Artificial Intelligence

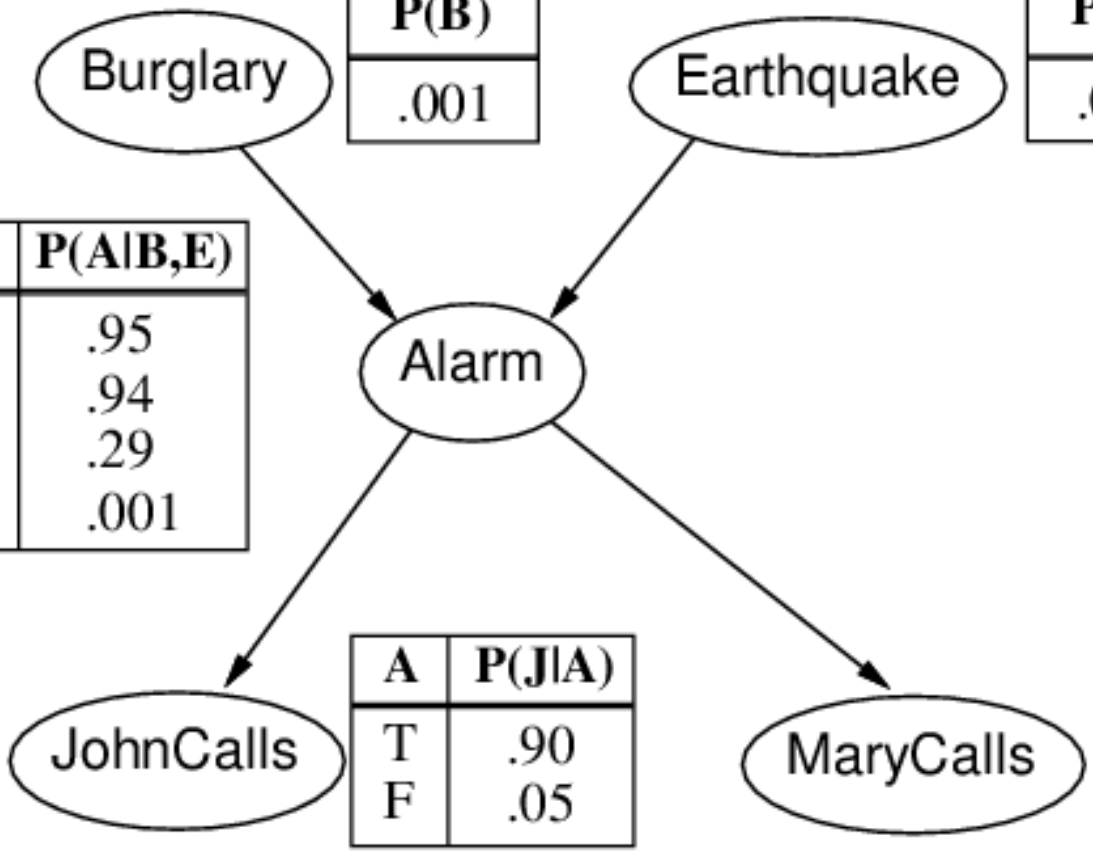
Probabilistic Reasoning (Part 3)

CS 444 – Spring 2019
Dr. Kevin Molloy
Department of Computer Science
James Madison University

Some Exercises

Consider this Bayesian network.

1. If no evidence is observed, are Burglary and Earthquake independent? Explain why/why not.
2. If we observe Alarm = true, are Burglary and Earthquake independent? Justify your answer by calculating whether the probabilities involved satisfy the definition of conditional independence.



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Compact Conditional Distributions

CPT grows exponentially with number of parents

CPT becomes infinite with continuous-valued parent or child

Solution: canonical distributions that are defined compactly

Determinitic nodes are the simplest case:

X = f(Parents(X)) for some function *f*

e.g. boolean functions

NorthAmerican ⟺ Canadian ∨ US ∨ Mexican

e.g. numerical relationships amongst continuous variable

CCD (Compact Conditional Distributions)

Noisy-OR distributions model multiple noninteracting causes

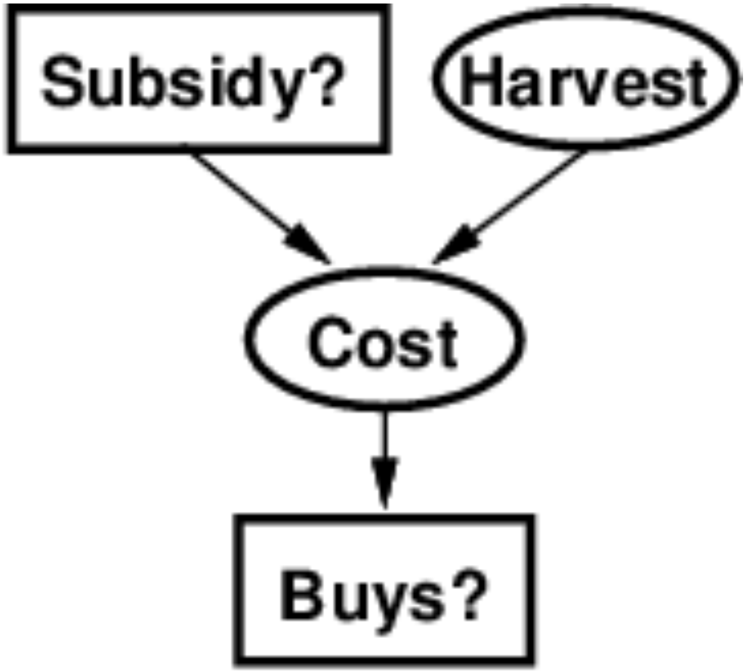
1. Parents U_1 … U_k include all causes (can add leak node)
2. Independent failure probability q_i for each cause alone

$\Longrightarrow$ P(X | U1 … Uj, ¬Uj+1 … ¬Uk) = 1 - $\prod_{i=1}^{j} q_i$

Number of parameters is linear in number of parents.

Cold	Flu	Malaria	P(Fever)	P(¬Fever)
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 = 0.2 x 0.1
T	F	F	0.4	0.6
T	F	T	0.94	0.06 = 0.6 x 0.1
T	T	F	0.88	0.12 = 0.6 x 0.2
T	T	T	0.988	0.012 = 0.6 x 0.2 x 0.1

Hybrid (Discrete + Continuous) Networks



Option 1: discretization

Option 2: finitely parameterized canonical families.

1. Continuous variable, discrete + continuous parents (e.g., **Cost**)
2. Discrete variable, continuous parents (e.g. **Buys**)

Continuous Child Variables

Need one conditional density function for child variables given continuous parents, for each possible assignment to discrete parents.

Most common is the linear Gaussian model, e.g.,:

P(Cost =c | Harvest = h, Subsidy? = true)

= N(ath + bt, σ)(c)

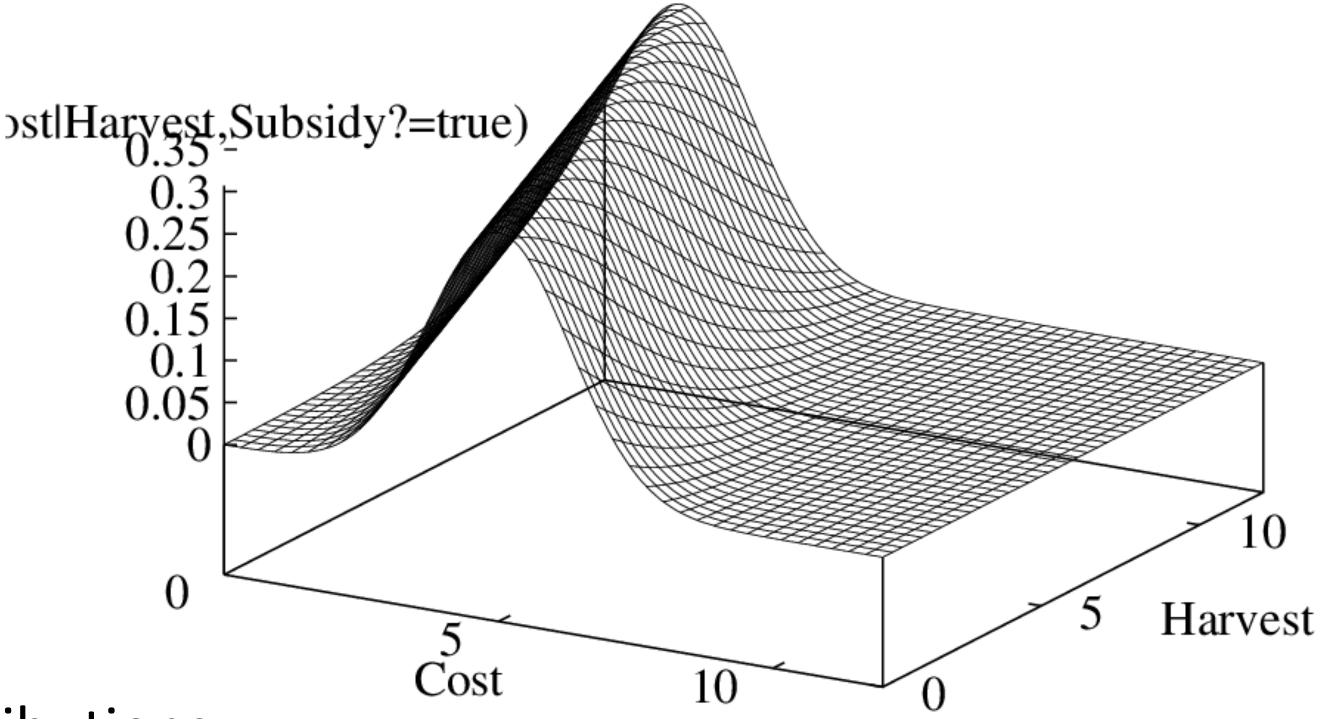
$$= \frac{1}{\sigma_t\sqrt{2\pi}} \exp(-\frac{1}{2}\left(\frac{c\ -\ a_t h +\ b_t}{\sigma_t}\right)^2)$$

Mean Cost varies linearly with Harvest, variance is fixed.

Linear variation is unreasonable over the full range

But works OK if the likely range of Harvest is narrow

Continuous Child Variables

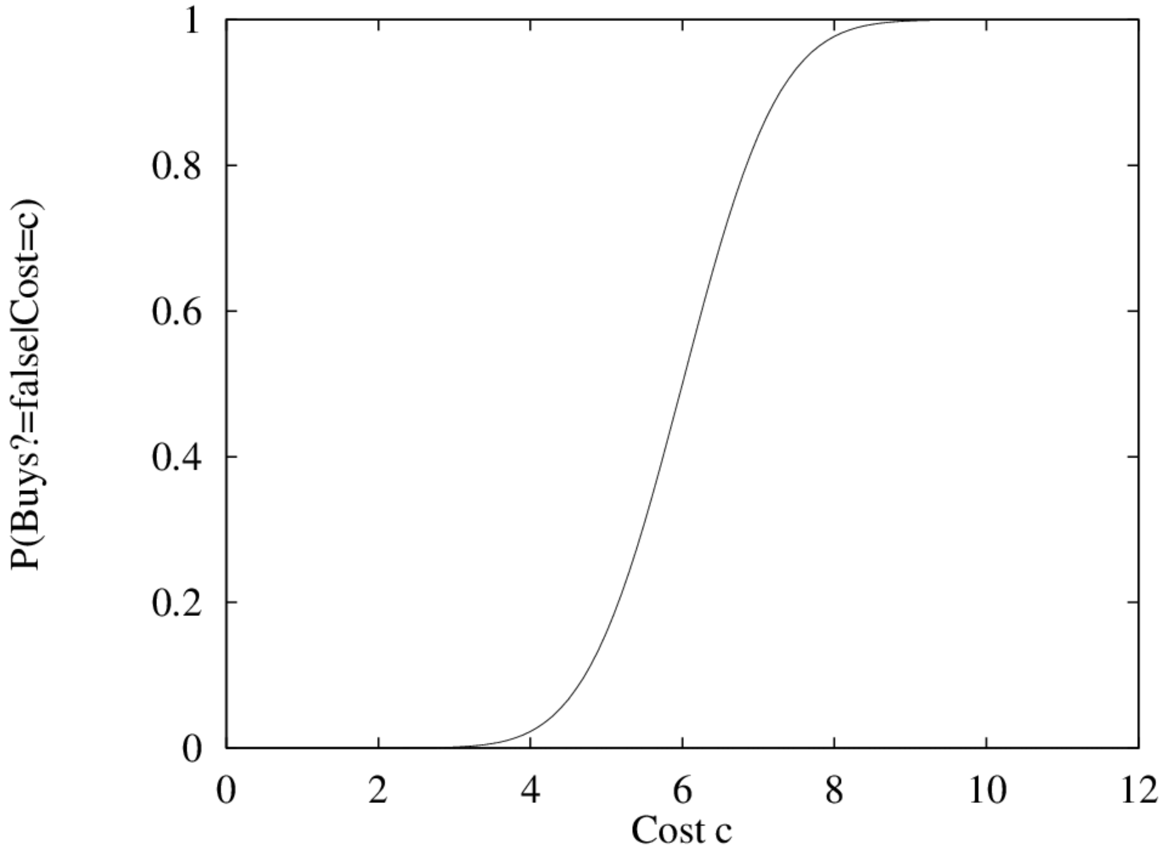


All continuous network with LG distributions:

⟹ full joint distribution is a multivariate Gaussian

Discrete Variable w/Continuous Parents

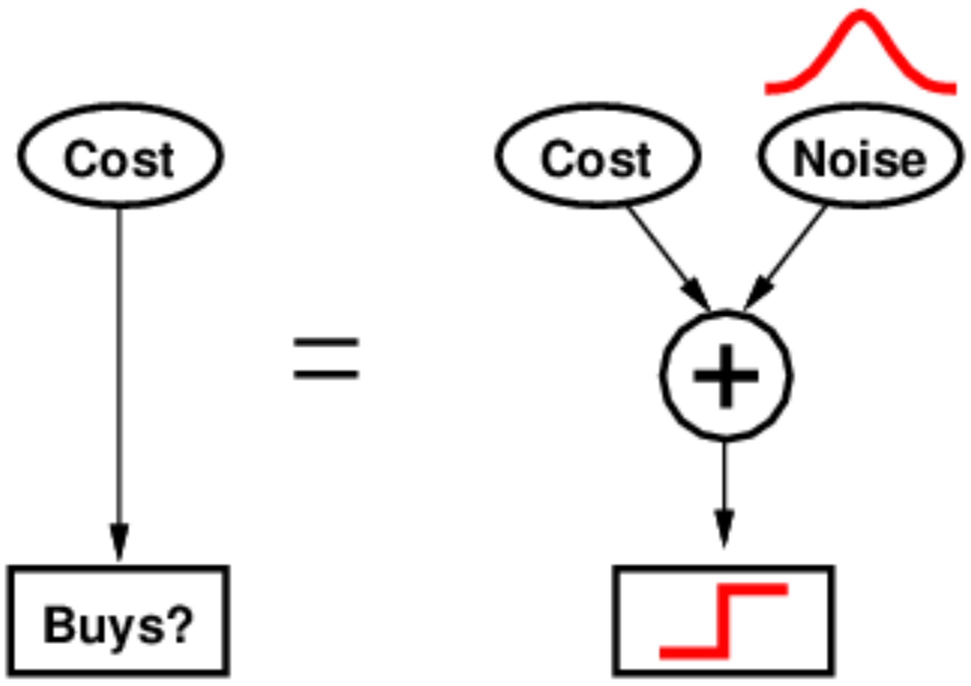
Probability of Buys? Given Cost should be a soft threshold:



Probit distsribution uses integral of Gaussian.

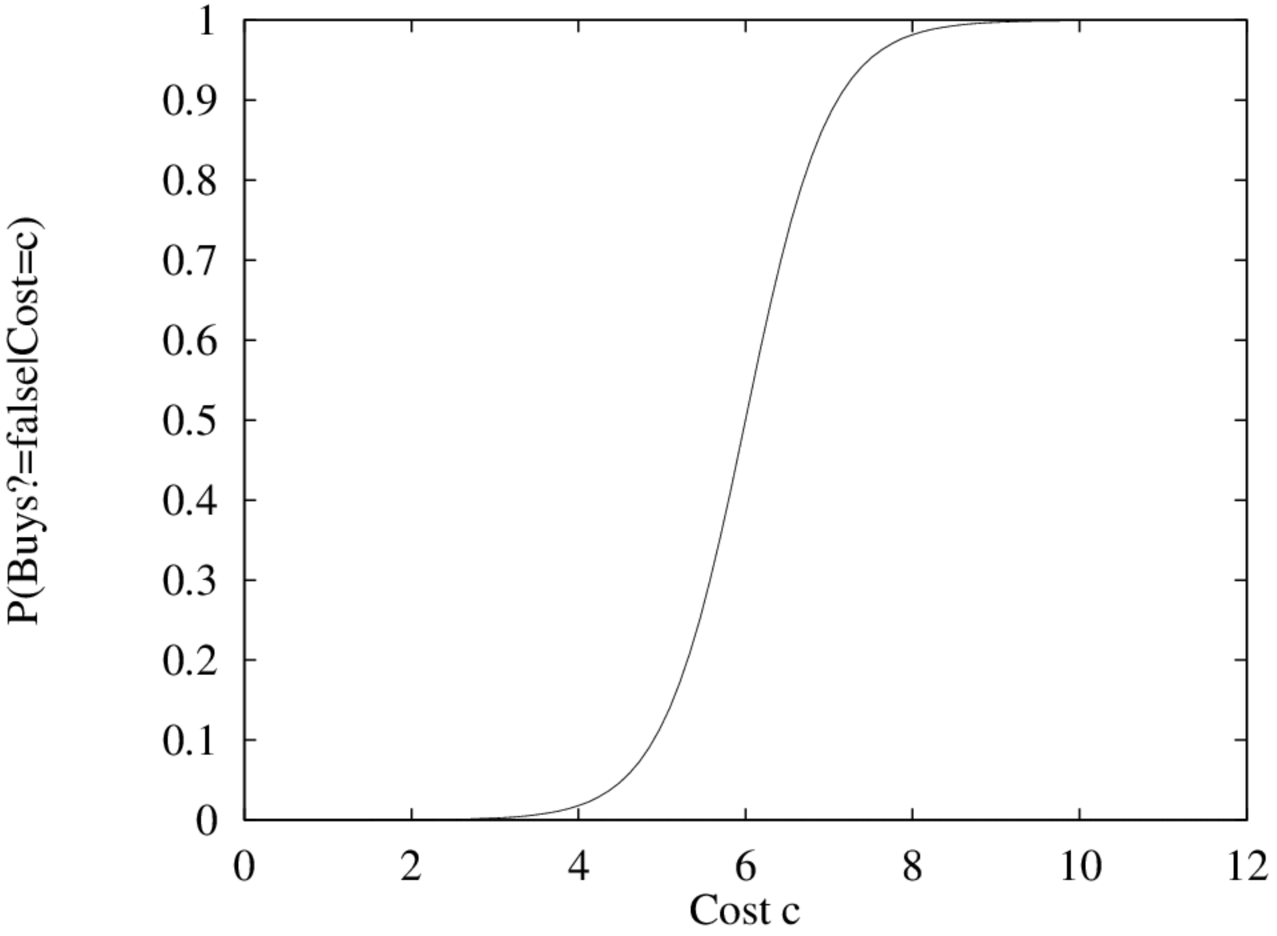
Why the probit?

Its sort of the right shape, can view as hard threshold subject to noise.



Discrete Variable

Sigmoid (or logit) distribution is also used (and frequently in neural networks).



Sigmoid has similar shape (but longer tails).

Summary on Bayesian Networks

Bayes nets provide a natural representation for (causally induced) conditional independence.

Topology + CPTs = compact representation of joint distributions

Generally easy for (non)expexrts to construct

Canonical distributions (noisy-OR) = compact representation of CPTs

Continuous variables $\Longrightarrow$ parameterized distribution (e.g. linear Gaussian)