Today

- Outline
- Hardness of Approximation
- Simulation Techniques
  - Simulation
  - Simplify Computation
  - Structural Approximation

Simulation

- Techniques
  - Gibbs Sampling (MCMC)
  - Forward Sampling
    - Logic Sampling
    - Likelihood Weighting
    - Bounded Variance
    - AA
  - Backward Sampling
  - Metropolis (MCMC)
- Numeric Integration
  - Gibbs+Metropolis: Samples drawn from Markov Chain
  - Forward+Backward: Independent samples

Approximate Marginal

Techniques:

- Bounded Conditioning (Horvitz)
- Single Fault Approximation
- Adaptive Hugin (Kjaerwulf)
- Search (Henrion, Poole)
- Approximate Mixtures
  - Poland, West, Bar-Shalom, many others (later in course)

Drop or combine terms

- Drop small probability terms.
- Combine similar terms.

Simplify Distribution

Assume Extra Independence

- Variational Approximation (Jordan) (May do this)
- LPE (Draper)
- Approximate Factoring (Boyen + Koller)

Solve a problem you know that you can solve.

Measuring Performance

Absolute Error

$$P\{X = x \mid E = e\} - \varepsilon \le \rho \le P\{X = x \mid E = e\} + \varepsilon$$

Relative Error

$$P\{X = x \mid E = e\}(1 - \varepsilon) \le \rho \le P\{X = x \mid E = e\}(1 + \varepsilon)$$

Epsilon-Delta Bound

$$P\{\varepsilon \text{ relative approximation to } \mathrm{P}\{\mathrm{X} \mid \mathrm{E}\}\} \ge 1 - \delta$$

Hardness Results: Relative Error

(Koller) The problem of finding ρ with a relative error ε to P{X=x} in a BN is NP-hard

Proof:

PIBNETB: P{'AND'} > 0

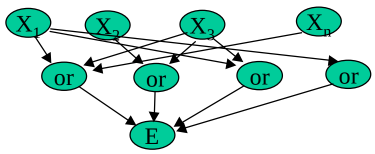
Use a "relative error" algorithm to estimate ρ to relative error ε

ρ > 0 iff P{'AND'} > 0

Hardness Results: Absolute Error

(Koller) The problem of finding an approximation ρ with absolute error ε < 1/2 to P{X=x|E=e} is NP-hard

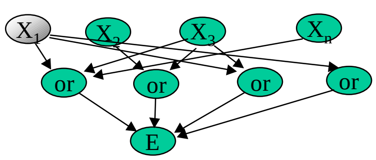
Proof:

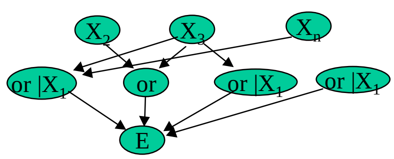


Estimate $P\{X_1|E=T\}$

If $\rho > 0$ Pick $X_1 = T$ [we know that $P\{X_1 = T|E = T\} > 0$]
o/w Pick $X_1 = F$

Absolute Error Proof Continued





Estimate $P\{X_2|E=T\}, \ldots, P\{X_n|E=T\}$
choosing X_i that maximizes $\rho\{X_i|E=T\}$,
$P\{X\} > 0$

Hardness: Absolute Epsilon-Delta

(Dagum+Luby)

The problem of finding ρ such that

$$P\left\{P\{X=x \mid E=e\}-\frac{1}{2}<\rho<P\{X=x \mid E=e\}+\frac{1}{2}\right\}>\frac{1}{2}$$

is NP-hard

Simulation

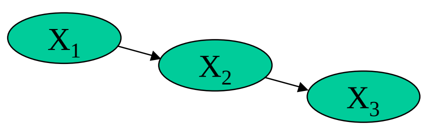
- Forward Sampling
  - Logic Sampling
  - Likelihood Weighting
  - Bounded Variance
  - AA
- Backward Sampling
- Gibbs Sampling (MCMC)
- Metropolis (MCMC)

Why Simulation?

Can solve problems that don't have closed-form solutions

For prediction problems, estimate accuracy increases as $\frac{1}{\sqrt{n}}$

Forward Sampling

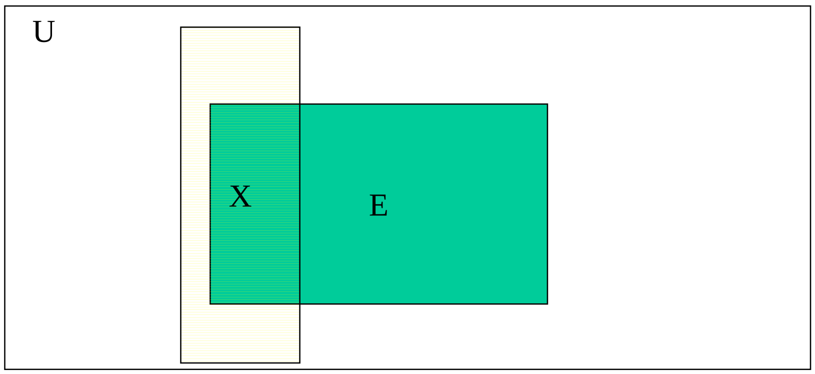


Sample from P{X}:

1. Sample x_1 from $P\{X_1\}$
2. Sample x_2 from $P\{X_2|X_1=x_1\}$
3. Sample x_3 from $P\{X_3| X_1=x_1,X_1=x_1\}$

Rejection Sampling

Estimate P(X|E)



Logic or *Rejection Sampling* (Henrion 87)

Goal: Estimate P{Q=f(X)|E=e}

Let X = unobserved, E = observed
and $Z = X \cup E$

Using Forward Sampling, collect N samples $Z[1],\ldots,Z[N]$

$$P(e) \approx \frac{1}{N}\sum_i I(E=e \mid Z[i]) \qquad P(e) \approx \frac{1}{N}\sum_i I(Q=q, E=e \mid Z[i])$$

$$P(q \mid e) \approx \frac{\sum_i I(Q=q, E=e \mid Z[i])}{\sum_i I(E=e \mid Z[i])}$$

Likelihood Weighting (Shachter+Peot, Fung+Chang)

Goal: Estimate P{Q=f(X)|E=e}

Let X = unobserved, E = observed

Path Prior: $\rho(X) = \prod_X P\{X_i \mid pa(X_i)\}$

Path Likelihood: $\omega(E) = \lambda(E) = \prod_E P\{E_i \mid pa(E_i)\}$

Likelihood Weighting, continued

Forward-sample X from the path prior: X[1], …, X[N]

Score:

$$P(e) \approx \frac{1}{N} \sum_i \omega(X[i])$$

$$P(Q=q,e) \approx \frac{1}{N} \sum_i I(q \mid X[i])\omega(X[i])$$

$$P(Q=q \mid e) \approx \frac{\sum_i I(q \mid X[i])\omega(X[i])}{\sum_i \omega(X[i])}$$

Likelihood Weighting, cont'd

As N -> infinity

$$\sum_i I(q \mid X[i])\omega(X[i]) = \int I(q \mid X)\rho(X)\omega(X)dX$$
$$= \int I(q \mid X)\prod_i P\{X_i \mid pa(X_i)\}\, dX$$
$$= \int I(q \mid X)P\{X, E\}dX$$
$$= P\{q, E\}$$

Importance Sampling

Sample from $\rho'(X)$

Score: $\omega(X) = \frac{\rho(X)}{\rho'(X)}\lambda(X)$

$$\sum_i I(q \mid X[i])\omega(X[i]) = \int I(q \mid X)\rho'(X)\frac{\rho(X)}{\rho'(X)}\lambda(X)dX$$
$$= \int I(q \mid X)\rho(X)\lambda(X)dX$$
$$= P\{q, E\}$$

Likelihood Weighting, Convergence

Say that P(Q=q) has mean μ and variance σ^2.
Chebychev:

$$P\left\{\left|\frac{\hat{\mu}-\mu}{\mu}\right| > \varepsilon\right\} < \frac{c\sigma^2}{N\varepsilon^2\mu^2} \le \delta$$

$$N \ge \frac{c\sigma^2}{\varepsilon^2\mu^2} \cdot \frac{1}{\delta}$$

Generalized {0,1} Estimator Theorem

(Dagum, Karp, Luby, Ross)

let $Z_1, \cdots, Z_N$ denote i.i.d. random variables, distributed according to Z. If $\varepsilon < 1$, and

$$N = \Psi \cdot \frac{k_Z}{(\mu_Z)^2}$$

$$\Psi = 4(e-2)\ln\left(\frac{2}{\delta}\right)\frac{1}{\varepsilon^2}$$

$$k_Z = \max\left\{\sigma_Z^2, \varepsilon\mu_Z\right\}$$

then

$$P\left[(1-\varepsilon)\mu_Z \le \sum_{i=1}^{N}\frac{Z_i}{N} \le (1+\varepsilon)\mu_Z\right] > 1-\delta$$

Likelihood Weighting, better bound

$$N = 4(e-2)\ln\left(\frac{2}{\delta}\right)\frac{1}{\varepsilon^2}\cdot\frac{\max\{\sigma^2,\varepsilon\mu\}}{(\mu)^2}$$

$$N = 4(e-2)\ln\left(\frac{2}{\delta}\right)\frac{1}{\varepsilon^2}\cdot\frac{(1-\mu)}{\mu} \qquad N = 4(e-2)\ln\left(\frac{2}{\delta}\right)\frac{1}{\mu\varepsilon}$$

LVB

Localized variance bound for estimating any joint probability.

Let $u(X_i)$ denote the maximum of $P\{X_i = x_i\}$

and $l(X_i)$ denote the minimum.

$$\Gamma_i = \max\left\{\frac{u(X_i)}{l(X_i)}, \frac{1-l(X_i)}{1-u(X_i)}\right\}$$

$$\Gamma = \prod_i \Gamma_i$$

$$N = 4(e-2)\ln\left(\frac{2}{\delta}\right)\frac{1}{\varepsilon^2}\cdot\frac{1}{\mu} = 4(e-2)\ln\left(\frac{2}{\delta}\right)\frac{1}{\varepsilon^2}\cdot\Gamma$$

Bounded Variance (Dagum+Luby)

Estimate P(x)

1. LW rejects samples because of I(x|X)
2. No good prior bound on μ and σ^2

Solution

- Estimate μ on the fly.
- Stopping rule based on cumulative sum.

Gibbs Sampling

For each observed variable E_i, set $E_i = e_i$.

Use any sampling technique (usually forward-sampling or importance-sampling) to set X_i to some random value.

Repeat

- Pick some unobserved variable X_i
- Sample $X_i[j] \sim P\{X_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$

Question, what is $P\{X_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$?

Procedures

Parallel:

- "Burn-in" by Gibbs for some period of time. Sample.
- Sample is in P{X|E}

Serial:

- If x[i] is from P(X|E), then, so is x[i+1].
- Say that x_1 is flipped:

$$P\{X[i+1]\} = \sum_{x_1} P\{x'_1 \mid x_2,...,x_n\} P\{x_1, x_2,...,x_n\}$$
$$= \sum_{x_1} P\{x'_1 \mid x_2,...,x_n\} P\{x_1 \mid x_2,...,x_n\} P\{x_2,...,x_n\}$$
$$= P\{x'_1 \mid x_2,...,x_n\} P\{x_2,...,x_n\} \sum_{x_i} P\{x_1 \mid x_2,...,x_n\}$$
$$= P\{x'_1, x_2,...,x_n\}$$

Pluses/Minuses

Brooks (98):

- (-) No good convergence result known for Gibbs
- (+) Includes evidence in sampling.