

Probabilistic Graphical Models

Monte-Carlo Inference

Siamak Ravanbakhsh

Winter 2021

Learning objectives

- the relationship between **sampling and inference**
- sampling from **univariate distributions**
- Monte Carlo sampling in **graphical models**

Monte Carlo inference

- calculating marginals $p(x_1 = \bar{x}_1) = \sum_{x_2, \dots, x_n} p(\bar{x}_1, x_2, \dots, x_n)$

Mote Carlo inference

- calculating marginals $p(x_1 = \bar{x}_1) = \sum_{x_2, \dots, x_n} p(\bar{x}_1, x_2, \dots, x_n)$
- approximate it by sampling $X^{(l)} \sim p(x)$

$$p(x_1 = \bar{x}_1) \approx \frac{1}{L} \sum_l \mathbb{I}(X_1^{(l)} = \bar{x}_1)$$

Mote Carlo inference

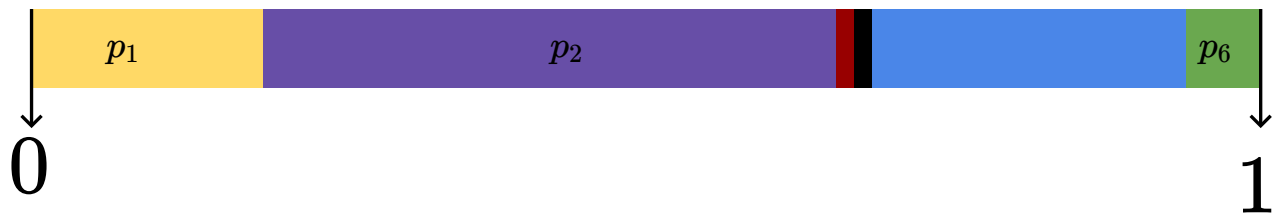
- calculating marginals $p(x_1 = \bar{x}_1) = \sum_{x_2, \dots, x_n} p(\bar{x}_1, x_2, \dots, x_n)$
- approximate it by sampling $X^{(l)} \sim p(x)$

$$p(x_1 = \bar{x}_1) \approx \frac{1}{L} \sum_l \mathbb{I}(X_1^{(l)} = \bar{x}_1)$$

- inference in exponential family $p_\theta(x) = \exp(\theta^\top \psi - A(\theta))$
 - is about finding the mean parameters $\mu = \mathbb{E}_{p_\theta}[\psi(x)]$
 - using L samples (particles) $\mu \approx \frac{1}{L} \sum_l \psi(X^{(l)})$

Sampling from **categorical** dist.

- access to *pseudo* random number generator for $X \sim U(0, 1)$
- given $p(X = d) = p_d \quad \forall 1 \leq d \leq D$



- generate $X \sim U(0, 1)$ and see where it falls
use binary search $\mathcal{O}(\log(D))$

Transforming probability densities

- given a random variable $X \sim p_X$
- what is the prob. density of $Y = \phi(X)$?

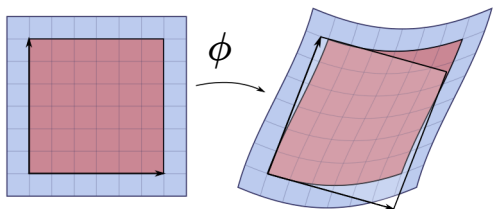
$$Y \sim p_Y(y) = p_X(\phi^{-1}(y)) \left| \frac{d\phi^{-1}(y)}{dy} \right|$$

corresponding x

how ϕ changes the volume around each point y

in multivariate case:

- determinant of the Jacobian matrix

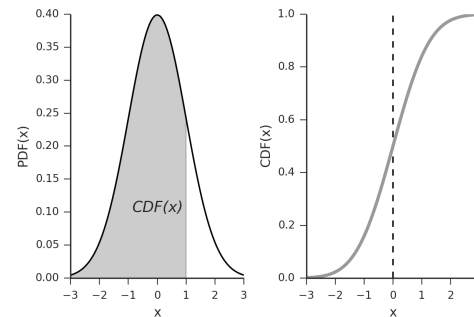


Inverse transform sampling

- let X be uniform $p_X = U(0, 1)$
- given a density p_Y

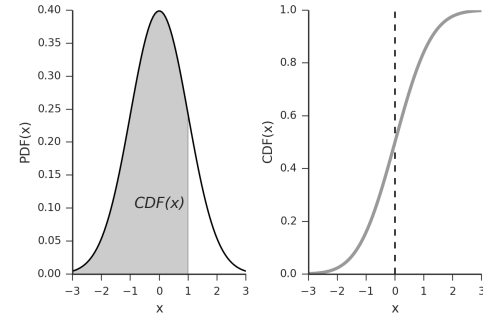
Inverse transform sampling

- let X be uniform $p_X = U(0, 1)$
- given a density p_Y
- let F_Y be its CDF $F_Y(y) = P(Y < y)$



Inverse transform sampling

- let X be uniform $p_X = U(0, 1)$
- given a density p_Y
- let F_Y be its CDF $F_Y(y) = P(Y < y)$
- transform X using $\phi(X) = F_Y^{-1}(X)$
- what is the density of $Y = \phi(X)$?

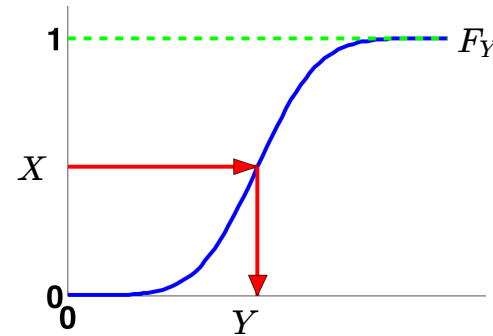
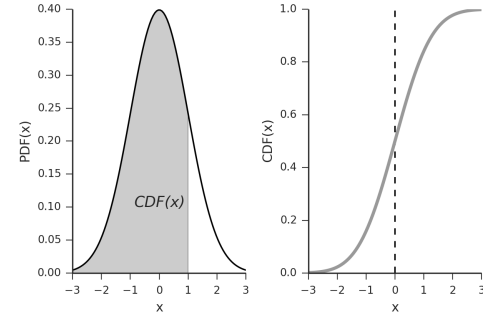


Inverse transform sampling

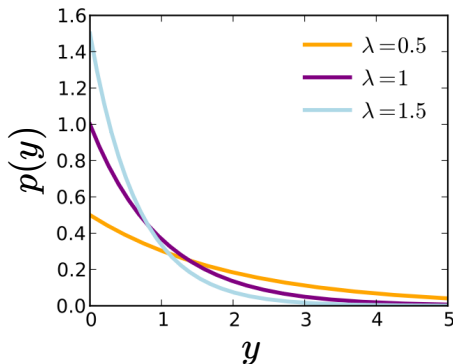
- let X be uniform $p_X = U(0, 1)$
- given a density p_Y
- let F_Y be its CDF $F_Y(y) = P(Y < y)$
- transform X using $\phi(X) = F_Y^{-1}(X)$
- what is the density of $Y = \phi(X)$?

$$Y \sim p_X(\phi^{-1}(y)) \left| \frac{d\phi^{-1}(y)}{dy} \right| = p_X(F_Y(y)) \left| \frac{dF_Y(y)}{dy} \right|$$

constant:
 $p_X = U(0, 1)$ $p_Y(y)$



Inverse transform sampling: **example**



Exponential distribution

$$p(y) = \lambda e^{-\lambda y}$$

$$F_Y(y) = 1 - e^{-\lambda y}$$

calculate the **inverse CDF**:

$$F_Y^{-1}(x) = -\frac{1}{\lambda} \ln(1 - x)$$

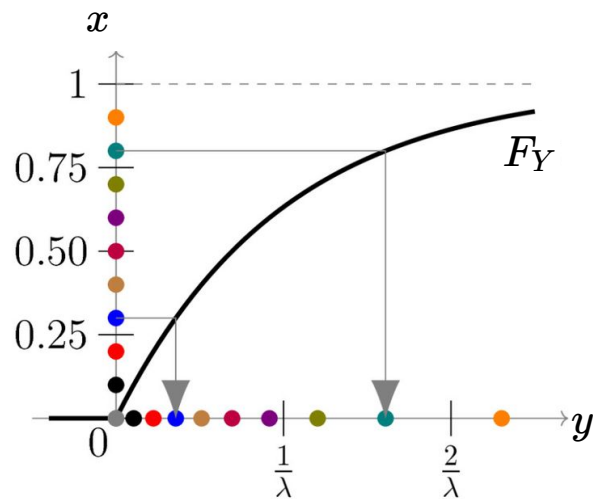
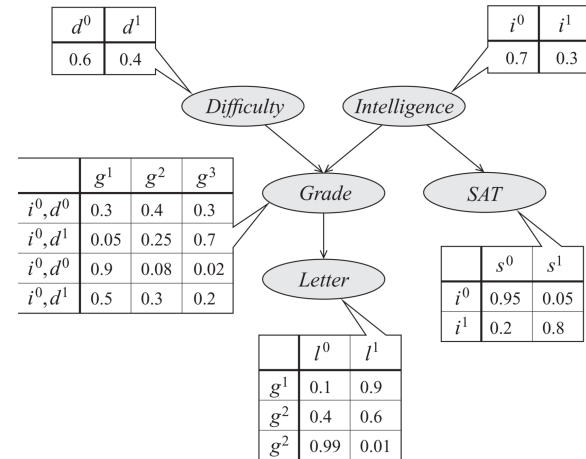


image:wikipedia

Sampling in graphical models

ancestral sampling for Bayes-nets

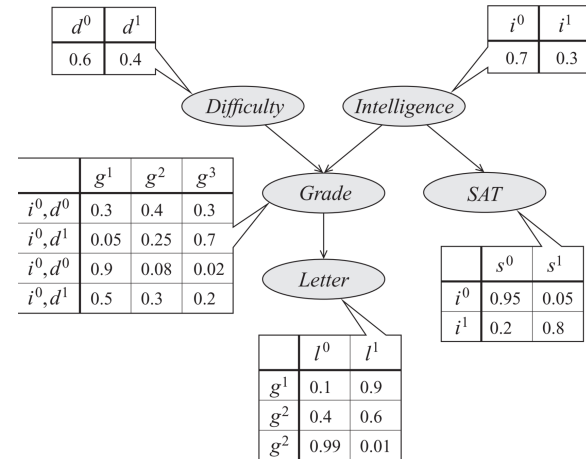


Sampling in graphical models

ancestral sampling for Bayes-nets

- find a *topological ordering*
 - e.g., D,I,G,S,L or I,S,D,G,L
- sample by conditioning on parents

$$G \sim P(g \mid I, D)$$

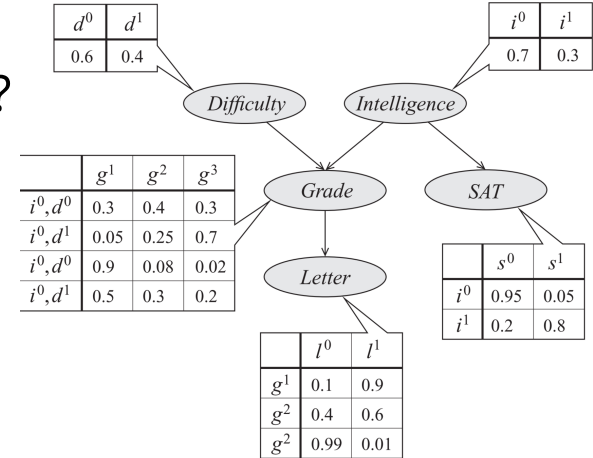


Introducing evidence

what if we have an evidence

- E.g., how to sample from the posterior?

$$p(D, I, S, L | G = g^0)$$



Introducing evidence

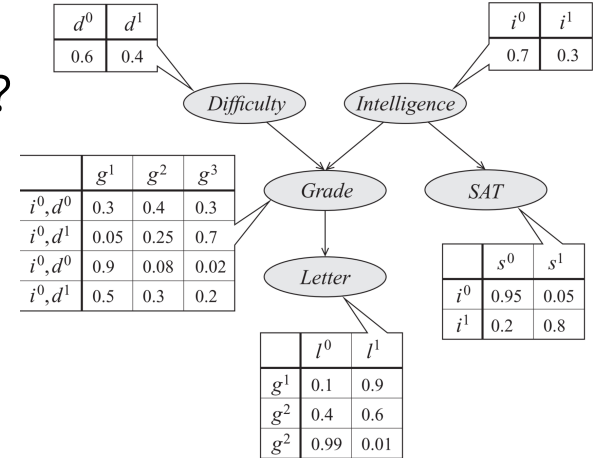
what if we have an evidence

- E.g., how to sample from the posterior?

$$p(D, I, S, L | G = g^0)$$

rejection sampling

- find a topological ordering
- sample by conditioning on parents
- only keep samples **compatible with evidence** ($G = g^0$)
 - wasteful if evidence has a low probability



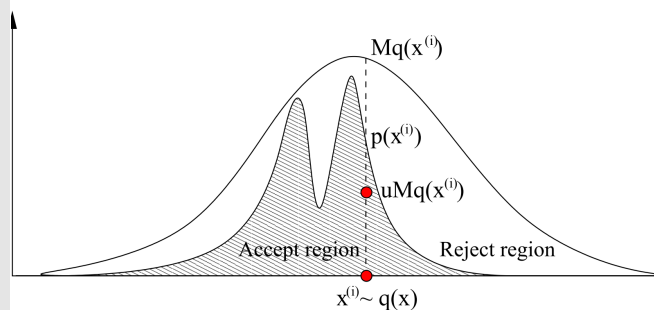
Rejection sampling

general form

to sample from $p(x) = \frac{1}{Z}\tilde{p}(x)$

use a **proposal** distribution $q(x)$
such that $Mq(x) > \tilde{p}(x)$ everywhere
sample $X \sim q(x)$

accept the sample with probability $\frac{\tilde{p}(x)}{Mq(x)}$



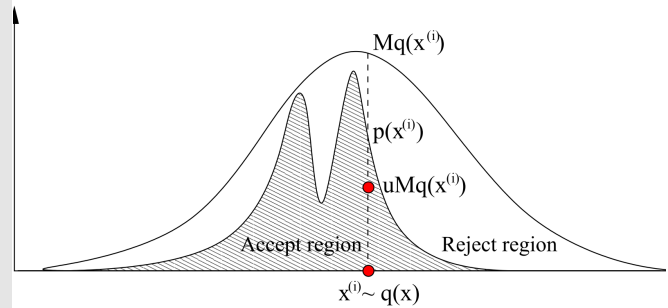
Rejection sampling

general form

to sample from $p(x) = \frac{1}{Z}\tilde{p}(x)$

use a **proposal** distribution $q(x)$
such that $Mq(x) > \tilde{p}(x)$ everywhere
sample $X \sim q(x)$

accept the sample with probability $\frac{\tilde{p}(x)}{Mq(x)}$



what is the probability of acceptance? $\int_x q(x) \frac{\tilde{p}(x)}{Mq(x)} dx = \frac{Z}{M}$

for high-dimensional dists. $\frac{Z}{M}$ becomes small!

- rejection sampling becomes wasteful

image: Murphy's book

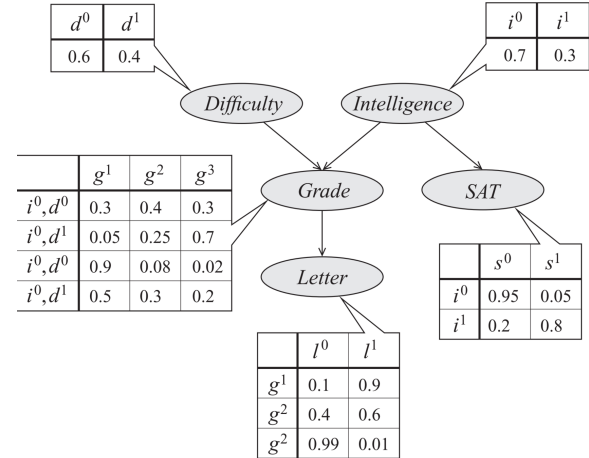
Likelihood weighting

what if we have an evidence?

- E.g., how to sample from the posterior?

$$p(D, I, S, L | G = g^1)$$

- find a topological ordering
- assign a weight to each particle $w^{(l)} \leftarrow 1$
- sample by conditioning on parents
- when sampling an **observed variable**
 - set it to its observed value $G = g^1$
 - update the sample's weight $w^{(l)} \leftarrow w^{(l)} \times p(G = g^1 | D = d^{(l)}, I = i^{(l)})$



current assignments to parents

Likelihood weighting

what if we have an evidence?

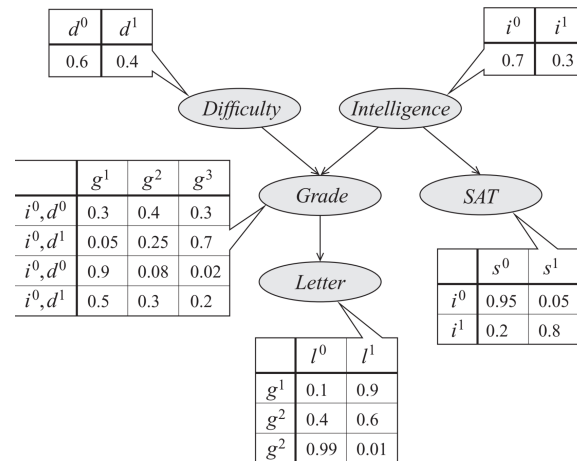
- e.g., how to sample from the posterior?

$$p(D, I, S, L | G = g^0)$$

now, we have **weighted particles**

answering inference queries:

$$p(S = s^0 | G = g^1) = \frac{\sum_l w_l \mathbb{I}(S^{(l)} = s^0)}{\sum_l w_l}$$



Likelihood weighting

what if we have an evidence?

- e.g., how to sample from the posterior?

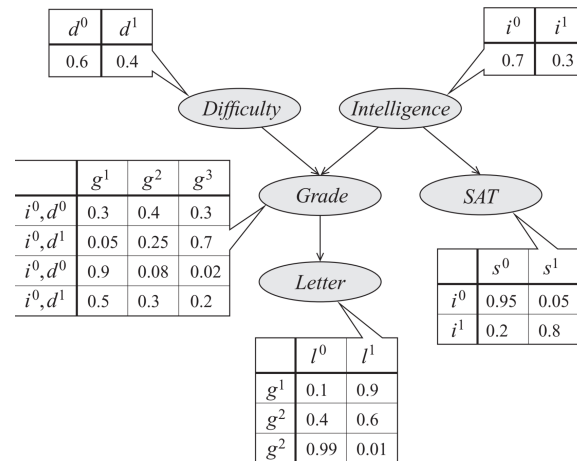
$$p(D, I, S, L | G = g^0)$$

now, we have **weighted particles**

answering inference queries:

$$p(S = s^0 | G = g^1) = \frac{\sum_l w_l \mathbb{I}(S^{(l)} = s^0)}{\sum_l w_l}$$

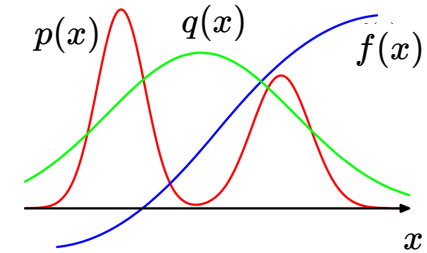
special case of **importance sampling**



Unnormalized importance sampling

Objective: Monte Carlo estimate $\mathbb{E}_p[f(x)]$

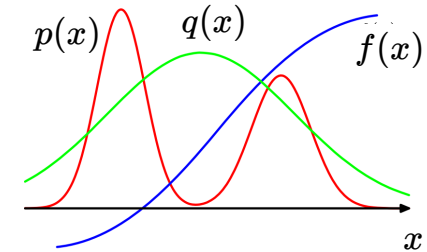
- difficult to sample from p (yet easy to evaluate)
- use a proposal distribution $q : p(x) > 0 \Rightarrow q(x) > 0$



Unnormalized importance sampling

Objective: Monte Carlo estimate $\mathbb{E}_p[f(x)]$

- difficult to sample from p (yet easy to evaluate)
- use a proposal distribution $q : p(x) > 0 \Rightarrow q(x) > 0$

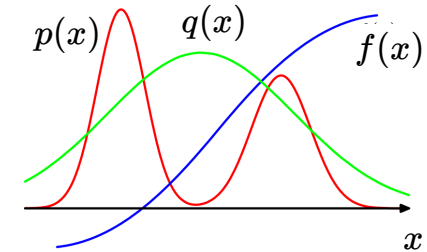


$$\text{since } \mathbb{E}_p[f(x)] = \int_x p(x) f(x) dx = \int_x q(x) \frac{p(x)}{q(x)} f(x) dx = \mathbb{E}_q\left[\frac{p(x)}{q(x)} f(x)\right]$$

Unnormalized importance sampling

Objective: Monte Carlo estimate $\mathbb{E}_p[f(x)]$

- difficult to sample from p (yet easy to evaluate)
- use a proposal distribution $q : p(x) > 0 \Rightarrow q(x) > 0$



$$\text{since } \mathbb{E}_p[f(x)] = \int_x p(x) f(x) dx = \int_x q(x) \frac{p(x)}{q(x)} f(x) dx = \mathbb{E}_q\left[\frac{p(x)}{q(x)} f(x)\right]$$

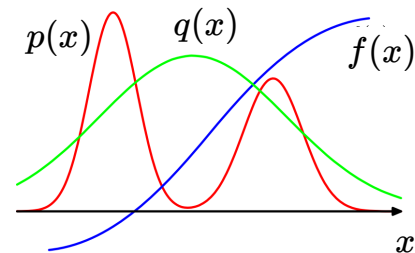
sample $X^l \sim q(x)$

assign an importance sampling weight $w(X^{(l)}) = \frac{p(X^{(l)})}{q(X^{(l)})}$

Unnormalized importance sampling

Objective: Monte Carlo estimate $\mathbb{E}_p[f(x)]$

- difficult to sample from p (yet easy to evaluate)
- use a proposal distribution $q : p(x) > 0 \Rightarrow q(x) > 0$



$$\text{since } \mathbb{E}_p[f(x)] = \int_x p(x) f(x) dx = \int_x q(x) \frac{p(x)}{q(x)} f(x) dx = \mathbb{E}_q\left[\frac{p(x)}{q(x)} f(x)\right]$$

sample $X^l \sim q(x)$

assign an importance sampling weight $w(X^{(l)}) = \frac{p(X^{(l)})}{q(X^{(l)})}$

$$\mathbb{E}_p[f(x)] \approx \frac{1}{L} \sum_l w(X^{(l)}) f(X^{(l)}) \text{ is an unbiased estimator}$$

can be more efficient than sampling from p itself! (why?)

normalized importance sampling

What if we can evaluate p , up to a constant? $p(x) = \frac{1}{Z} \tilde{p}(x)$

Examples

- posterior in directed models $p(x | E = e) = \frac{1}{p(e)} p(x, e)$
- prior in undirected models $p(x) = \frac{1}{Z} \prod_I \phi_I(x_I)$

normalized importance sampling

What if we can evaluate p , up to a constant? $p(x) = \frac{1}{Z}\tilde{p}(x)$

Examples posterior in directed models $p(x | E = e) = \frac{1}{p(e)}p(x, e)$

prior in undirected models $p(x) = \frac{1}{Z} \prod_I \phi_I(x_I)$

define $w(x) = \frac{\tilde{p}(x)}{q(x)}$ then $\mathbb{E}_q[w(x)] = \int_x \tilde{p}(x) dx = Z$

since $\mathbb{E}_p[f(x)] = \int_x p(x)f(x)dx = \frac{1}{Z} \int_x q(x) \frac{\tilde{p}(x)}{q(x)} f(x)dx = \frac{1}{Z} \mathbb{E}_q[w(x)f(x)] = \frac{\mathbb{E}_q[w(x)f(x)]}{\mathbb{E}_q[w(x)]}$

normalized importance sampling

What if we can evaluate p , up to a constant? $p(x) = \frac{1}{Z} \tilde{p}(x)$

Examples

posterior in directed models $p(x | E = e) = \frac{1}{p(e)} p(x, e)$

prior in undirected models $p(x) = \frac{1}{Z} \prod_I \phi_I(x_I)$

define $w(x) = \frac{\tilde{p}(x)}{q(x)}$ then $\mathbb{E}_q[w(x)] = \int_x \tilde{p}(x) dx = Z$

since $\mathbb{E}_p[f(x)] = \int_x p(x) f(x) dx = \frac{1}{Z} \int_x q(x) \frac{\tilde{p}(x)}{q(x)} f(x) dx = \frac{1}{Z} \mathbb{E}_q[w(x) f(x)] = \frac{\mathbb{E}_q[w(x) f(x)]}{\mathbb{E}_q[w(x)]}$

sample $X^{(l)} \sim q(x)$

assign an importance sampling weight $w(X^{(l)}) = \frac{\tilde{p}(X^{(l)})}{q(X^{(l)})}$

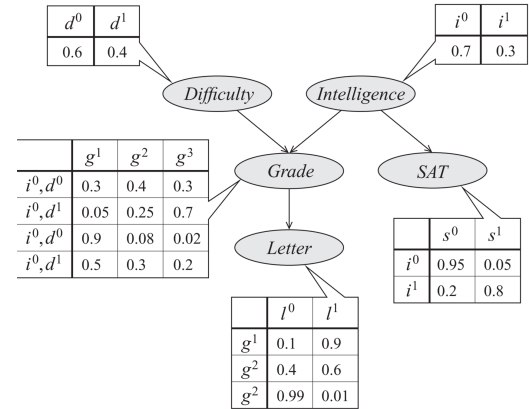
$\mathbb{E}_p[f(x)] \approx \frac{\sum_l w(X^{(l)}) f(X^{(l)})}{\sum_l w(X^{(l)})}$ is a biased estimator (e.g., consider $L=1$)

Revisiting likelihood weighting

likelihood weighting:

$$p(S = s^0 \mid G = g^2, I = i^1) = \frac{\sum_l w_l \mathbb{I}(S^{(l)} = s^0)}{\sum_l w_l}$$

equivalent to:



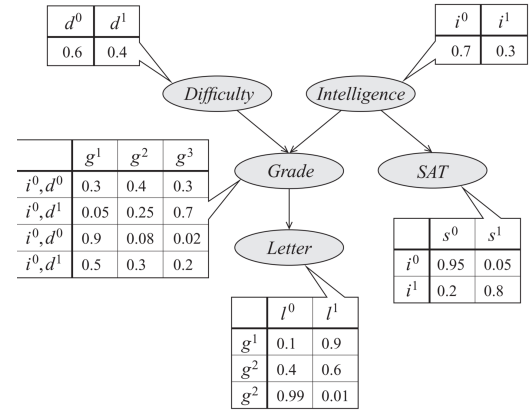
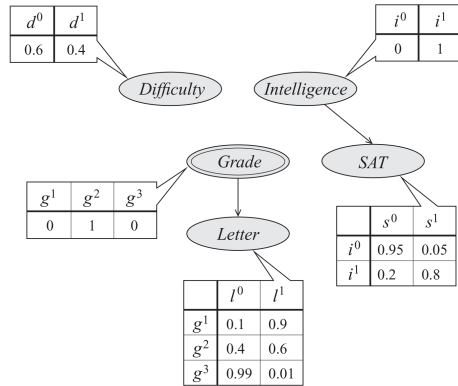
Revisiting likelihood weighting

likelihood weighting:

$$p(S = s^0 \mid G = g^2, I = i^1) = \frac{\sum_l w_l \mathbb{I}(S^{(l)} = s^0)}{\sum_l w_l}$$

equivalent to:

mutilated Bayes-net as **proposal q**



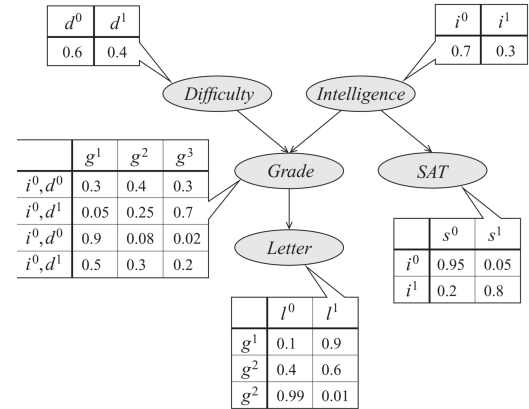
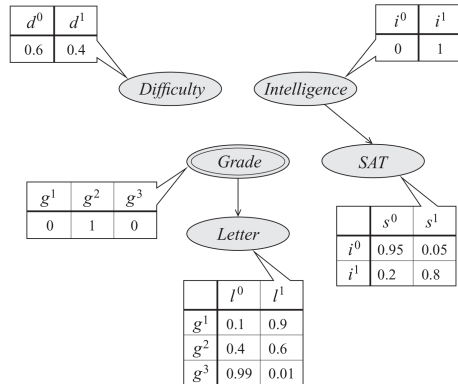
Revisiting likelihood weighting

likelihood weighting:

$$p(S = s^0 \mid G = g^2, I = i^1) = \frac{\sum_l w_l \mathbb{I}(S^{(l)} = s^0)}{\sum_l w_l}$$

equivalent to:

mutilated Bayes-net as **proposal q**



$$w_l = \frac{\tilde{p}(X)}{q(X)} = p(G = g^2 \mid I = i^{(l)}, D = d^{(l)}) \times P(I = i^1)$$

similar to initial algorithm for likelihood weighting

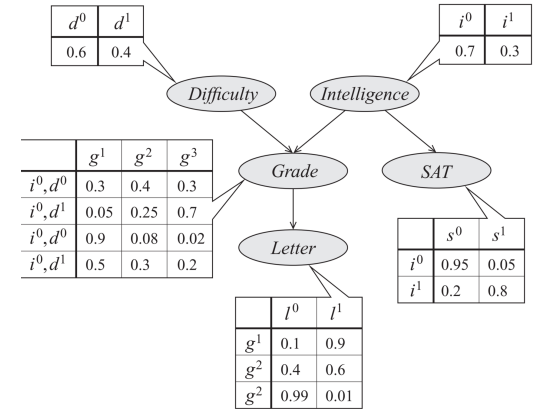
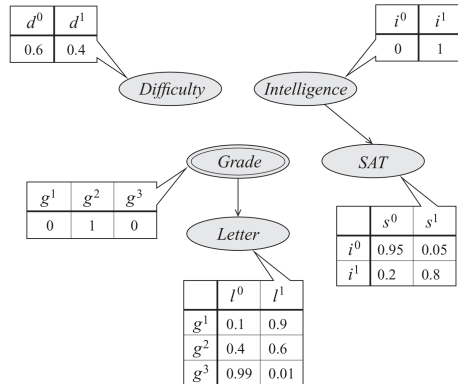
Revisiting likelihood weighting

likelihood weighting:

$$p(S = s^0 \mid G = g^2, I = i^1) = \frac{\sum_l w_l \mathbb{I}(S^{(l)} = s^0)}{\sum_l w_l}$$

equivalent to:

mutilated Bayes-net as **proposal q**



$$w_l = \frac{\tilde{p}(X)}{q(X)} = p(G = g^2 \mid I = i^{(l)}, D = d^{(l)}) \times P(I = i^1)$$

similar to initial algorithm for likelihood weighting

- evidence only affects sampling for the descendants
- what if all evidence appears at leaf nodes?

Summary

Monte-carlo sampling for approximate inference:

- sampling from univariates:
 - categorical distribution
 - inverse transform sampling
- marginals in directed models:
 - ancestral sampling
- more sophisticated: (incorporating evidence)
 - rejection sampling
 - importance sampling (likelihood weighting)