Sampling Techniques for Probabilistic and Deterministic Graphical models

ICS 276, Spring 2013

Bozhena Bidyuk

Rina Dechter

Reading" Darwiche chapter 15, related papers

Overview

1. Probabilistic Reasoning/Graphical models
2. Importance Sampling
3. Markov Chain Monte Carlo: Gibbs Sampling
4. Sampling in presence of Determinism
5. Rao-Blackwellisation
6. AND/OR importance sampling

Overview

1. Probabilistic Reasoning/Graphical models
2. Importance Sampling
3. Markov Chain Monte Carlo: Gibbs Sampling
4. Sampling in presence of Determinism
5. Cutset-based Variance Reduction
6. AND/OR importance sampling

Probabilistic Reasoning; Graphical models

- Graphical models:
  - Bayesian network, constraint networks, mixed network
- Queries
- Exact algorithm
  - using inference,
  - search and hybrids
- Graph parameters:
  - tree-width, cycle-cutset, w-cutset

Queries

- **Probability of evidence (or partition function)**

$$P(e) = \sum_{X-\text{var}(e)} \prod_{i=1}^{n} P(x_i \mid pa_i)|_e \qquad Z = \sum_X \prod_i \psi_i(C_i)$$

- **Posterior marginal (beliefs):**

$$P(x_i \mid e) = \frac{P(x_i, e)}{P(e)} = \frac{\sum_{X-\text{var}(e)-X_i} \prod_{j=1}^{n} P(x_j \mid pa_j)|_e}{\sum_{X-\text{var}(e)} \prod_{j=1}^{n} P(x_j \mid pa_j)|_e}$$

- **Most Probable Explanation**

$$\bar{\mathbf{x}}^* = \arg\max_{\bar{\mathbf{x}}} P(\bar{\mathbf{x}}, \mathbf{e})$$

Approximation

- Since inference, search and hybrids are too expensive when graph is dense; (high treewidth) then:

- Bounding inference: (week 8)
  - mini-bucket and mini-clustering
  - Belief propagation

- **Bounding search: (week 7)**
  - **Sampling**

- Goal: an anytime scheme

Overview

1. Probabilistic Reasoning/Graphical models
2. Importance Sampling
3. Markov Chain Monte Carlo: Gibbs Sampling
4. Sampling in presence of Determinism
5. Rao-Blackwellisation
6. AND/OR importance sampling

Outline

- Definitions and Background on Statistics
- Theory of importance sampling
- Likelihood weighting
- State-of-the-art importance sampling techniques

A sample

- Given a set of variables X={X$_1$,...,X$_n$}, a sample, denoted by S^t is an instantiation of all variables:

$$S^t = (x_1^t, x_2^t, \ldots, x_n^t)$$

How to draw a sample ?
Univariate distribution

- Example: Given random variable X having domain {0, 1} and a distribution P(X) = (0.3, 0.7).
- Task: Generate samples of X from P.
- How?
  - draw random number r $\in$ [0, 1]
  - If (r < 0.3) then set X=0
  - Else set X=1

How to draw a sample? Multi-variate distribution

- Let X={X_1,..,X_n} be a set of variables
- Express the distribution in product form

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

- Sample variables one by one from left to right, along the ordering dictated by the product form.
- Bayesian network literature: Logic sampling

Sampling for Prob. Inference Outline

- **Logic Sampling**
- Importance Sampling
  - Likelihood Sampling
  - Choosing a Proposal Distribution
- Markov Chain Monte Carlo (MCMC)
  - Metropolis-Hastings
  - Gibbs sampling
- Variance Reduction

Logic Sampling: No Evidence (Henrion 1988)

Input: Bayesian network

$X = \{X_1, \ldots, X_N\}$, N- #nodes, T - # samples

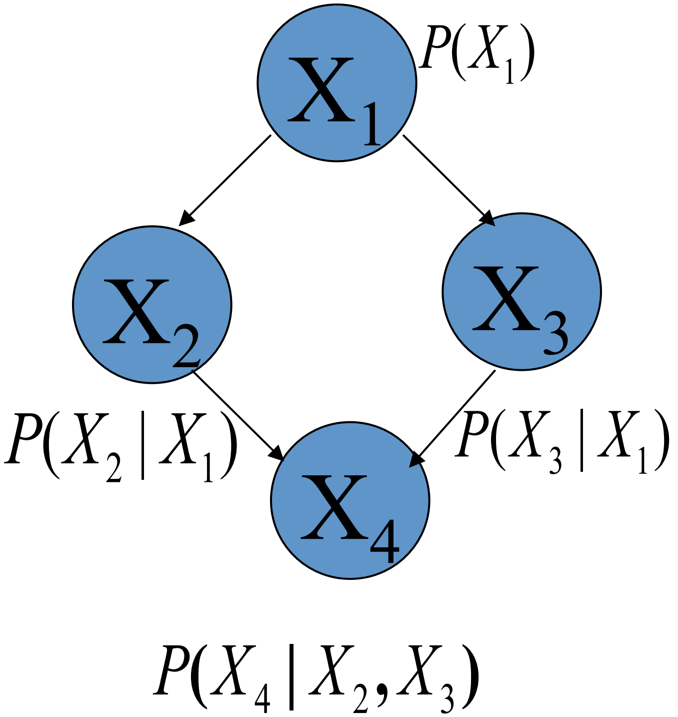
Output: T samples

Process nodes in topological order – first process the ancestors of a node, then the node itself:

```
For t = 0 to T
    For i = 0 to N
        $X_i \leftarrow$ sample $x_i^t$ from $P(x_i \mid pa_i)$
```

Logic sampling (example)

$$P(X_1, X_2, X_3, X_4) = P(X_1) \times P(X_2 \mid X_1) \times P(X_3 \mid X_1) \times P(X_4 \mid X_2, X_3)$$



No Evidence

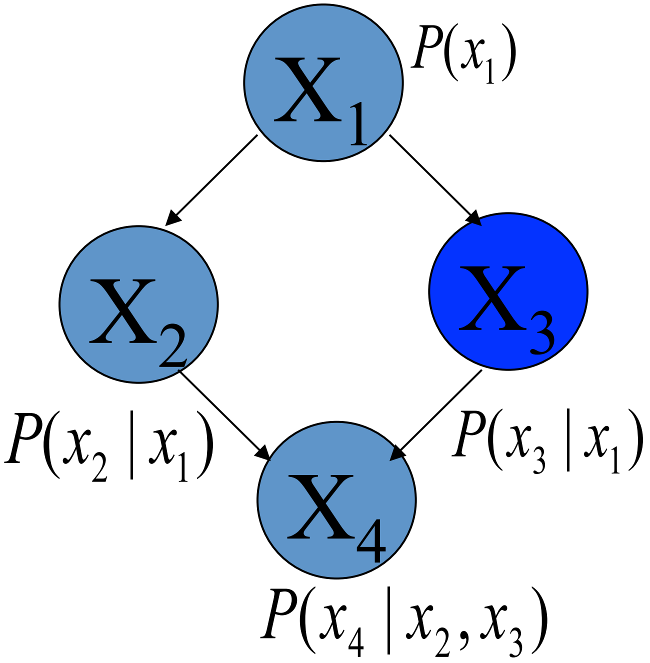
// generate sample k

1. Sample x_1 from $P(x_1)$
2. Sample x_2 from $P(x_2 \mid X_1 = x_1)$
3. Sample x_3 from $P(x_3 \mid X_1 = x_1)$
4. Sample x_4 from $P(x_4 \mid X_2 = x_2, X_3 = x_3)$

Logic Sampling w/ Evidence

```
Input: Bayesian network
      X= {$X_1$,…,$X_N$}, N- #nodes
      E – evidence, T - # samples
Output: T samples consistent with E
1. For t=1 to T
2.    For i=1 to N
3.       $X_i \leftarrow$ sample $x_i^t$ from $P(x_i \mid pa_i)$
4.       If $X_i$ in E and $X_i \neq x_i$, reject sample:
5.          Goto Step 1.
```

Logic Sampling (example)



$\text{Evidence}: X_3 = 0$

```
// generate sample $k$
1. Sample $x_1$ from $P(x_1)$
2. Sample $x_2$ from $P(x_2 | x_1)$
3. Sample $x_3$ from $P(x_3 | x_1)$
4. If $x_3 \neq 0$, reject sample
   and start from 1, otherwise
5. Sample $x_4$ from $P(x_4 | x_{2,} x_3)$
```

Expected value and Variance

Expected value: Given a probability distribution P(X) and a function g(X) defined over a set of variables X = {X_1, X_2, … X_n}, the expected value of g w.r.t. P is

$$E_P[g(x)] = \sum_x g(x)P(x)$$

Variance: The variance of g w.r.t. P is:

$$Var_P[g(x)] = \sum_x \left[g(x) - E_P[g(x)]\right]^2 P(x)$$

Monte Carlo Estimate

- **Estimator:**
  - An estimator is a function of the samples.
  - It produces an estimate of the unknown parameter of the sampling *distribution.*

Given i.i.d. samples $\mathrm{S}^1, \mathrm{S}^2, \ldots \mathrm{S}^{\mathrm{T}}$ drawn from P,

the Monte carlo estimate of $\mathrm{E}_P[\mathrm{g(x)}]$ is given by :

$$\hat{g} = \frac{1}{T} \sum_{t=1}^{T} g(S^t)$$

Example: Monte Carlo estimate

- Given:
  - A distribution P(X) = (0.3, 0.7).
  - g(X) = 40 if X equals 0

    = 50 if X equals 1.
- Estimate $E_P[g(x)]=(40x0.3+50x0.7)=47$.
- Generate k samples from P: 0,1,1,1,0,1,1,0,1,0

$$\hat{g} = \frac{40 \times \# samples(X = 0) + 50 \times \# samples(X = 1)}{\# samples}$$

$$= \frac{40 \times 4 + 50 \times 6}{10} = 46$$

Outline

- Definitions and Background on Statistics
- **Theory of importance sampling**
- Likelihood weighting
- State-of-the-art importance sampling techniques

Importance sampling: Main idea

- Express query as the expected value of a random variable w.r.t. to a distribution Q.
- Generate random samples from Q.
- Estimate the expected value from the generated samples using a monte carlo estimator (average).

Importance sampling for P(e)

$Let\ Z = X \setminus E,$

Let Q(Z) be a (proposal) distribution, satisfying

$P(z,e) > 0 \Rightarrow Q(z) > 0$

Then, we can rewrite P(e) as :

$$P(e) = \sum_{z} P(z,e) = \sum_{z} P(z,e)\frac{Q(z)}{Q(z)} = E_{Q}\left[\frac{P(z,e)}{Q(z)}\right] = E_{Q}[w(z)]$$

Monte Carlo estimate **:**

$$\hat{P}(e) = \frac{1}{T}\sum_{t=1}^{T} w(z^{t}), \text{where } z^{t} \leftarrow Q(Z)$$

Properties of IS estimate of P(e)

- **Convergence:** by law of large numbers

$$\hat{P}(e) = \frac{1}{T}\sum_{i=1}^{T} w(z^i) \xrightarrow{a.s.} P(e) \text{ for T} \rightarrow \infty$$

- **Unbiased.**

$$E_Q[\hat{P}(e)] = P(e)$$

- **Variance:**

$$Var_Q\left[\hat{P}(e)\right] = Var_Q\left[\frac{1}{T}\sum_{i=1}^{N} w(z^i)\right] = \frac{Var_Q[w(z)]}{T}$$

Properties of IS estimate of P(e)

- Mean Squared Error of the estimator

$$MSE_Q\left[\hat{P}(e)\right] = E_Q\left[\left(\hat{P}(e) - P(e)\right)^2\right]$$

$$= \left(P(e) - E_Q[\hat{P}(e)]\right)^2 + Var_Q\left[\hat{P}(e)\right]$$

$$= Var_Q\left[\hat{P}(e)\right]$$

$$= \frac{Var_Q[w(x)]}{T}$$

This quantity enclosed in the brackets is zero because the expected value of the estimator equals the expected value of g(x)

Estimating P(X$_i$|e)

Let $\delta_{\mathrm{x_i}}$ (z) be a dirac - delta function, which is 1 if z contains x$_i$ and 0 otherwise.

$$P(x_i \mid e) = \frac{P(x_i, e)}{P(e)} = \frac{\sum_z \delta_{x_i}(z) P(z,e)}{\sum_z P(z,e)} = \frac{E_Q\left[\frac{\delta_{x_i}(z)P(z,e)}{Q(z)}\right]}{E_Q\left[\frac{P(z,e)}{Q(z)}\right]}$$

Idea : Estimate numerator and denominator by IS.

Ratio estimate : $\overline{P}(x_i \mid e) = \frac{\hat{P}(x_i, e)}{\hat{P}(e)} = \frac{\sum_{\mathrm{k=1}}^{\mathrm{T}} \delta_{x_i}(z^k) w(z^k, e)}{\sum_{\mathrm{k=1}}^{\mathrm{T}} w(z^k, e)}$

Estimate is biased : $\mathrm{E}\left[\overline{P}(x_i \mid e)\right] \neq P(x_i \mid e)$

Properties of the IS estimator for $P(X_i|e)$

- Convergence: By Weak law of large numbers

$$\overline{P}(x_i \mid e) \rightarrow P(x_i \mid e) \text{ as T} \rightarrow \infty$$

- Asymptotically unbiased

$$\lim_{T\rightarrow\infty} E_P[\overline{P}(x_i \mid e)] = P(x_i \mid e)$$

- Variance
  - Harder to analyze
  - Liu suggests a measure called “Effective sample size”

Generating samples from Q

- No restrictions on “how to”
- Typically, express Q in product form:
  - $Q(Z)=Q(Z_1)xQ(Z_2|Z_1)x....xQ(Z_n|Z_1,..Z_{n-1})$
- Sample along the order $Z_1,..,Z_n$
- Example:
  - $Z_1 \leftarrow Q(Z_1)=(0.2,0.8)$
  - $Z_2 \leftarrow Q(Z_2|Z_1)=(0.1,0.9,0.2,0.8)$
  - $Z_3 \leftarrow Q(Z_3|Z_1,Z_2)=Q(Z_3)=(0.5,0.5)$

Outline

- Definitions and Background on Statistics
- Theory of importance sampling
- **Likelihood weighting**
- State-of-the-art importance sampling techniques

Likelihood Weighting

(Fung and Chang, 1990; Shachter and Peot, 1990)

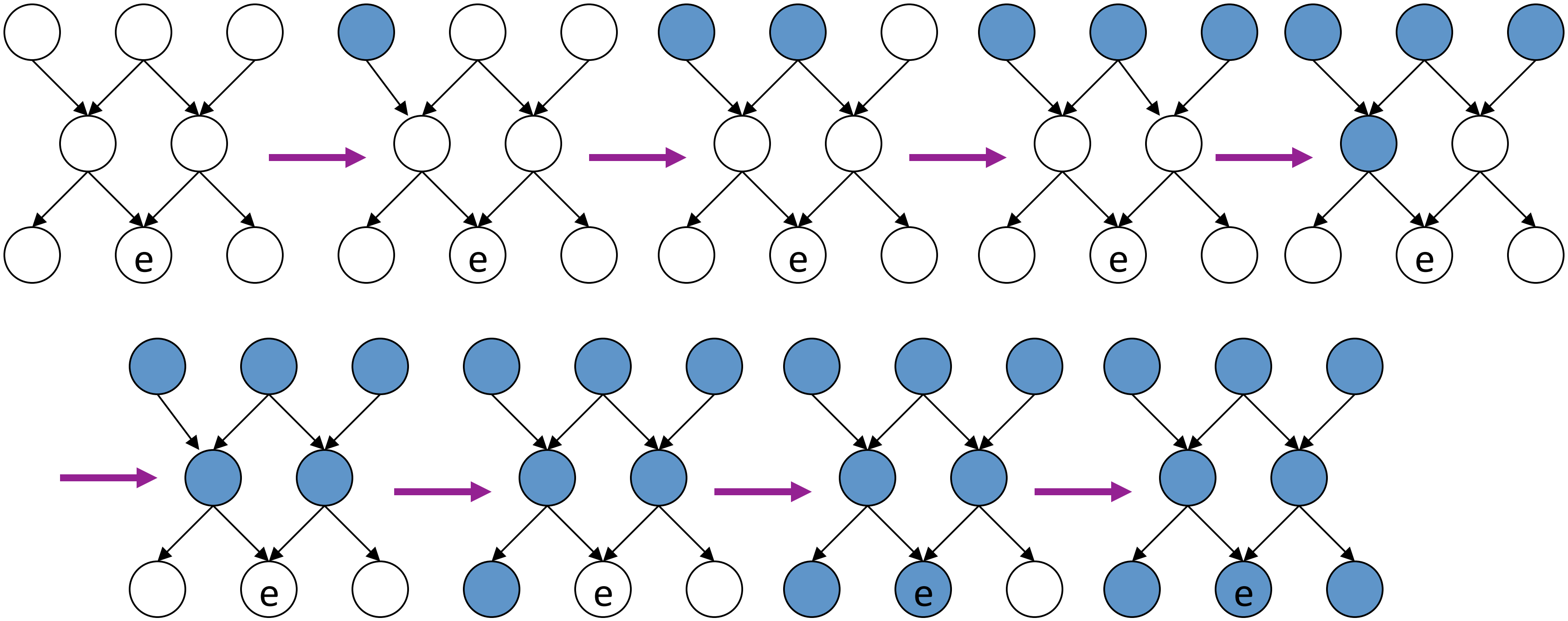
Is an instance of importance sampling!

“Clamping” evidence+
logic sampling+
weighing samples by evidence likelihood

Works well for *likely evidence!*

Likelihood Weighting: Sampling

Sample in topological order over **X** !



Clamp evidence, Sample $x_i \leftarrow P(X_i | pa_i)$, $P(X_i | pa_i)$ is *a look-up* in CPT!

Likelihood Weighting: Proposal Distribution

$$Q(X \setminus E) = \prod_{X_i \in X \setminus E} P(X_i \mid pa_i, e)$$

Notice: Q is another Bayesian network

Example:

Given a Bayesian network: $\mathrm{P}(\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3) = \mathrm{P}(\mathrm{X}_1) \times \mathrm{P}(\mathrm{X}_2 \mid \mathrm{X}_1) \times \mathrm{P}(\mathrm{X}_3 \mid \mathrm{X}_1, \mathrm{X}_2)$ and

Evidence $\mathrm{X}_2 = \mathrm{x}_2$.

$$\mathrm{Q}(\mathrm{X}_1, \mathrm{X}_3) = P(X_1) \times P(X_3 \mid X_1, X_2 = x_2)$$

Weights **:**

Given a sample: $x = (x_1, .., x_n)$

$$w = \frac{P(x, e)}{Q(x)} = \frac{\cancel{\prod_{X_i \in X \setminus E} P(x_i \mid pa_i, e)} \times \prod_{E_j \in E} P(e_j \mid pa_j)}{\cancel{\prod_{X_i \in X \setminus E} P(x_i \mid pa_i, e)}}$$

$$= \prod_{E_j \in E} P(e_j \mid pa_j)$$

Likelihood Weighting: Estimates

Estimate P(e): $\hat{P}(e) = \frac{1}{T}\sum_{t=1}^{T} w^{(t)}$

Estimate Posterior Marginals:

$$\hat{P}(x_i \mid e) = \frac{\hat{P}(x_i, e)}{\hat{P}(e)} = \frac{\sum_{t=1}^{T} w^{(t)} g_{x_i}(x^{(t)})}{\sum_{t=1}^{T} w^{(t)}}$$

$g_{x_i}(x^{(t)}) = 1 \text{ if } \mathrm{x_i} = x_i^t \text{and equals zero otherwise}$

Likelihood Weighting

- Converges to exact posterior marginals
- Generates Samples Fast
- Sampling distribution is close to prior (especially if $E \subset$ Leaf Nodes)
- Increasing sampling variance

$\Rightarrow$Convergence may be slow

$\Rightarrow$Many samples with $P(x^{(t)})=0$ rejected

Outline

- Definitions and Background on Statistics
- Theory of importance sampling
- Likelihood weighting
- **Error estimation**
- State-of-the-art importance sampling techniques

Bounds on the Absolute Error

The absolute error of an estimate $\mathrm{Av}_n(\breve{\alpha})$

is the absolute difference it has with the true probability $\Pr(\alpha)$ we are trying to estimate.

For any $\epsilon > 0$, we have

$$\mathbb{P}\Big(|\mathrm{Av}_n(\breve{\alpha}) - \Pr(\alpha)| < \epsilon\Big) \geq 1 - \frac{\Pr(\alpha)\Pr(\neg\alpha)}{n\epsilon^2}$$

The estimate $\mathrm{Av}_n(\breve{\alpha})$ computed by direct sampling will fall within the interval $(\Pr(\alpha) - \epsilon, \Pr(\alpha) + \epsilon)$ with probability at least $1 - \Pr(\alpha)\Pr(\neg\alpha)/n\epsilon^2$

Bounds on the Absolute Error

A sharper bound which does not depend on the probability $\Pr(\alpha)$

Hoeffding's inequality

Let $\mathrm{Av}_n(f)$ be a sample mean, where the function f has expectation μ and values in $\{0, 1\}$. For any $\epsilon > 0$, we have:

$$\mathbb{P}\Big(|\mathrm{Av}_n(f) - \mu| \leq \epsilon\Big) \geq 1 - 2e^{-2n\epsilon^2}$$

For any $\epsilon > 0$, we have:

$$\mathbb{P}\Big(|\mathrm{Av}_n(\breve{\alpha}) - \Pr(\alpha)| \leq \epsilon\Big) \geq 1 - 2e^{-2n\epsilon^2}$$

The estimate $\mathrm{Av}_n(\breve{\alpha})$ computed by direct sampling will fall within the interval $(\Pr(\alpha) - \epsilon, \Pr(\alpha) + \epsilon)$ with probability at least $1 - 2e^{-2n\epsilon^2}$

Bounds on the Relative Error

For any $\epsilon > 0$, we have:

$$\mathbb{P}\left(\frac{|\mathrm{Av}_n(\breve{\alpha}) - \mathrm{Pr}(\alpha)|}{\mathrm{Pr}(\alpha)} \leq \epsilon\right) \geq 1 - 2e^{-2n\epsilon^2\mathrm{Pr}(\alpha)^2}$$

Require the probability $\mathrm{Pr}(\alpha)$ (or some lower bound on it).

Bounds on the Relative Error

The relative error of an estimate $\mathrm{Av}_n(\breve{\alpha})$

$$\frac{|\mathrm{Av}_n(\breve{\alpha}) - \Pr(\alpha)|}{\Pr(\alpha)}$$

The bound on the absolute error becomes tighter as the probability of an event becomes more extreme. Yet, the corresponding bound on the relative error becomes looser as the probability of an event becomes more extreme.

Example

For an event with probability .5 and a sample size of 10000, there is a 95% chance that the absolute error is $\approx$ 4.5%. However, for the same confidence level, the relative error increases to $\approx$ 13.4% if the event has probability .1, and increases again to $\approx$ 44.5% if the event has probability .01

Outline

- Definitions and Background on Statistics
- Theory of importance sampling
- Likelihood weighting
- **State-of-the-art importance sampling techniques**

Proposal selection

- One should try to select a proposal that is as close as possible to the posterior distribution.

$$Var_Q\left[\hat{P}(e)\right] = \frac{Var_Q[w(z)]}{T} = \frac{1}{N}\sum_{z \in Z}\left(\frac{P(z,e)}{Q(z)} - P(e)\right)^2 Q(z)$$

$$\frac{P(z,e)}{Q(z)} - P(e) = 0, \text{ to have a zero - variance estimator}$$

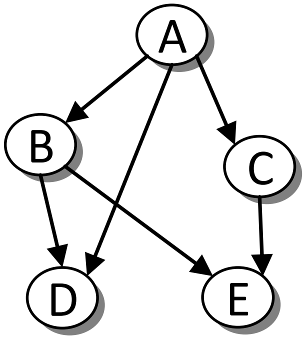
$$\therefore \frac{P(z,e)}{P(e)} = Q(z)$$

$$\therefore Q(z) = P(z \mid e)$$

Perfect sampling using Bucket Elimination

- Algorithm:
  - Run Bucket elimination on the problem along an ordering o=(X_N,..,X_1).
  - Sample along the reverse ordering: (X_1,..,X_N)
  - At each variable X_i, recover the probability P(X_i| x_1,...,x_{i-1}) by referring to the bucket.

Bucket Elimination



Query: $P(a|e=0) \propto P(a,e=0)$ Elimination Order: d,e,b,c

$$P(a,e=0) = \sum_{c,b,e=0,d} P(a)P(b|a)P(c|a)P(d|a,b)P(e|b,c)$$

$$= P(a)\sum_{c} P(c|a)\sum_{b} P(b|a)\sum_{e=0} P(e|b,c)\sum_{d} P(d|a,b)$$

Original Functions

D: $P(d|a,b)$

E: $P(e|b,c)$

B: $P(b|a)$

C: $P(c|a)$

A: $P(a)$

Messages

$$\boxed{f_D(a,b)} = \sum_{d} P(d|a,b)$$

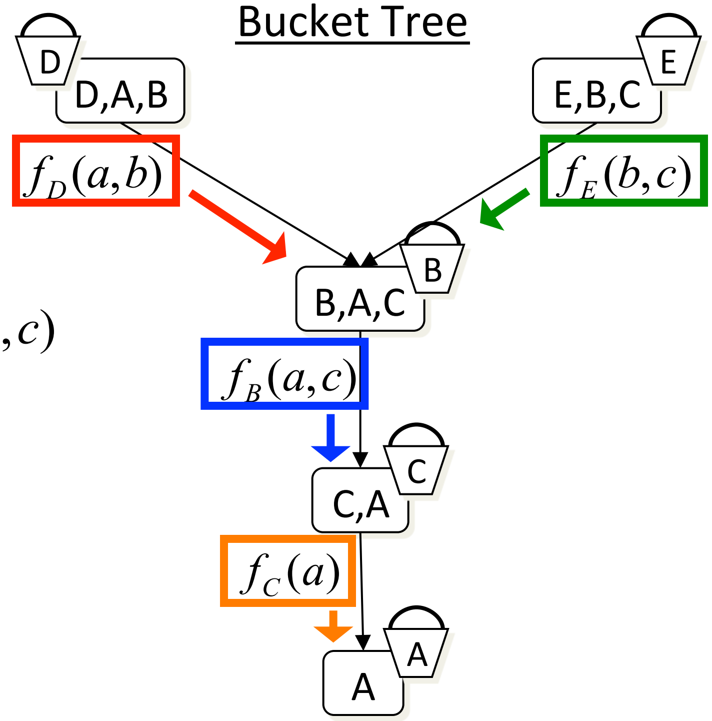
$$\boxed{f_E(b,c)} = P(e=0|b,c)$$

$$\boxed{f_B(a,c)} = \sum_{b} P(b|a)f_D(a,b)f_E(b,c)$$

$$\boxed{f_C(a)} = \sum_{c} P(c|a)f_B(a,c)$$

$$P(a,e=0) = p(A)f_C(a)$$

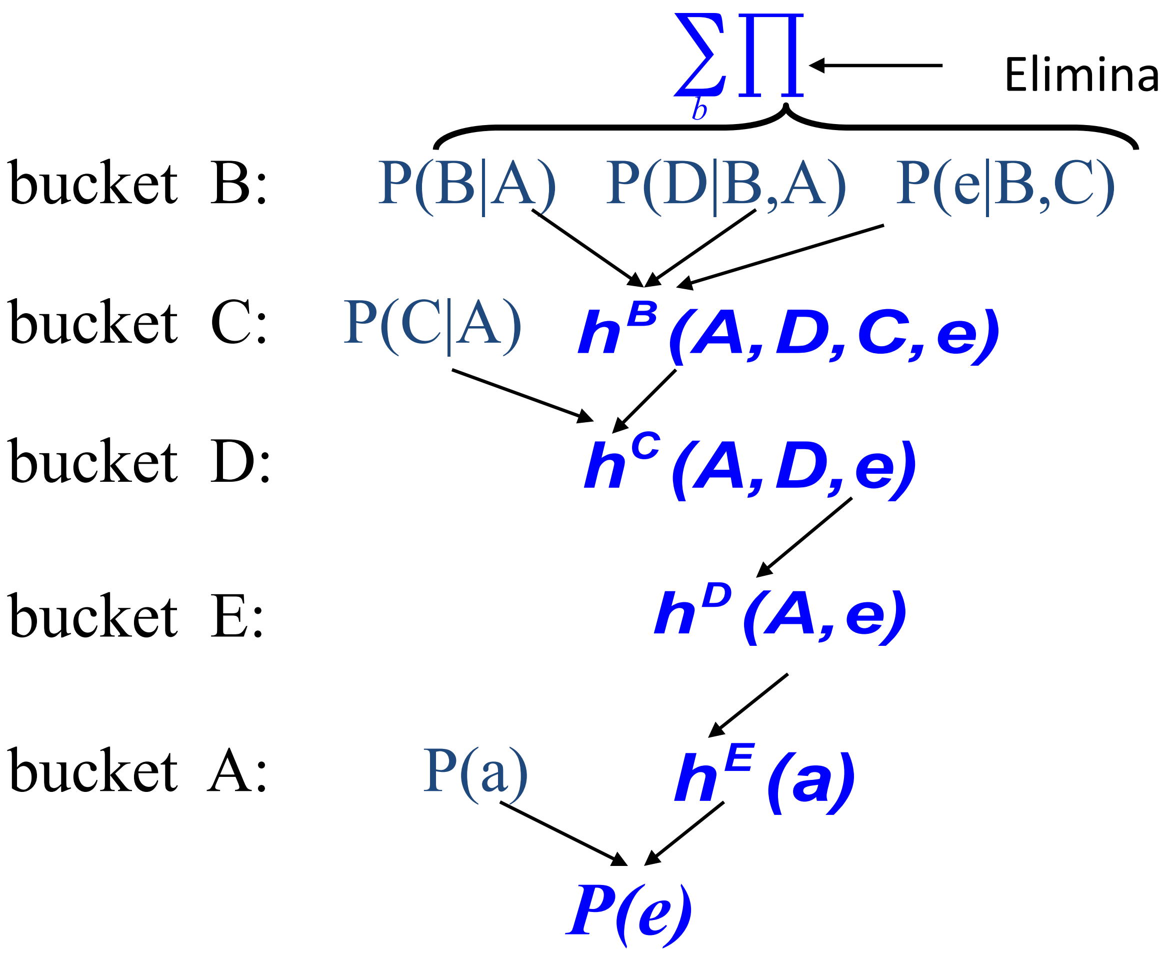
Bucket Tree



Time and space exp(w*)

Bucket elimination (BE)

Algorithm *elim-bel* (Dechter 1996)



$\underbrace{\sum_b \prod}$ ← Elimination operator

bucket B: P(B|A) P(D|B,A) P(e|B,C)

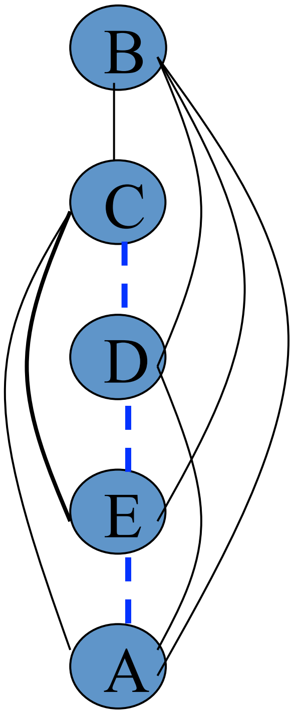
bucket C: P(C|A) $h^B(A,D,C,e)$

bucket D: $h^C(A,D,e)$

bucket E: $h^D(A,e)$

bucket A: P(a) $h^E(a)$

$P(e)$



Sampling from the output of BE

(Dechter 2002)

$\mathrm{Set\ A = a, D = d, C = c\ in\ the\ bucket}$

$\mathrm{Sample: B = b \leftarrow Q(B \mid a, e, d)} \propto P(B \mid a)P(d \mid B, a)P(e \mid b, c)$

bucket B: $P(B|A) \quad P(D|B,A) \quad P(e|B,C)$

bucket C: $P(C|A) \quad \boldsymbol{h^{B}(A,D,C,e)}$

$\mathrm{Set\ A = a, D = d\ in\ the\ bucket}$

$\mathrm{Sample: C = c \leftarrow Q(C \mid a, e, d)} \propto P(C \mid A) \cdot \mathrm{h^{B}(a, d, C, e}$

bucket D: $\boldsymbol{h^{C}(A,D,e)}$

$\mathrm{Set\ A = a\ in\ the\ bucket}$

$\mathrm{Sample: D = d \leftarrow Q(D \mid a, e) \propto h^{C}(a, D, e)}$

bucket E: $\boldsymbol{h^{D}(A,e)}$

$\mathrm{Evidence\ bucket: ignore}$

bucket A: $P(A) \quad \boldsymbol{h^{E}(A)}$

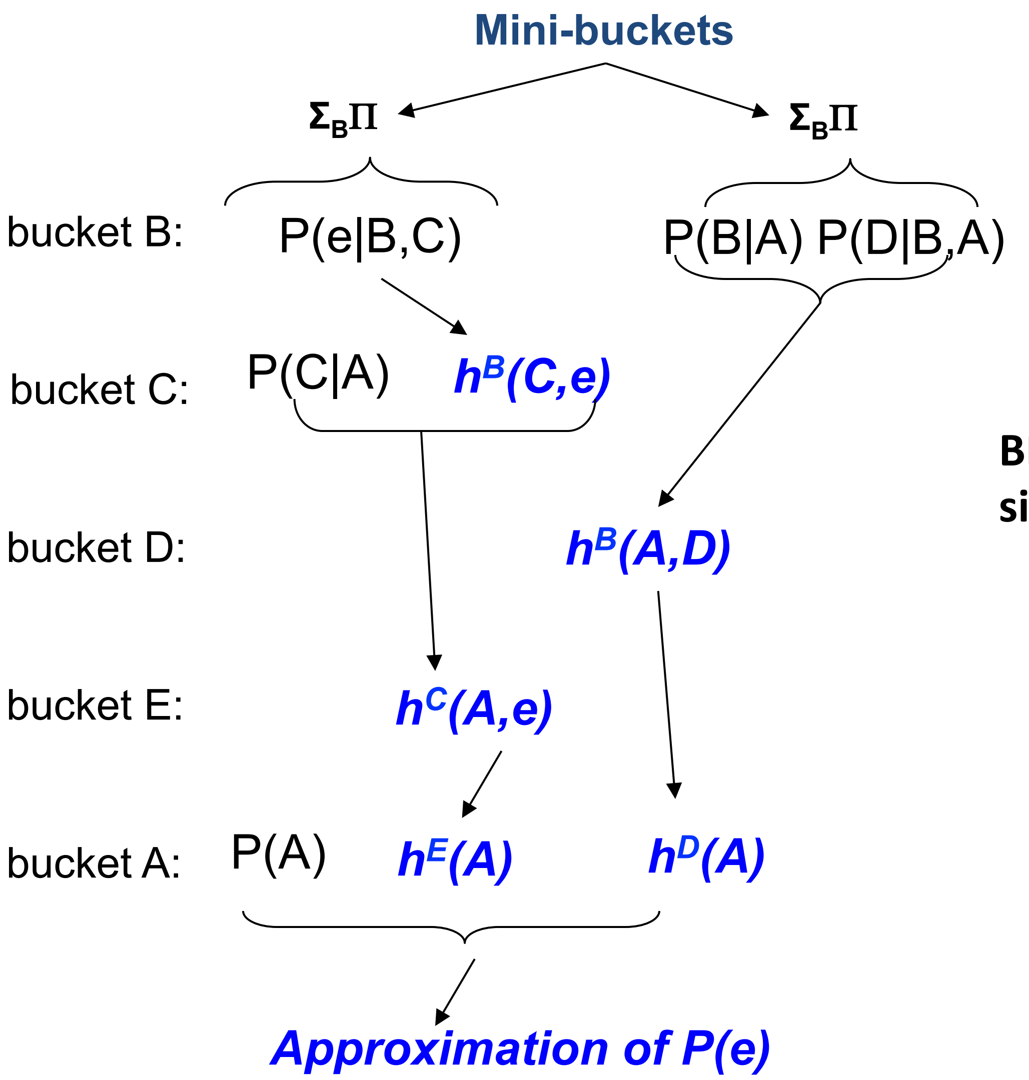
$$\mathbf{Q(A) \propto P(A) \times h^{E}(A)}$$

$$\mathit{Sample}\ \mathbf{: A = a \leftarrow Q(A)}$$

Mini-buckets: “local inference”

- Computation in a bucket is time and space exponential in the number of variables involved
- Therefore, partition functions in a bucket into “mini-buckets” on smaller number of variables
- Can control the size of each “mini-bucket”, yielding polynomial complexity.

Mini-Bucket Elimination



Mini-buckets

$\Sigma_B \Pi$ $\Sigma_B \Pi$

bucket B: $P(e|B,C)$ $P(B|A)\ P(D|B,A)$

bucket C: $P(C|A)$ $h^B(C,e)$

bucket D: $h^B(A,D)$

bucket E: $h^C(A,e)$

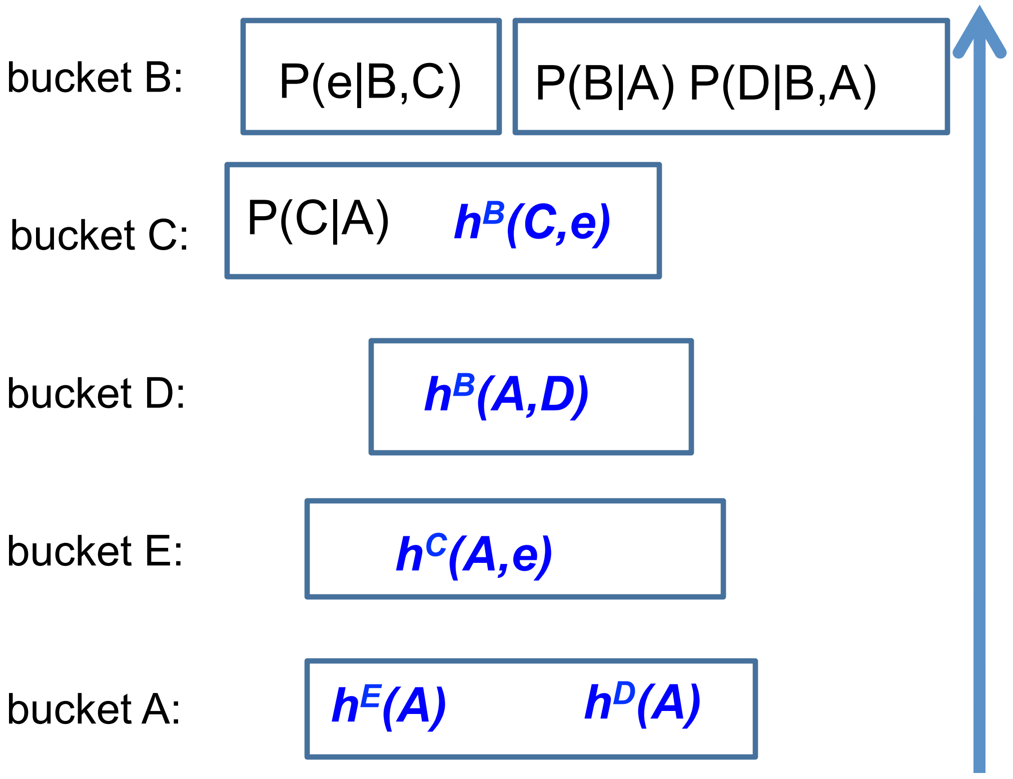
bucket A: $P(A)$ $h^E(A)$ $h^D(A)$

Approximation of $P(e)$

Space and Time constraints: Maximum scope size of the new function generated should be bounded by 2

BE generates a function having scope size 3. So it cannot be used.

Sampling from the output of MBE



bucket B: P(e|B,C) P(B|A) P(D|B,A)

bucket C: P(C|A) $h^B(C,e)$

bucket D: $h^B(A,D)$

bucket E: $h^C(A,e)$

bucket A: $h^E(A)$ $h^D(A)$

Sampling is same as in BE-sampling except that now we construct Q from a randomly selected "mini-bucket"

IJGP-Sampling (Gogate and Dechter, 2005)

- Iterative Join Graph Propagation (IJGP)
  - A Generalized Belief Propagation scheme (Yedidia et al., 2002)
- IJGP yields better approximations of P(X|E) than MBE
  - (Dechter, Kask and Mateescu, 2002)
- Output of IJGP is same as mini-bucket "clusters"
- **Currently the best performing IS scheme!**

Current Research question

- Given a Bayesian network with evidence or a Markov network representing function P, generate another Bayesian network representing a function Q (from a family of distributions, restricted by structure) such that Q is closest to P.
- Current approaches
  - Mini-buckets
  - Ijgp
  - Both
- Experimented, but need to be justified theoretically.

Algorithm: Approximate Sampling

1) Run IJGP or MBE
2) At each branch point compute the edge probabilities by consulting output of IJGP or MBE

- Rejection Problem:
  - Some assignments generated are non solutions

Adaptive Importance Sampling

```
Initial Proposal = $Q^1(Z) = Q(Z_1)\times Q(Z_2 \mid pa(Z_2))\times\ldots\times Q(Z_n \mid pa(Z_n))$
$\hat{P}(E=e)=0$
For i = 1 to k do
    Generate samples $z^1,\ldots,z^N$ from $Q^k$
    $\hat{P}(E=e) = \hat{P}(E=e) + \frac{1}{N}\sum_{j=1}^{N} w_k(z^i)$
    Update $Q^{k+1} = Q^k + \eta(k)\left[Q^k - Q'\right]$
End
Return $\frac{\hat{P}(E=e)}{k}$
```

Adaptive Importance Sampling

- General case
- Given k proposal distributions
- Take N samples out of each distribution
- Approximate P(e)

$$\hat{P}(e) = \frac{1}{k}\sum_{j=1}^{k}\left[Avg - weight - jth - proposal\right]$$

Estimating Q'(z)

$$\mathrm{Q}^{'}(Z) = Q'(Z_1) \times Q'(Z_2 \mid pa(Z_2)) \times \ldots \times Q'(Z_n \mid pa(Z_n))$$

where each $\mathrm{Q'}(\mathrm{Z_i} \mid \mathrm{Z_1},..,\mathrm{Z_{i\text{-}1}})$

is estimated by importance sampling

Overview

1. Probabilistic Reasoning/Graphical models
2. Importance Sampling
3. Markov Chain Monte Carlo: Gibbs Sampling
4. Sampling in presence of Determinism
5. Rao-Blackwellisation
6. AND/OR importance sampling