Announcements

• All exams are ready, regrades, conflicts
• You may pick up exams after class today
• Do not pick up your exam today IF
  – you have not yet seen your exam
  – and you might want a regrade
  – (then email me with your available times for a meeting)
Approximate Inference in Non-Polytrees

• Monte Carlo sampling algorithms
  – Generate examples from the conditional distribution
  – Conditionalize on the evidence
  – Estimate the distribution by counting

• Principled use of randomization

• Construct a generator of samples from the desired distribution and count appropriately

• Often this is far easier than trying to construct the conditional distribution explicitly
Conditional Independence among Variables in a Bayesian Net

A Variable is conditionally independent of all other nodes in the network given its parents, children, and children’s parents (co-parents). This is its Markov blanket.

A Variable is conditionally independent of its non-descendants given its parents.

(Text Figure 14.4)

A set of Variables X is conditionally independent of a set of Variables Y given a set of Evidence Variables E if all paths connecting an x to a y are “d-separated”
X is conditionally independent of everything else given its Markov blanket (used in Gibbs sampling)
X is conditionally independent of its non-descendants given its parents (used in rejection and likelihood sampling)
d-separation
Monte Carlo Algorithms I
Rejection Sampling

• Start w/ all RV’s unknown
• Choose a *topological* ordering (parents before children)
• Get random numbers to assign values
• Descendents are always unknown when assigning a RV
• Generate lots of samples
• Count, but ignore (reject) samples violating evidence
• Can require lots of samples; when?
Monte Carlo algorithms II
Likelihood Weighting

• Generate a sample randomly, assigning values topologically as before
• Except at evidence nodes luckily you get the evidence value
  – how “lucky” was this?
  – compute $Pr(\text{evidence} \mid \text{parents})$
  – this value and its frequency reflects previous random numbers
• Generate lots of samples
• Reject NOTHING. All samples are consistent w/ evidence!
• But count each fractionally, weighted by its total likelihood (product of the “lucky” evidence assignments)
• Some may count for very little
Monte Carlo algorithms III
Gibbs Sampling

• Gibbs: simple version of Markov Chain Monte Carlo (MCMC)
• MCMC is a general, important, and simple algorithm
• Why it works is somewhat involved
• Fix evidence variables to their values
• Initialize others randomly
• Iterate through nodes (must visit all many times)
• Resample each non-evidence variable given its Markov blanket
• Each change generates a (slightly different) example (with values assigned to each random variable)
• After ‘burn in,’ accumulate each k’th sample
• Count
Inference (the big picture)

• Learning = building a model
• Inference = using the model
• Statistical inference
  – Prior distribution (learned)
  – Evidence (specific observations)
  – Posterior distribution (Prior adjusted to reflect evidence)
• Compare w/ Logical Inference / Reasoning
Reasoning vs. Statistical Inference

• An Amazon customer C orders diapers in bulk, fancy espresso machine,...
• What else might we tempt C to buy (limited screen space, C’s attention,...)
• Reasoning (with lots of prior symbolic knowledge)
  – Conclude C is well-off, refined, self-indulgent, but thrifty
  – Espresso machines last for years so not another of those
  – But to complement, maybe a burr coffee grinder, upscale whole bean coffee, etc.
  – Diapers get used up, so maybe more diapers but not right away
    • Try some lower price some higher quality
    • Monitor click through & ordering decisions
  – More of a reach: on-sale high end audio system, wine-of-the-month club, etc.
Reasoning vs. Statistical Inference

• An Amazon customer C orders diapers in bulk, fancy espresso machine,...
• What else might we tempt C to buy (limited screen space, C’s attention,...)
• Statistics (with lots of prior training examples)
  – Build a (large & deep) probability distribution
    • RVs: purchases and next purchases
    • Parameters estimated from copious data
  – Conditionalize on C’s purchases
  – Suggest a few next purchases ordered by conditional probability
  – With a little knowledge:
    • BN w/ unobservable RVs for how wealthy, how thrifty, ...
    • Optimize parameters to fit data
    • Improves generalization significantly
We saw how the parameters are set. Where does the BN structure come from?

- Make it up
  - (elicit it from experts)
- Learn from data
  - Works well only for special classes of nets
  - Can require too much data
  - Can be ill-posed
BN Construction by Hand

• Identify variables
• Order them*
• While there are variables to add
  – Pick the next in the ordering
  – Identify its parents in the net
    • Hold all others constant (in every configuration)
    • If net variable influences it, net var is a parent
  – Draw all arcs and add CPT

* order can matter a lot
Heuristic

Usually the most compact representation results when direct influence mirrors physical causality

So order causes before effects
Dentist Example

3 Boolean Random Variables:

C – Patient has a cavity
A – Patient reports a toothache
B – Dentist’s probe catches on tooth
Effect of Order on BN structure

Order: C A B

Order: A B C

Order: B C A
Effect of Order on BN structure

Order: C A B

Order: A B C

Order: B C A

5 parameters

7 parameters
no savings over
the joint

5 parameters
BN Structure
consider ordering M,J,E,B,A

B – a burglary is in progress
E – an earthquake is in progress
A – the alarm is sounding
J – John calls
M – Mary calls

All variables are Boolean
How many numbers?

No savings over the Joint
What conditional independence assumptions does it make?
Ordering is Important

• Original net B E A J M
  – 10 parameters
  – Saves 21 over the joint
• Ordering M J E B A
  – 31 parameters
  – No savings over the joint
• “Causality” ordering is only heuristic
• Suppose we want to test all orderings of n random variables...