Learning Bayesian Networks

- Objectives
  - ❑ Acquisition of BNs
  - ❑ Technical context of BN learning
  - ❑ Criterion of sound structure learning
  - ❑ BN structure learning in 2 steps
  - ❑ BN CPT estimation
- Reference
  - ❑ R.E. Neapolitan: Learning Bayesian Networks (2004)

Acquisition of BNs

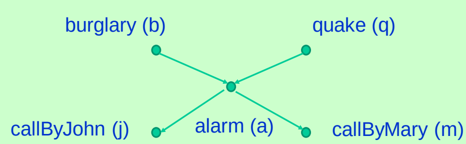
- Elicitation based acquisition
  - ❑ Determine the set V of env variables and their domains.
  - ❑ Determine the graphical dependence structure.
  - ❑ Determine CPTs one for each variable.
  - ❑ Time consuming for domain experts & agent developer.
- Learning based acquisition
  - ❑ Input: a training set *R* of examples in an application env
  - ❑ Output: a BN for inference about the env
  - ❑ The focus of study

Task Decomposition

- Denote a BN by S = (V, G, Pb), where V is a set of env variables, G = (V, E) is a DAG, and Pb is a set of CPTs: Pb = $\{P(v|\pi(v))|v\in V\}$.
- Task: Learning a BN from a training set R
- Subtasks
  - ❑ Identification of V
  - ❑ Definition of variable domains
  - ❑ Construction of dependency structure G
    - Referred to as structure learning
  - ❑ Estimation of CPTs

Review on BN Semantics

1. A variable v in BN is conditionally independent of its non-descendants given its parents π(v).
2. Variables x and y are dependent given their common descendant(s).
   - ❑ Ex Burglar-quake



Technical Context of BN Learning

- Env is characterized by a unknown full joint P*(V).
- A unknown BN S* = (V, G*, Pb*) encodes exactly the same conditional independencies of P*(V) by G*.
- A set R of training examples are obtained from the env (i.e., P*(V)) by independent trials.
- Task: From R, learn a BN S = (V, G, Pb) that models P*(V) as accurately and as concisely as possible.

Structure Markov Equivalence

- Suppose a full joint P*(V) over env V can be encoded by a BN S = (V, G, Pb).
  - Is DAG structure G unique?
- Ex A P*(V) can be encoded by a BN with G below:
  G: child_age → foot_size → shoe_size
- Two DAGs are Markov equivalent if they entail the same conditional independencies.
- Ex G”: child_age → foot_size ← shoe_size
  - Are G” and G Markov equivalent?

Criterion of Sound Structure Learning

- Let S = (V,G,Pb) and S’ = (V,G’,Pb’) be BNs s.t.
  a) G and G’ are Markov equivalent, and
  b) Pb and Pb’ are derived from the same env.

  Then S and S’ model the same full joint over V.
- Ex Env V = {a, b} with true full joint P*(V):
  G_1: a→b; G_2: a←b; G_3 is disconnected.

a	b	P*(a,b)
t	t	0.05
t	f	0.35
f	t	0.50
f	f	0.10

- If full joint P*(V) can be encoded by BN S = (V,G,Pb) but S’ = (V,G’,Pb’) is learned, then structure learning is sound if G and G’ are Markov equivalent.

Learning DAG Skeleton

- Let G = (V, E) be a DAG. The undirected graph G’ = (V, E’), where E’ is obtained by removing direction of each link in E, is the skeleton of G.
- Structure learning can be performed in two steps:
  - Learning a skeleton G’ = (V, E’).
  - Direct links in G’ to obtain DAG G = (V, E).
- In skeleton learning, how do we know whether a pair of variables should be adjacent?
- [Theorem] Variables x, y $\in$ V are adjacent in G’ iff there exists no $Z \subset V \setminus \{x,y\}$ s.t. I(x, Z, y) holds.

Entropy

- In inductive learning, we measured amount of info contained in the value of a variable. That measure is known as entropy.
- Let X be a set of variables with JPD P(X). The entropy of X is

$$H(X) = - \Sigma_{\underline{x}} P(\underline{x}) \log_2(P(\underline{x})).$$

- Interpretations
  - ❑ H(X) is the amount of info in an assignment $\underline{x}$ of X.
  - ❑ H(X) is the measure of uncertainty associated with X.

How to Determine I(x,Z,y)?

- [Theorem] For variables x, y∈V and Z⊂V, the following holds:

$$I(x,Z,y) \Leftrightarrow H(x,Z,y) = H(x,Z)+H(Z,y)-H(Z).$$

- Algorithm to test I(x,Z,y) using training set r

```
estimate P(x,Z,y) from r;
marginalize P(x,Z,y) to obtain P(x,Z), P(Z,y) and P(Z);
compute H(x,Z,y), H(x,Z), H(Z,y), and H(Z);
compute diff = |H(x,Z,y) – (H(x,Z)+H(Z,y)-H(Z))|;
if diff < threshold, return I(x,Z,y);
else return ¬I(x,Z,y);
```


How to Choose Z to Test I(x,Z,y)?

- Given x,y∈V, Z⊂V\{x,y}, & ¬I(x, Z, y), it is possible
  - ❑ I(x, Z^-, y) for some $Z^- \subset Z$, or
  - ❑ I(x, Z^+, y) for some $Z^+ \supset Z$.
- Implication: In the worst case, to determine I(x,Z,y), all subsets of V\{x,y} must be tested.
- [Theorem] In a BN, x,y∈V, Z⊂V\{x,y}, and I(x,Z,y). Then either I(x,π(x),y) or I(x,π(y),y) holds.
- Idea a) To find Z s.t. I(x,Z,y), it suffices to limit search to Z ⊂ Adj(x) and Z ⊂ Adj(y).
- Idea b) Test smaller subsets Z first.

Structure Learning Algorithm

```
learnBnDag(V, R) {
  G’ = complete undirected graph over V;
  for each link <x,y> in G’,
      associated <x,y> with a set Sxy = null;
  G’  = getSkeleton(G’, R);
  G = directLink(G’);
  return G;
}
```

```
getSkeleton(G', R) {
   k = 0; done = false;
   while done = false, do {
       done = true;
       for each node x in G',
         get Adj(x);
         if |Adj(x)|-1 < k, continue;
         done = false;
         for each node y in Adj(x),
           for each subset Z of Adj(x)\{y} with |Z|=k,
             if I(x,Z,y), then {Sxy = Z; rm <x,y> in G'; break;}
       k++;
    }
}
```

13

Types of Chains in the Structure

- Undirected chain

❑ A chain x-z-y where x and y are not adjacent is called a uncoupled meeting.

- Directed chain

1. A chain x→z→y is a head-to-tail meeting at z.
2. A chain x←z→y is a tail-to-tail meeting at z.
3. A chain x→z←y is a head-to-head meeting at z.

- When isolated, are the above directed meetings Markov equivalent?