Forward Sampling

Sargur Srihari
srihari@buffalo.edu

Topics

- Forward Sampling
- Sampling from a Bayesian Network
- Sampling from a Discrete Distribution
- Use of samples
- Analysis of Error
- Conditional Probability Queries

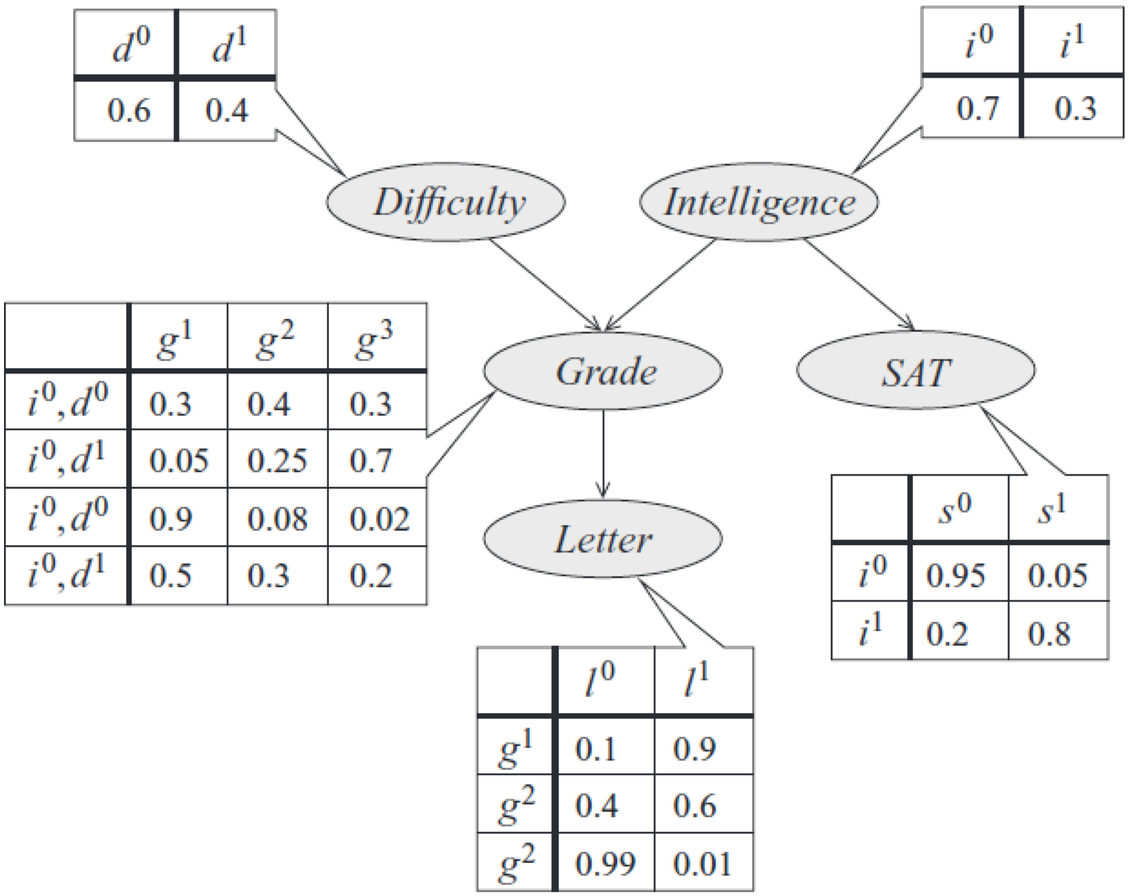
Forward Sampling: Plan of Discussion

- Forward sampling
  - It is the simplest particle generation approach
  - We generate samples $\xi[1],..,\xi[M]$ from $P(\chi)$
- Plan of discussion
  - How we can easily generate particles from $P_{\mathcal{B}}(\chi)$ by sampling from a Bayesian network
  - No of particles needed to get a good approximation of the expectation of a target function f
  - Difficulties in generating samples from posterior $P_{\mathcal{B}}(\chi \mid e)$
    - In undirected models even generating a sample from the prior distribution is a difficult task

Sampling from a Bayesian Network

- A very simple process
- Sample nodes in some order so that by the time we sample a node we have values for all of its parents
- We can then sample from the distribution specified by the CPD
- Need ability to sample from the distributions underlying the CPD
  - Straightforward for discrete case
  - Subtler for continuous measures

Forward Sampling example



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^0, d^0	0.9	0.08	0.02
i^0, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

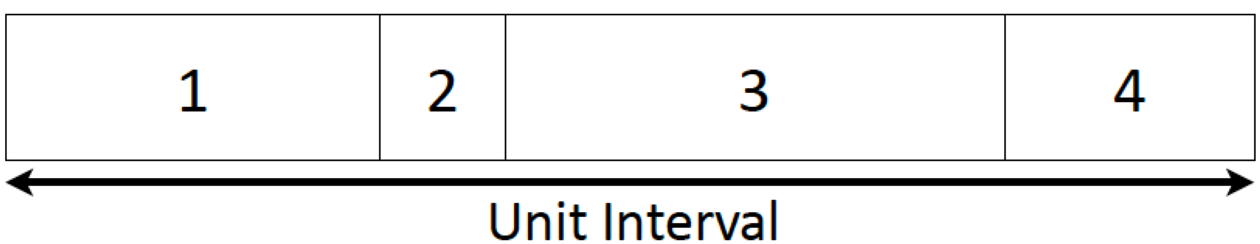
	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^2	0.99	0.01

1) sample D {.6 vs .4} -> d^0
2) sample I {.7 vs .3} -> i^0
3) sample G
   -> $P(g^1)=0.3$
   $P(g^2)=0.4$
   $P(g^3)=0.3$

etc. etc.

Sampling from a Discrete Distribution

- Split $[0,1]$ interval into bins whose sizes are determined by the probabilities $P(x^i)$, $i=1,..,k$
  - Partition the interval into k subintervals
- Generate a sample s uniformly from the interval
- If s is in the i^{th} interval then sampled value is x^i
- Example: sample from $P(x=\{1,2,3,4\})=\{0.3,0.1,0.4,0.2\}$



Use of Samples

- Using basic convergence bounds, we know that from a set of particles $\mathcal{D}=\{\xi[1]),\ldots,(\xi[M]\}$ generated via this sampling process we can estimate the expectation of any function as

$$\hat{E}_{D}(f)=\frac{1}{M}\sum_{m=1}^{M} f(\xi[m])$$

- In the case where our task is to compute $P(\boldsymbol{y})$, this estimate is simply the fraction of particles where we have seen the event $\boldsymbol{y}$

$$\hat{P}_{D}(\boldsymbol{y})=\frac{1}{M}\sum_{m=1}^{M} I\{\boldsymbol{y}[m]=\boldsymbol{y}\}$$

Analysis of Error

- Quality of the estimate depends heavily on the no. of particles generated
- How many particles are needed to obtain a certain performance guarantee?
- Focus analysis on the case where we need to estimate $P(\boldsymbol{y})$

How many samples are required?

1. From Hoeffding bound
   - Error is bounded by ε with probability of at least $1-\delta$
   - The estimator with (ε,δ) reliability is $\boxed{M \geq \frac{\ln(2/\delta)}{2\varepsilon^2}}$

2. From Chernoff bound
   - The no of samples needed to guarantee a certain error probability δ is $\boxed{M \geq 3\frac{\ln(2/\delta)}{P(y)\varepsilon^2}}$
   - Thus no. of required samples grows inversely with the probability $P(\boldsymbol{y})$

Conditional Probability Queries

- So far we have discussed the problem of estimating marginal probabilities
  - i.e., Probability of event $\boldsymbol{Y}=\boldsymbol{y}$ relative to the original joint distribution
- In general we are interested in conditional distributions of the form $p(\boldsymbol{y}|\boldsymbol{E}=\boldsymbol{e})$
- This estimation task is significantly harder
- One approach is called Rejection sampling

Rejection Sampling

- Generate samples from posterior $P(\boldsymbol{\chi}|\boldsymbol{e})$
  - Can do this by generating samples $\boldsymbol{x}$ from $P(\mathbf{X})$
  - Reject any sample that is not compatible with e
  - Resulting samples are from posterior $P(\boldsymbol{\chi}|\boldsymbol{e})$
- Problem: no of unrejected particles is small
  - Expected no of unrejected from original M is $MP(\boldsymbol{e})$
    - Ex: If $P(\mathbf{e})=0.001$ & $M=10{,}000$, only 10 unrejected samples
  - Small probabilities are rule rather than exception
    - Any set of symptoms has a low probability
    - As no of variables k increases, probability of evidence k decreases exponentially with k