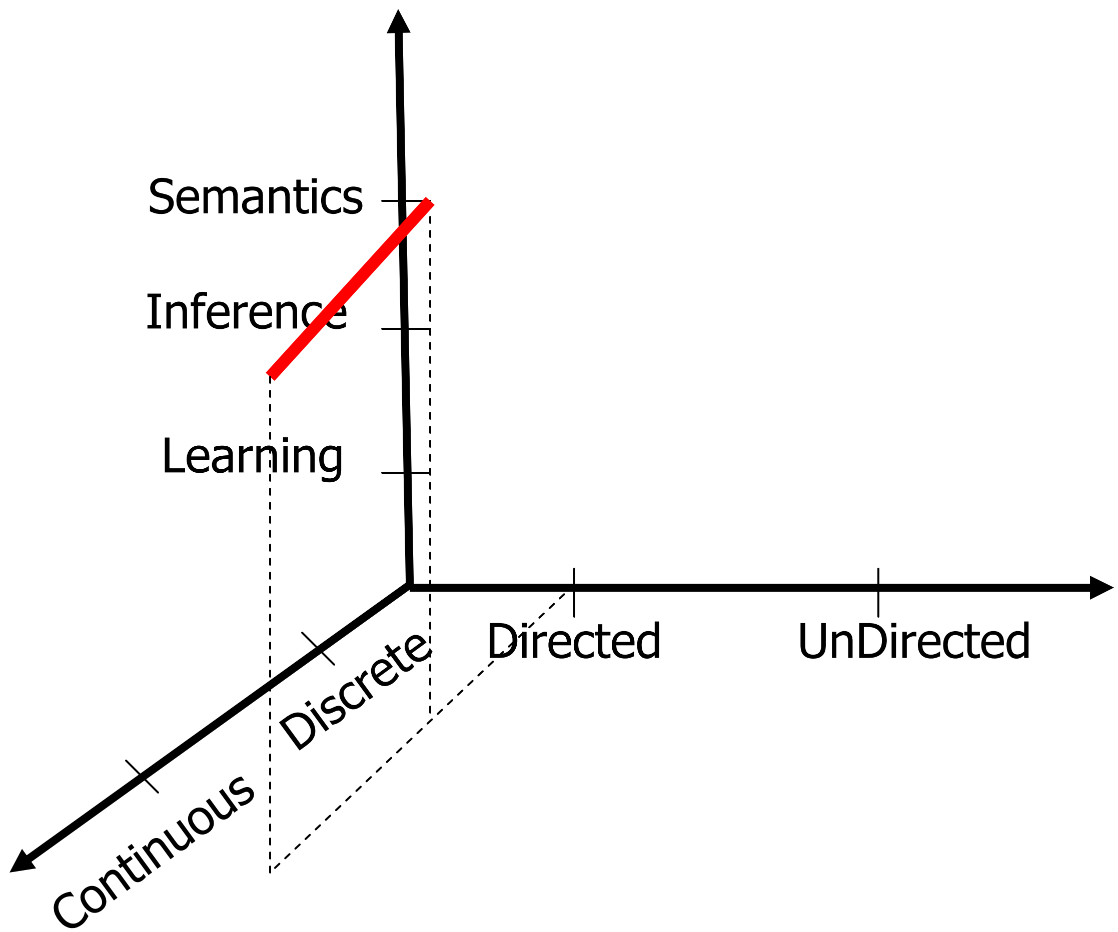
Bayesian Belief Networks: Representation II

KF, Chapter 6

Space of Topics



Outline

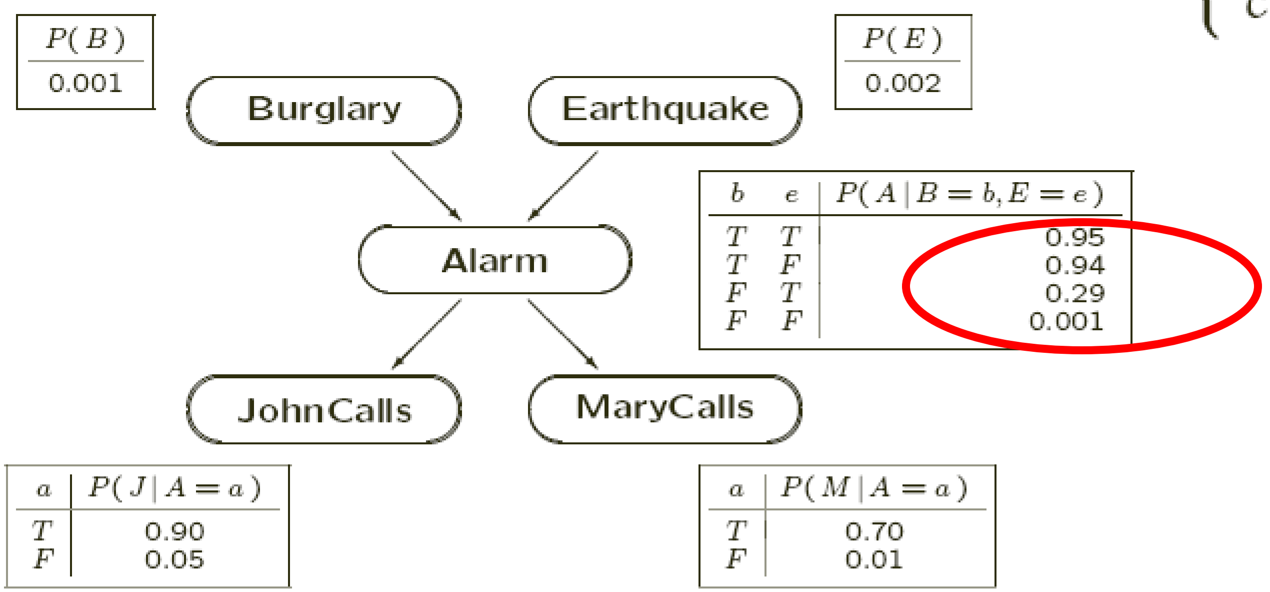
Alternative Conditional Probabilities

- Special forms for Discrete variables
- Continuous variables – Gaussians
  - Linear Gaussian Model
  - Hybrid Versions

Example Bayesian Net

Directed Acyclic Graph:

$$\mathcal{BN} = \left\{ \begin{array}{lll} \mathcal{N} & \text{Nodes} & \equiv \text{Variables} \\ \mathcal{A} & \text{Arcs} & \equiv \text{Dependencies} \\ \mathcal{C} & \text{CPTables} & \equiv \text{“weights”} \end{array} \right\}$$



- Discrete variables
- Explicit table

- **Nodes**: one for each random variable
- **Arcs**: one for each direct influence between two random variables
- **CPT**: each node stores a conditional probability table

  P(Node | Parents(Node))

  to quantify effects of “parents” on child

Simple forms of CPTable

- In gen'l: CPTable is function mapping
  values of parents to *distribution over child*

$$f: \left[\prod_{U \in Parents(X)} Dom(U)\right] \times Dom(X) \mapsto [0.1]$$

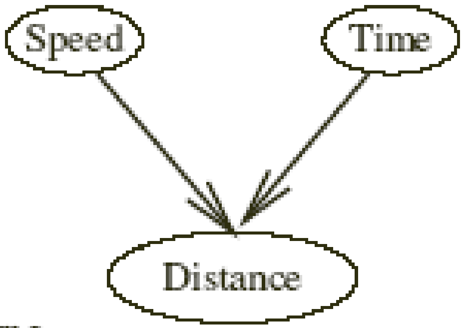
(Actually, $f': \prod_{U \in Parents(X)} Dom(U) \mapsto$ dist over X)

Cold	Flu	Malaria	$P(\text{Fever} \mid C,F,M)$	$P(\neg\text{Fever} \mid C,F,M)$
F	F	F	0.0	1.0
F	F	T	0.9	0.1
F	T	F	0.8	0.2
F	T	T	0.98	0.02
T	F	F	0.4	0.6
T	F	T	0.94	0.06
T	T	F	0.88	0.12
T	T	T	0.988	0.012

- Standard: Include $\prod_{U \in Parents(X)} |Dom(U)|$ rows, each with $|Dom(X)| - 1$ entries
- But... can be structure within CPTable: Deterministic, Noisy-Or, (Decision Tree), ...

Deterministic Node

- Given value of parent(s), specify unique value for child (logical, functional)



$$P(\texttt{Distance} \mid \texttt{Rate},\ \texttt{Time}) = \begin{cases} 1.0 & \text{if } \texttt{Distance} = \texttt{Rate} \cdot \texttt{Time} \\ 0.0 & \text{otherwise} \end{cases}$$

As if each row has just one 1, rest 0s:

Rate	Time	P(Dist=0 | R,T)	P(Dist=1 | R,T)	P(Dist=2 | R,T)
0	1	**1.0**	0.0	0.0
1	0	**1.0**	0.0	0.0
1	1	1.0	**1.0**	0.0
1	2	0.0	0.0	**1.0**
2	1	0.0	0.0	**1.0**
⋮		⋮		

Noisy-OR CPTable

- Each cause is independent of the others
- All possible causes are listed

Want: No Fever if none of Cold, Flu or Malaria

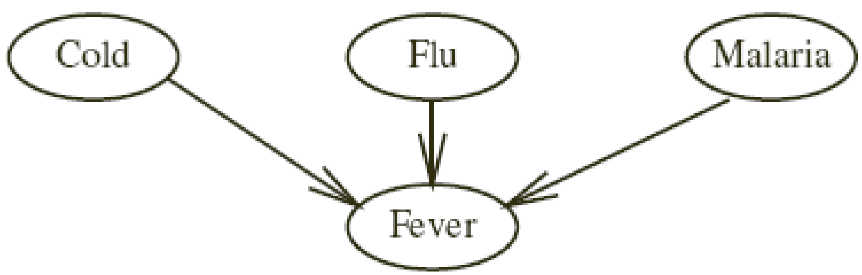
$P(\neg Fev \mid \neg Col, \neg Flu, \neg Mal) = 1.0$

\+ Whatever inhibits cold from causing fever

is independent of

whatever inhibits flu from causing fever

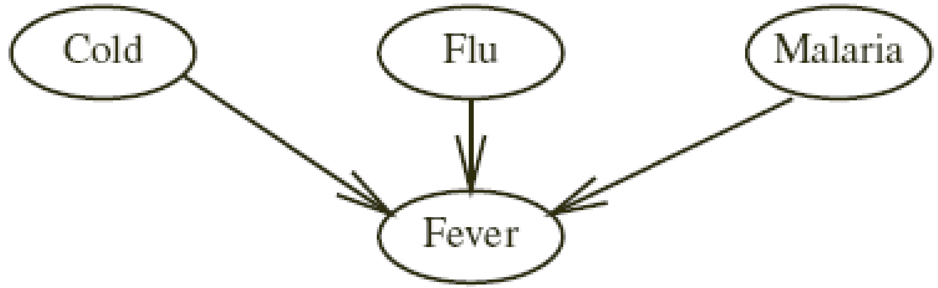
$P(\neg Fev \mid Cold, Flu) \approx P(\neg Fev \mid Cold)\, P(\neg Fev \mid Flu)$



Noisy-OR "CPTable" (2)

- $P(\texttt{Fev} \mid \neg\texttt{Col}, \neg\texttt{Flu}, \neg\texttt{Mal}) = 0$

$$P(\neg\texttt{Fev} \mid \texttt{Col}) \approx q_{col} = 0.6$$
$$P(\neg\texttt{Fev} \mid \texttt{Flu}) \approx q_{flu} = 0.2$$
$$P(\neg\texttt{Fev} \mid \texttt{Mal}) \approx q_{mal} = 0.1$$



- Independent inhibiters:
  $P(\neg\texttt{Fev} \mid \texttt{Col}, \texttt{Flu}) \approx P(\neg\texttt{Fev} \mid \texttt{Col}) \times P(\neg\texttt{Fev} \mid \texttt{Flu})$

$$P(\neg\texttt{Fever} \mid \pm_i d_i) = \prod_{i:+d_i} q_i$$

Cold	Flu	Malaria	$P(\neg\texttt{Fever} \mid \texttt{c,f,m})$	$P(\texttt{Fever} \mid \texttt{c,f,m})$
F	F	F	1.0	0.0
F	F	T	0.1	0.9
F	T	F	0.2	0.8
F	T	T	$0.02 = 0.2 \times 0.1$	0.98
T	F	F	0.6	0.4
T	F	T	$0.06 = 0.6 \times 0.1$	0.94
T	T	F	$0.12 = 0.6 \times 0.2$	0.88
T	T	T	$0.012 = 0.6 \times 0.2 \times 0.1$	0.988

Noisy-Or (Gen'l)

- Fever if Cold, Flu or Malaria

$$\text{Want} \begin{cases} P(\texttt{Fev} \mid \neg\texttt{Col}, \neg\texttt{Flu}, \neg\texttt{Mal}) &= 0 \\ P(\neg\texttt{Fev} \mid \texttt{Col}) &\approx q_{col} = 0.6 \\ P(\neg\texttt{Fev} \mid \texttt{Flu}) &\approx q_{flu} = 0.2 \\ P(\neg\texttt{Fev} \mid \texttt{Mal}) &\approx q_{mal} = 0.1 \end{cases}$$

- Define 3 numbers ("noise" parameters)

  q_{col}, q_{flu}, q_{mal}

$$\text{Def'n: } P(\neg\texttt{Fev} \mid \pm S_1, \pm S_2, \pm S_3) = \prod_{i:+S_i} q_i$$

$$P(\neg\texttt{Fev} \mid \neg\texttt{Col}, \neg\texttt{Flu}, \neg\texttt{Mal}) = 1.0$$

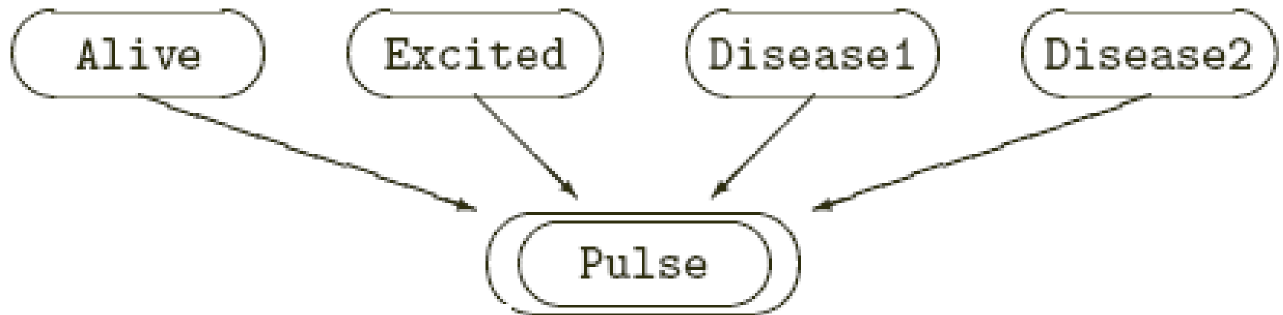
$$P(\neg\texttt{Fev} \mid \neg\texttt{Col}, \texttt{Flu}, \neg\texttt{Mal}) = q_{flu} = 0.2$$

$$P(\texttt{Fev} \mid \neg\texttt{Col}, \texttt{Flu}, \neg\texttt{Mal}) = 0.8$$

Assumes:
- each cause has independent chance of causing effect
- all causes listed
  (Leak node, to handle ALL OTHER CAUSES. . .)
- inhibiting factors independent

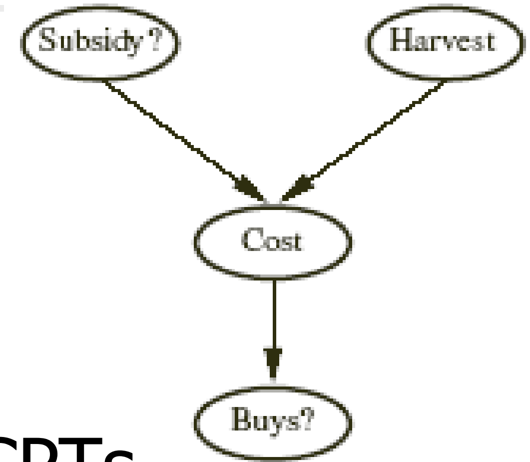
Note: Only k parameters, not 2^k

DecisionTree CPTable



A	E	D1	D2	χ s.t. $P($Pulse=χ | A,E,D1,D2$) = 1.0$
Y	Y	Y	Y	vhigh
Y	Y	Y	N	vhigh
Y	Y	N	Y	vhigh
Y	Y	N	N	vhigh
Y	N	Y	Y	high
Y	N	Y	N	med
Y	N	N	Y	med
Y	N	N	N	ok
N	Y	Y	Y	none
N	Y	Y	N	none
N	Y	N	Y	none
N	Y	N	N	none
N	N	Y	Y	none
N	N	Y	N	none
N	N	N	Y	none
N	N	N	N	none

Hybrid (discrete+continuous) Networks



- **Discrete**: Subsidy?, Buys?

  Continuous: Harvest, Cost

Option 1: Discretization

but possibly large errors, large CPTs

Option 2: Finitely parameterized canonical families

Problematic cases to consider. . .

- Continuous variable, discrete+continuous parents

  Cost
- Discrete variable, continuous parents

  Buys?

Continuous Child Variables

- For each "continuous" child E,
  - with continuous parents C
  - with discrete parents D
- Need conditional density function

  $P(E = e \mid C = c, D = d) = P_{D=d}(E = e \mid C = c)$

  for each assignment to discrete parents D=d
- Common: linear Gaussian model

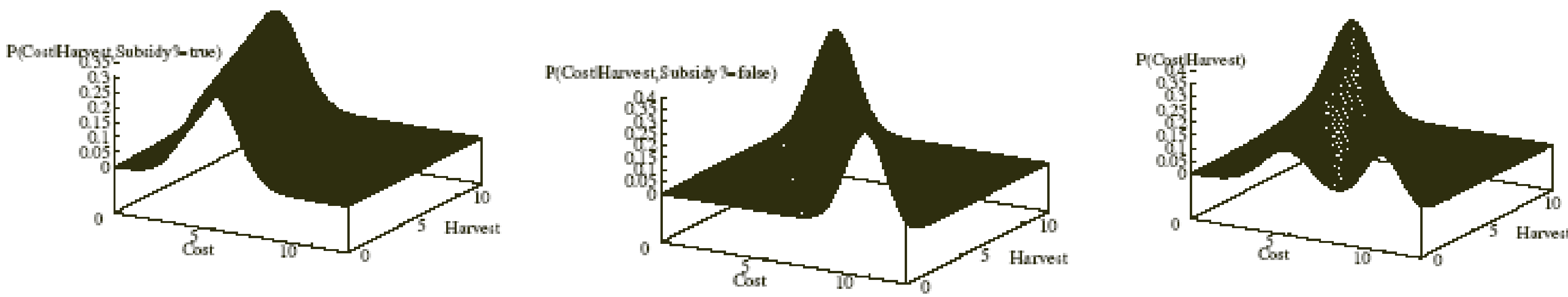
$$P(\texttt{Cost} = \texttt{c} \mid \texttt{Harvest} = \texttt{h},\ \texttt{Subsidy?} = \texttt{true})$$
$$= \mathcal{N}[a_t h + b_t, \sigma_t](c)$$
$$= \frac{1}{\sigma_t \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{c - (a_t h + b_t)}{\sigma_t}\right)^2\right)$$

$$P(\texttt{Cost} = \texttt{c} \mid \texttt{Harvest} = \texttt{h},\ \texttt{Subsidy?} = \texttt{false})$$
$$= \mathcal{N}[a_f h + b_f, \sigma_f](c)$$

If everything is Gaussian. . .

- All nodes continuous w/ LG dist'ns

  ⇒ full joint is a multivariate Gaussian



- Discrete+continuous LG network

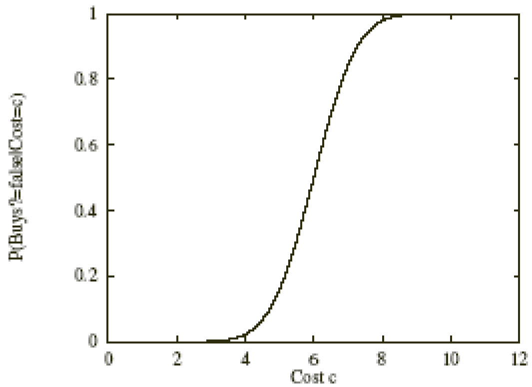
⇒ conditional Gaussian network

multivariate Gaussian over all continuous variables
for each combination of discrete variable values

Discrete variable w/Continuous Parents

- Probability of Buys? given Cost

  $\approx^?$ "soft" threshold:



- Probit distribution uses integral of Gaussian:

$$\Phi(x) = \int_{-\infty}^{x} \mathcal{N}[0,1](x)\, dx$$

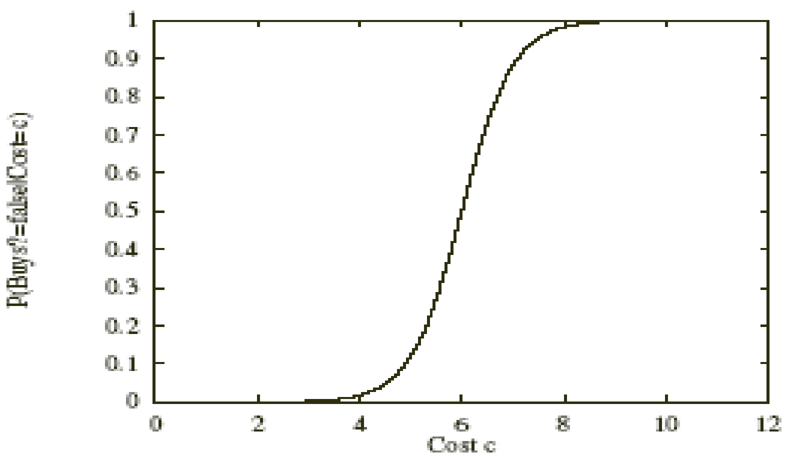
$$P(\texttt{Buys?} = \texttt{true} \mid \texttt{Cost} = c) = \Phi\left(\frac{\mu - c}{\sigma}\right)$$

$\approx$ hard threshold, whose location is subject to noise

Logit vs Probit

- Logit (Sigmoid) used in neural networks:

$$P(\texttt{Buys?} = \texttt{true} \mid \texttt{Cost} = \texttt{c}) = \frac{1}{1 + exp(-2\frac{\mu - c}{\sigma})}$$



- Shapes:

  Logit $\approx$ Probit

  but Logit has much longer tails

Summary

- Alternative encoding can be helpful
  - Efficiently represent discrete variables
  - Represent different types of variables