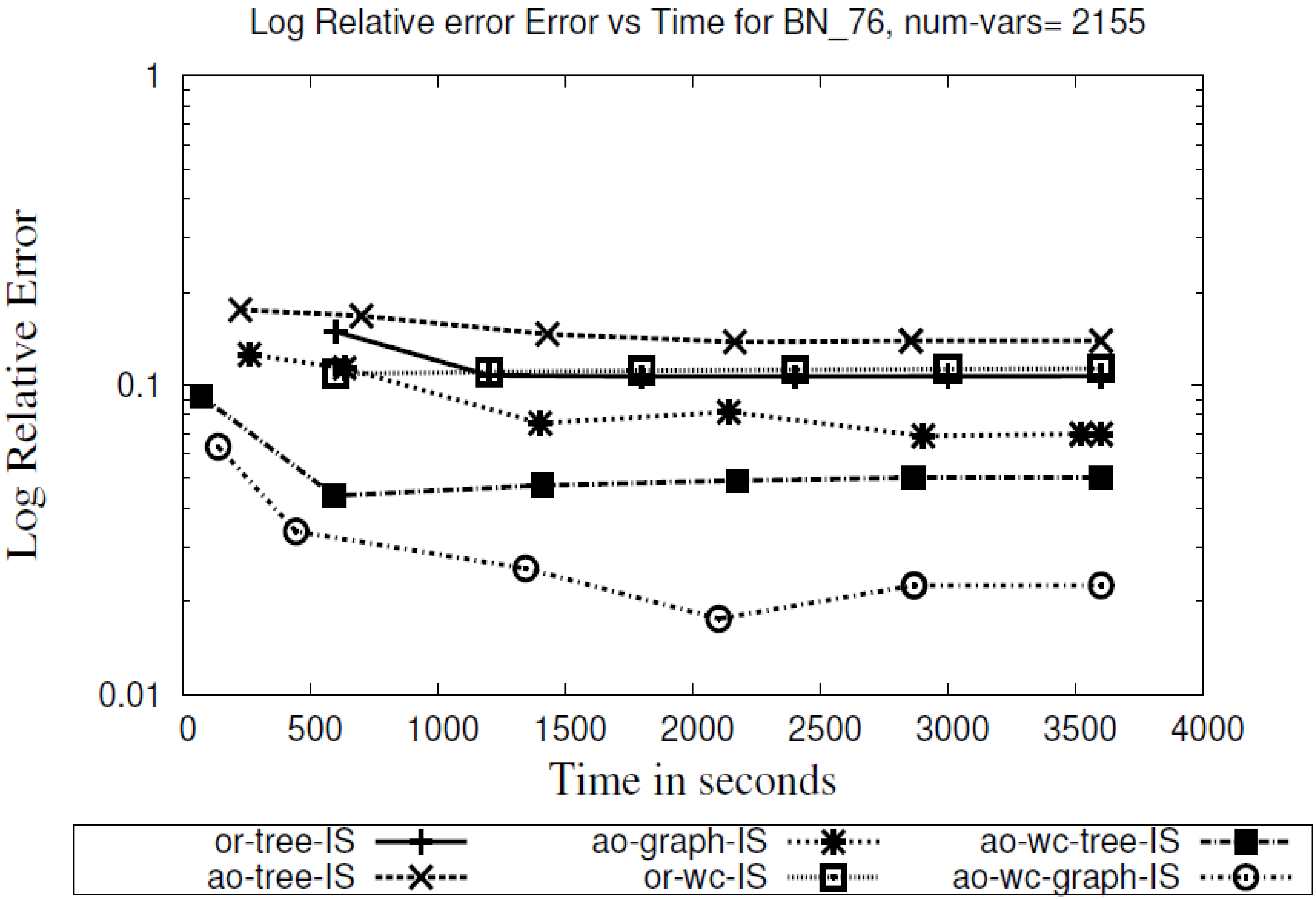
Experiments

- 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)

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

Time Bound: 1hr


Log Relative error Error vs Time for BN_76, num-vars= 2155
Log Relative Error
Time in seconds
1
0.1
0.01
0
500
1000
1500
2000
2500
3000
3500
4000
or-tree-IS
ao-tree-IS
ao-graph-IS
or-wc-IS
ao-wc-tree-IS
ao-wc-graph-IS

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.