Bayesian Network

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

- Know and memorize the theorems of probability
  - Recover the probability concepts
  - Recover the probability theorems
  - Recover the concepts of the marginal and the conditional independencies
- Understand Bayesian networks
  - Know the syntax and the semantics of Bayesian networks
  - Know how to factorize Bayesian networks
  - Able to calculate a probability with given conditions
- Understand the inference of Bayesian networks
  - Able to calculate parameters of Bayesian networks
  - Able to list the exact inference of Bayesian networks

INFERENCE ON BAYESIAN NETWORKS

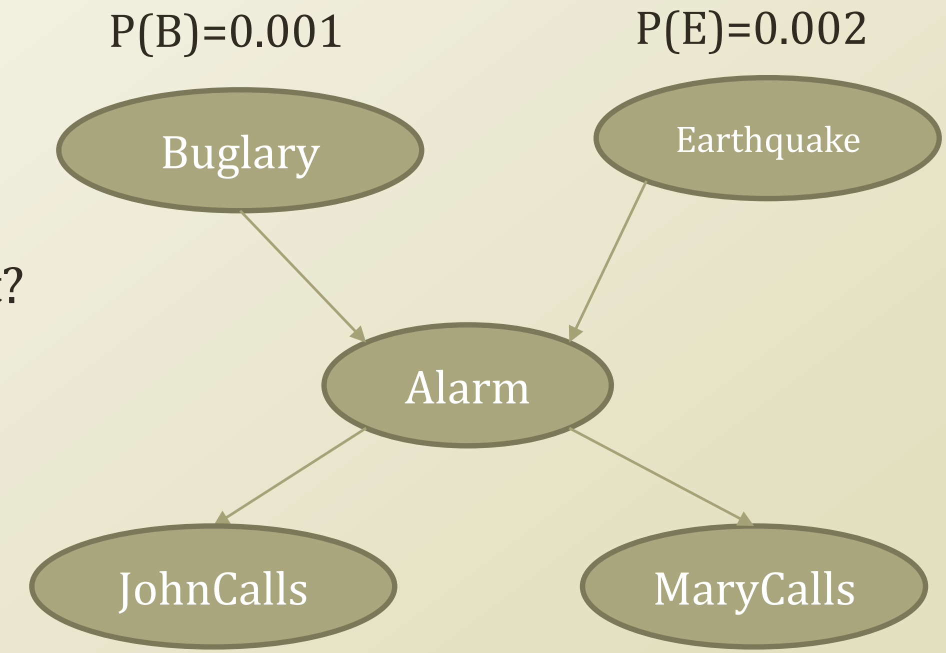
Inference Question 1: Likelihood

P(B=true, MC=true)=?

- Given a set of evidence, what is the likelihood of the evidence set?
  - $X = \{X_1 \ldots X_N\}$: all random variables
  - $X_V = \{X_{k+1} \ldots X_N\}$: evidence variables
    - x_V: evidence values
  - $X_H = X - X_V = \{X_1 \ldots X_k\}$: hidden variables
- General form
  - $P(x_V) = \sum_{X_H} P(X_H, X_V)$

$$= \sum_{x_1} \ldots \sum_{x_k} P(x_1 \ldots x_k, x_V)$$

  - Likelihood of x_V



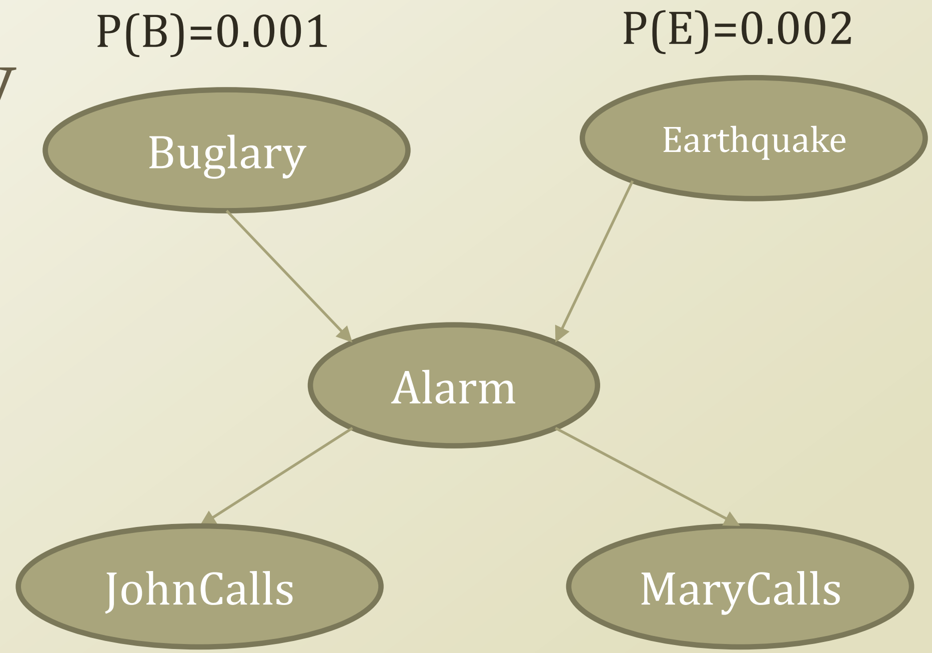
B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|A)
T	0.90
F	0.05

A	P(M|A)
T	0.70
F	0.01

Inference Question 2: Conditional Probability

P(A|B=true, MC=true)=?



- Given a set of evidence, what is the conditional probability of interested hidden variables?
  - $X_H = \{Y, Z\}$
    - Y: interested hidden variables
    - Z: uninterested hidden variables
- General form
  - $P(Y|x_V) = \sum_z P(Y, Z = z|x_V)$

    $= \sum_z \frac{P(Y, Z, x_V)}{P(x_V)}$

    $= \sum_z \frac{P(Y, Z, x_V)}{\sum_{y,z} P(Y = y, Z = z, x_V)}$
  - Conditional probability of Y given x_V

B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

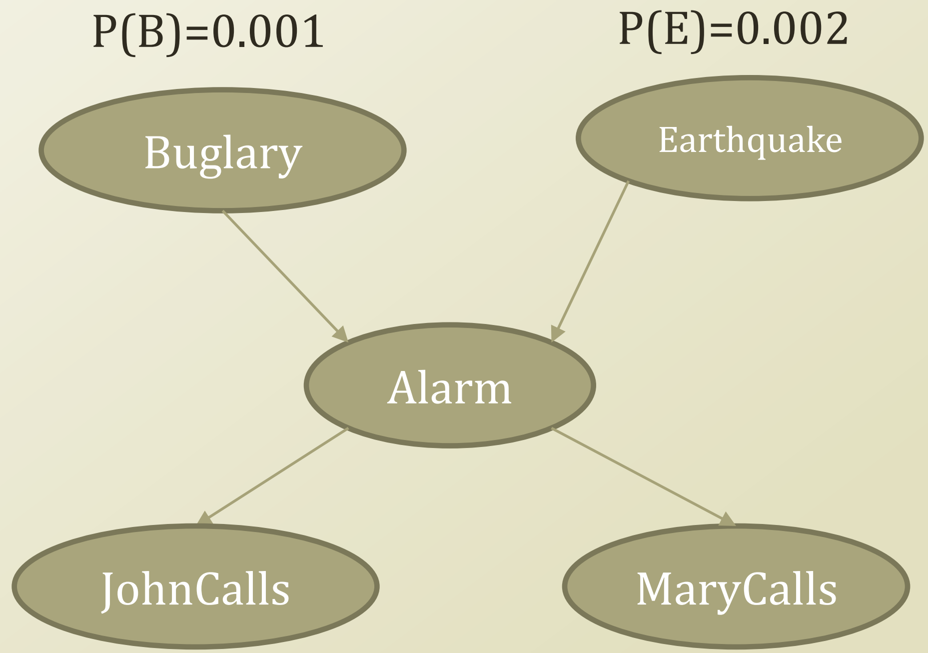
A	P(J|A)
T	0.90
F	0.05

A	P(M|A)
T	0.70
F	0.01

$\text{argmax}_a P(A|B=\text{true}, MC=\text{true})=?$

Inference Question 3: Most Probable Assignment

- Given a set of evidence, what is the most probable assignment, or explanation, given the evidence?
  - Some variables of interests
  - Need to utilize the inference question 2
    - Conditional probability
  - Maximum a posteriori configuration of *Y*
- Applications of *a posteriori*
  - Prediction
    - B, E → A
  - Diagnosis
    - A → B, E



B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|A)
T	0.90
F	0.05

A	P(M|A)
T	0.70
F	0.01

Marginalization and Elimination

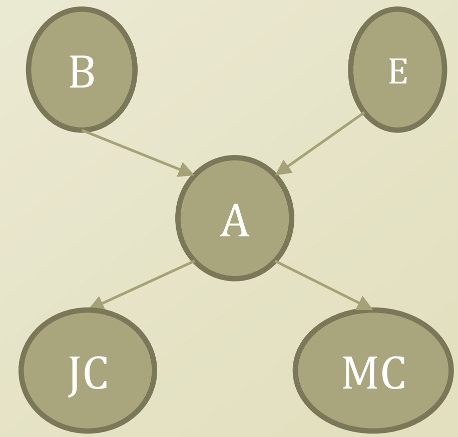
- Computing joint probabilities is a key
  - How to compute them?
  - Many, many, many multiplications and summations
    - P(a=true,b=true,mc=true)=$\sum_{JC}\sum_{E} P(a,b,E,JC,mc)$

$$=\sum_{JC}\sum_{E} P(JC|a)P(mc|a)P(a|b,E)P(E)P(b)$$

  - In big Oh notation?
- Is there any better method?
  - What-if we move around the summation?
    - P(a,b,mc)=$\sum_{JC}\sum_{E} P(a,b,E,JC,mc)$

$$=P(b)P(mc|a)\sum_{JC} P(JC|a)\sum_{E} P(a|b,E)P(E)$$

  - Did we reduced the computation complexity?



P(E)=0.002

P(B)=0.001

B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|A)
T	0.90
F	0.05

A	P(M|A)
T	0.70
F	0.01

Variable Elimination

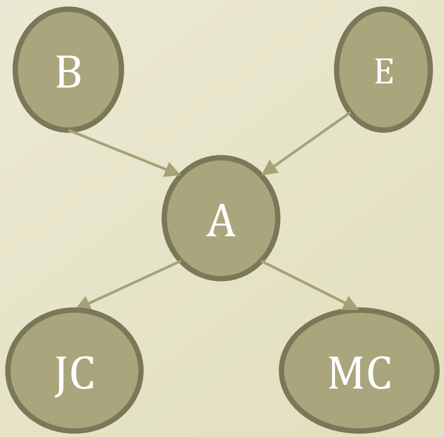
- Preliminary
  - $P(e|jc,mc) = \alpha P(e,jc,mc)$
- Joint probability (e=jc=mc=true)
  - $P(e,jc,mc,B,A) =$
    $\alpha P(e) \sum_B P(b) \sum_A P(a|b,e)\, P(jc|a) P(mc|a)$
    - Line up the terms by the topological order
    - Consider a probability distribution as a function
      - $f_E(E=t) = 0.002$
  - $= \alpha f_E(e) \sum_B f_B(b) \sum_A f_A(a,b,e)\, \underbrace{f_J(a) f_M(a)}$

A	$f_{JM}(A)$
T	0.63
F	0.0005

A	$f_J(A)$
T	0.90
F	0.05

A	$f_M(A)$
T	0.70
F	0.01

  - $= \alpha f_E(e) \sum_B f_B(b) \sum_A f_A(a,b,e)\, f_{JM}(a)$



P(B)=0.001

B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

P(E)=0.002

A	P(J|A)
T	0.90
F	0.05

A	P(M|A)
T	0.70
F	0.01

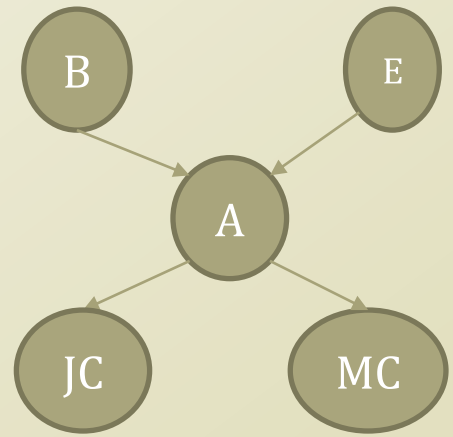
Variable Elimination cont.

- $= \alpha f_E(e) \sum_B f_B(b) \sum_A f_A(a,b,e) f_{JM}(a)$

A	B	E	$f_{AJM}(A,B,E)$
T	T	T	0.95*0.63
T	T	F	0.94*0.63
T	F	T	0.29*0.63
T	F	F	0.001*0.63
F	T	T	0.05*0.0005
F	T	F	0.06*0.0005
F	F	T	0.71*0.0005
F	F	F	0.999*0.0005

A	B	E	$f_A(A,B,E)$
T	T	T	0.95
T	T	F	0.94
T	F	T	0.29
T	F	F	0.001
F	T	T	0.05
F	T	F	0.06
F	F	T	0.71
F	F	F	0.999

A	$f_{JM}(A)$
T	0.63
F	0.0005



P(B)=0.001

P(E)=0.002

B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|A)
T	0.90
F	0.05

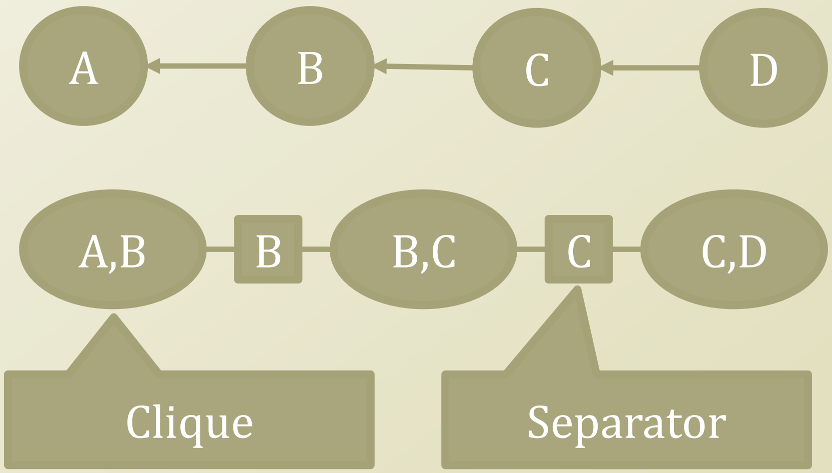
A	P(M|A)
T	0.70
F	0.01

- $= \alpha f_E(e) \sum_B f_B(b) \sum_A f_{AJM}(a,b,e)$
- $= \alpha f_E(e) \sum_B f_B(b) f_{\bar{A}JM}(b,e)$
- $= \alpha f_E(e) \sum_B f_{B\bar{A}JM}(b,e)$
- $= \alpha f_E(e) f_{\bar{B}\bar{A}JM}(e)$
- $= \alpha f_{E\bar{B}\bar{A}JM}(e)$

B	E	$f_{\bar{A}JM}(B,E)$
T	T	0.95*0.63+0.05*0.0005
T	F	0.94*0.63+0.06*0.0005
F	T	0.29*0.63+0.71*0.0005
F	F	0.001*0.63+0.999*0.0005

Potential Functions

- $P(A, B, C, D)$
- $= P(A|B)P(B|C)P(C|D)P(D)$
- Let's define a potential function
  - Potential function:
    a function which is not a probability function yet, but once normalized it can be a probability distribution function
  - Potential function on nodes
    - $\psi(a, b), \psi(b, c), \psi(c, d)$
  - Potential function on links
    - $\phi(b), \phi(c)$
- How to setup the function?
  - $P(A, B, C, D) = P(U) = \frac{\prod_N \psi(N)}{\prod_L \phi(L)} = \frac{\psi(a,b)\psi(b,c)\psi(c,d)}{\phi(b)\phi(c)}$
    - $\psi(a, b) = P(A|B), \psi(b, c) = P(B|C), \psi(c, d) = P(C|D)P(D)$
    - $\phi(b) = 1, \phi(c) = 1$
  - $P(A, B, C, D) = P(U) = \frac{\prod_N \psi(N)}{\prod_L \phi(L)} = \frac{\psi^*(a,b)\psi^*(b,c)\psi^*(c,d)}{\phi^*(b)\phi^*(c)}$
    - $\psi^*(a, b) = P(A, B), \psi^*(b, c) = P(B, C), \psi^*(c, d) = P(C, D)$
    - $\phi^*(b) = P(B), \phi^*(c)$ =P(C)

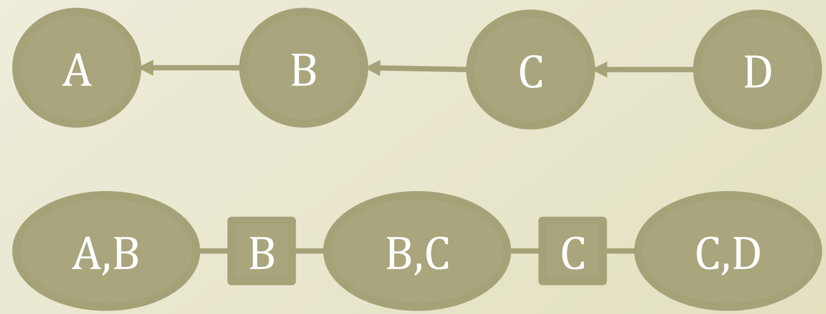


Marginalization is also applicable:

$$\psi(w) = \sum_{v-w} \psi(v)$$

Constructing a potential of a subset (w) of all variables (v)

Absorption in Clique Graph



- Only applicable to the tree structure of clique graph
- Let's assume
  - $P(B) = \sum_A \psi(A,B)$
  - $P(B) = \sum_C \psi(B,C)$
  - $P(B) = \phi(B)$
  - How to find out the ψs and the ϕs?
    - When the ψs change by the observations: P(A,B) → P(A=1,B)
    - A single ψ change can result in the change of multiple ψs
    - The effect of the observation propagates through the clique graph
    - Belief propagation!
- How to propagate the belief?
  - Absorption (update) rule
  - Assume $\psi^*(A,B)$,$\psi(B,C)$, and $\phi(B)$
  - Define the update rule for separators
    - $\phi^*(B) = \sum_A \psi^*(A,B)$
  - Define the update rule for cliques
    - $\psi^*(B,C)=\psi(B,C)\frac{\phi^*(B)}{\phi(B)}$

Why does this work?

$$\sum_C \psi^*(B,C) = \sum_C \psi(B,C)\frac{\phi^*(B)}{\phi(B)} = \frac{\phi^*(B)}{\phi(B)}\sum_C \psi(B,C) = \frac{\phi^*(B)}{\phi(B)}\phi(B) = \sum_A \psi^*(A,B)$$

Guarantees the local consistency

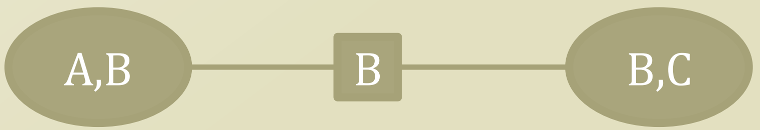
→ Global consistency after iterations

Simple Example of Belief Propagation

- Initialized the potential functions
  - $\psi(a,b) = P(a|b), \psi(b,c) = P(b|c)P(c)$
  - $\phi(b) = 1$
- Example 1. P(b)=?
  - $\phi^*(b) = \sum_a \psi(a,b) = 1$
  - $\psi^*(b,c) = \psi(b,c)\frac{\phi^*(b)}{\phi(b)} = P(b|c)P(c) = P(b,c)$
  - $\phi^{**}(b) = \sum_c \psi(b,c) = \sum_c P(b,c) = P(b)$
  - $\psi^*(a,b) = \psi(a,b)\frac{\phi^{**}(b)}{\phi^*(b)} = \frac{P(a|b)P(b)}{1} = P(a,b)$
  - $\phi^{***}(b) = \sum_a \psi^*(a,b) = P(b)$
- Example 2. P(b|a=1,c=1)=?
  - $\phi^*(b) = \sum_a \psi(a,b)\delta(a=1) = P(a=1|b)$
  - $\psi^*(b,c) = \psi(b,c)\frac{\phi^*(b)}{\phi(b)} = P(b|c=1)P(c=1)\frac{P(a=1|b)}{1}$
  - $\phi^{**}(b) = \sum_c \psi(b,c)\,\delta(c=1) = P(b|c=1)P(c=1)P(a=1|b)$
  - $\psi^*(a,b) = \psi(a,b)\frac{\phi^{**}(b)}{\phi^*(b)} = \mathrm{P}(\mathrm{a}=1|\mathrm{b})\frac{P(b|c=1)P(c=1)P(a=1|b)}{P(a=1|b)} = P(b|c=1)P(c=1)P(a=1|b)$
  - $\phi^{***}(b) = \sum_a \psi^*(a,b)\,\delta(a=1) = P(b|c=1)P(c=1)P(a=1|b)$

Bayesian Network

1) $\phi^*(b)$ 2) $\psi^*(b,c)$
4) $\psi^*(a,b)$ 3) $\phi^{**}(b)$
5) $\phi^{***}(b)$



Clique Graph