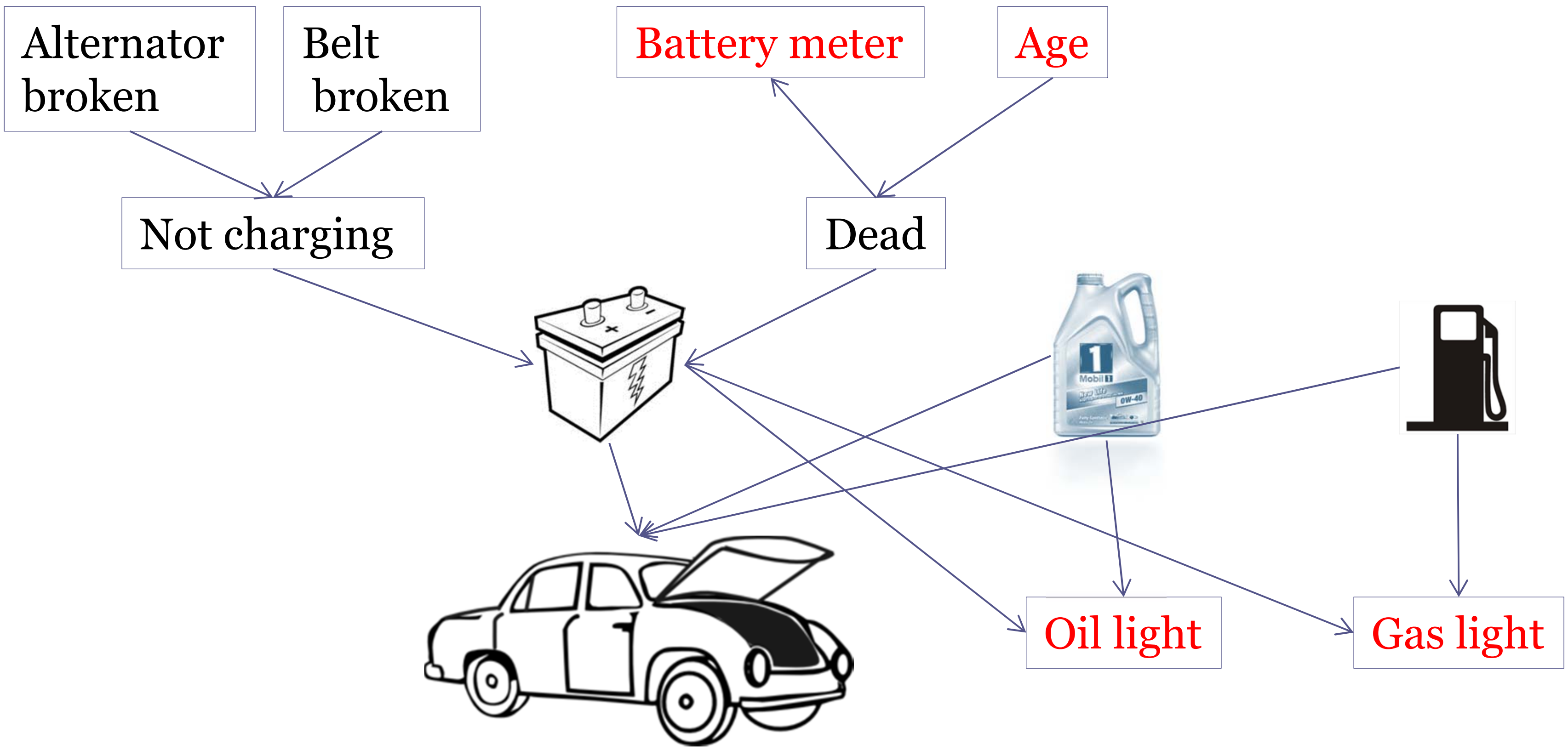
Intro to Artificial Intelligence

Lecture 3: Probability in AI

Ahmed Sallam $\left\{\frac{http://sallam.cf}{http://sallamah.weebly.com}\right\}$

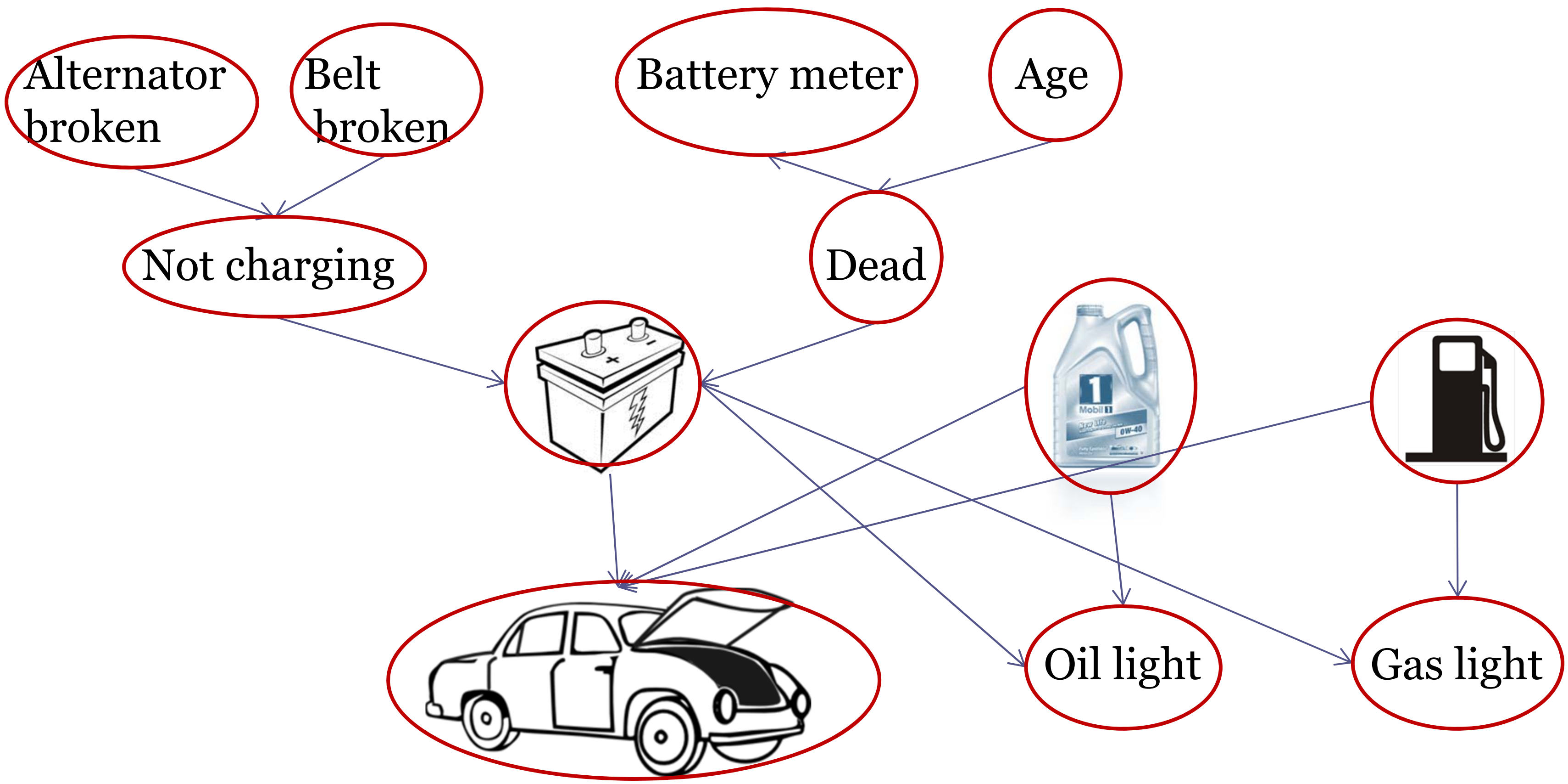
Complex problem

- Many Reasons !
  - Can we diagnose ?



- Bayes Network can assist to diagnose this complex problem

Bayes Network



- Set of nodes
  - Represent events you might/not know
  - Called random variables
- Arcs
  - Represent influence of child by its parent
  - Influence is Probabilistic not Deterministic

- Suppose these 12 variables are binary then we have 2^{16} values to choose between.
- That's why we need Bayes Network.

Bayes network

- Applications
  - Diagnostic.
  - Prediction.
  - Machine learning.
  - Finance
  - Robotics
- AI advanced techniques
  - Particle filters
  - Kalman filters
  - Hidden Markov chain
    - MDPs and POMDPs

Goals

- How to construct Bayes network!
- How to compute the probability!
- How to observe reasons!

Assumptions

- Binary events
- Simple bayes network

Probabilities

Basics

- P(H)+P(T)=1
- If P(H)=0.5 then P(T)=?
  - P(T) =0.5 (fair coin)
- If P(H)=0.25 then P(T)=?
  - P(T)=0.75 (unfair coin)

Independent events

- If we have fair coin then P(H,H, H) = ?
  - P(H, H, H) = 1/8
- If X_i = result of i flip, X_i={H, T} , and we have a fair coin then p(X_1=X_2=X_3=X_4)=?
  - P(X_1=X_2=X_3=X_4)=2/16
- P({X_1, X_2, X_3, X_4} contains ≥ 3 heads)=?
  - = 5/16

H	H	H
H	H	T
H	T	H
H	T	T
T	H	H
T	H	T
T	T	H
T	T	T

H	H	H	H
H	H	H	T
H	H	T	H
H	H	T	T
H	T	H	H
H	T	H	T
H	T	T	H
H	T	T	T
T	H	H	H
T	H	H	T
T	H	T	H
T	H	T	T
T	T	H	H
T	T	H	T
T	T	T	H
T	T	T	T

Independent events rules

- Joint probability:
  If $(X \perp Y)$ then P(X,Y) = P(X)*P(Y)
  - If P(H)=0.5 then P(H, H, H) = ?
  - P(H, H, H) = P(H)*P(H)*P(H)=1/8

- If X_i = result of i flip, X_i={H, T} , P(X_i)=0.5 ∀ i then P(X_1=X_2=X_3=X_4)=?
  - P(X_1=X_2=X_3=X_4)=P(H,H,H,H)+ P(T, T, T, T) =P(H)*P(H)*P(H)*P(H) + P(T)*P(T)*P(T)*P(T) =1/16+1/16

- Complement:
  If P(A)=p then P(¬A)=1-p

Dependent events

- suppose X_1 is a fair coin, and X_2 is selected from a set of unfair coins such that:
  - $P(X_1)$ $\rightarrow H$: $P(X_2{=}H \mid X_1{=}H){=}0.9$
    $\rightarrow T$: $P(X_2{=}T \mid X_1{=}T){=}0.8$
- $P(X_2{=}H){=}$?
  - $= 0.55$
  - $= P(X_1 = H) * P(X_2 = H \mid X_1 = H) + P(X_1 = T) * P(X_2 = H \mid X_1 = T)$
  - $= 0.5 * 0.9 + 0.5 * 0.2$
- X_1 is called conditional variable and $p(X_2)$ is called conditional probability.

Dependent events rules

- Total probability:
  - If Y is depend on X then
    $P(Y) = \sum_i P(X_i) * P(Y|X_i)$

- Complement:
  - $P(\neg Y|X) = 1 - P(Y|X)$
  - $P(Y|\neg X) = 1 - P(Y|X)$

Quiz

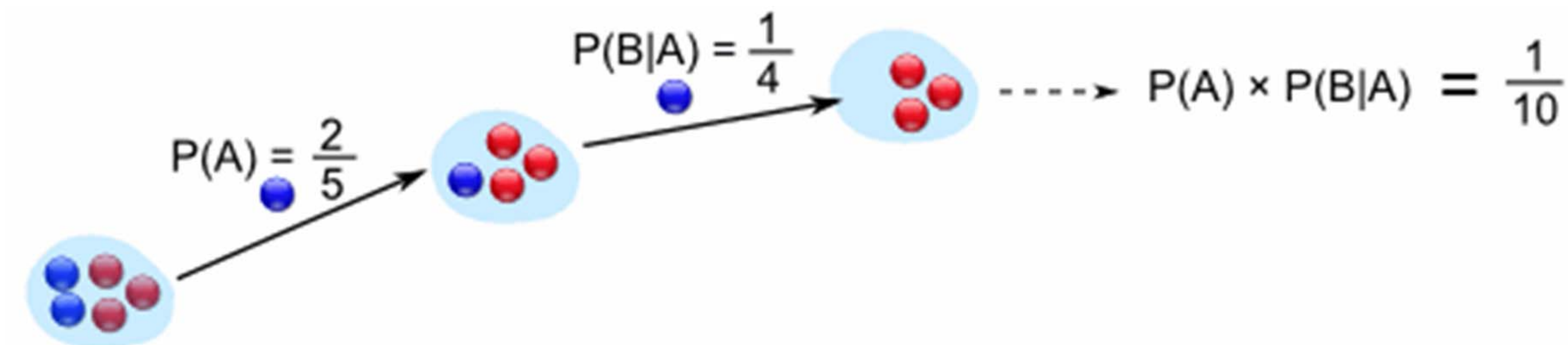
- Sunny/Rainy days example.
  - $P(D_1=S)=0.9$
  - $P(D_2=S|D_1=S)=0.8$
  - $P(D_2=S|D_1=R)=0.6$

- $P(D_2=R|D_1=S)= ?$
  - $1-0.8=0.2$
- $P(D_2=R|D_1=R)= ?$
  - $1-0.6=0.4$
- $P(D_2=S) =?$
  - $= (0.9*0.8)+(0.1*0.6)=0.78$

Dependent events rules cont.

- Joint probability :
  - If B is depend on A then P(B,A)=? =P(A) * P(B|A)

If X,Y independent
P(B, A)=P(A)*P(B)



- P(A) is called the prior, P(B|A) is called the posterior.

Quiz

- Cancer example.
  - $P(C)=0.01$ $P(\neg C)= 0.99$
  - $P(+|C)=0.9$ $P(-|C)= 0.1$
  - $P(+|\neg C)=0.2$ $P(-|\neg C)=0.8$

- $P(+, C)=?$
  - $= 0.01*0.9$
- $P(-, C)=?$
  - $=0.01*.1$

Dependent events rules cont.

- Continue with the cancer example
- P(C|+)=?
  - $\because P(C,+) = P(+) * P(C|+)$
  - $=P(C) * P(+|c)$
  - $\therefore P(C|+) = \frac{P(C,+)}{P(+)}$

$$= \frac{P(C)*P(+|C)}{P(C)*P(+|C) + P(\neg C)*P(+|\neg C)}$$

$$= \frac{0.01*0.9}{0.01*.9 + 0.99*0.2} = 0.0434$$

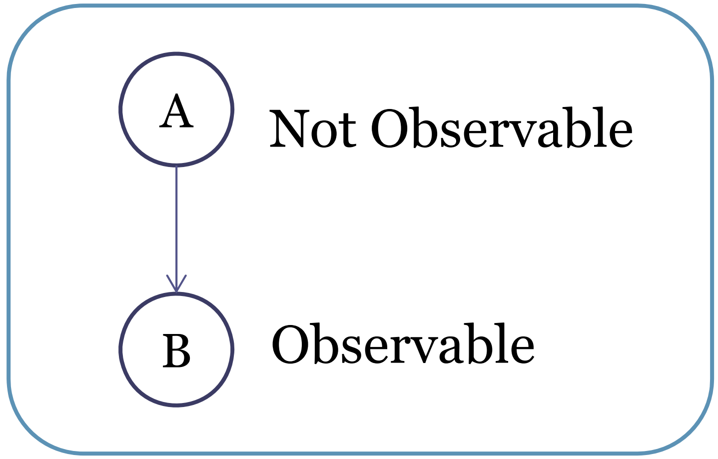
P(C)=0.01
P(+|C)=0.9
P(+| ¬ C)=0.2

Bayes rule

- $P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$
- P(A) called the prior, P(A|B) called the posterior, P(B|A) is called likelihood, and P(B) is called marginal likelihood.

- P(B) is calculated with the total probability
  - $P(B) = \sum_i P(B|A_i) * P(A_i)$

Bayes network



Bayes network

- We have P(A), P(B|A), P(B|¬A)
- We need to get P(A|B). i.e., in the cancer example we interest to correctly diagnose the existence of cancer (event A) after performing the test (event B).
- This is called Diagnostic reasoning = inverse of causal reasoning

Bayes Network

Bayes rule adaptation

- $P(A|B) = \frac{P(B|A)*P(A)}{P(B)}$ =P(A)*P(B|A)+P(¬A)*P(¬ B|A)
  - P(B) is complex to calculate.

- $P(\neg A|B) = \frac{P(B|\neg A)*P(\neg A)}{P(B)}$
  - P(A|B)+P(¬A|B)=1

Normalizer

- P\`(A|B) =P(B|A)*P(A)
  - P(A|B) =η P\`(A|B)

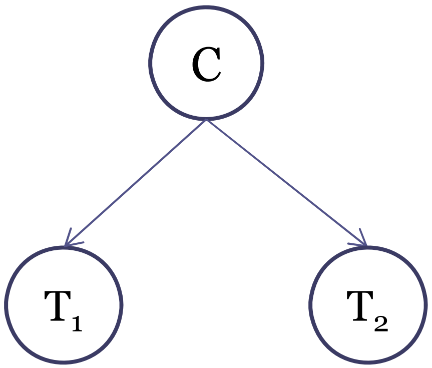
Eta

- P\`(¬A|B) =P(B| ¬ A)*P(¬ A)
  - P(¬ A|B) =η P\`(¬ A|B)

- η= ?
  - η(P\`(A|B)+ P\`(¬ A|B)) = P(A|B)+P(¬ A|B)
  - η(P\`(A|B)+ P\`(¬ A|B)) = 1
  - η = (P\`(A|B)+ P\`(¬ A|B))$^{-1}$

Bayes network

- P(C)=0.01
- P(+|C)=0.9
- P(-|¬C)=0.8
- P(¬C)=0.99
- P(-|C)=0.1
- P(+|¬C)=0.2



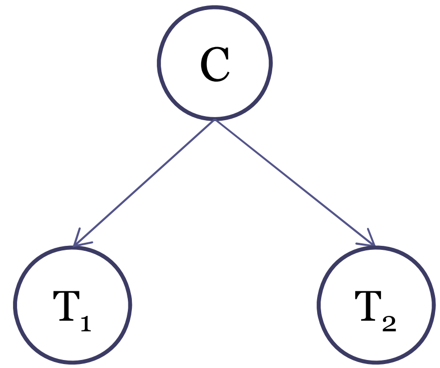
- P(C | ++) = ??

	P(..)	P(+ | ..)	P(+| ..)	P`(.. | ++)	P(..|++)
C	0.01	0.9	0.9	0.01*0.9*0.9 =0.0081	0.1698
¬C	0.99	0.2	0.2	0.99*0.2*0.2 =0.0396	0.8302
				0.0477	

Bayes network cont.

- $P(C)=0.01$
- $P(+|C)=0.9$
- $P(-|\neg C)=0.8$
- $P(\neg C)=0.99$
- $P(-|C)=0.1$
- $P(+|\neg C)=0.2$

C causes T_1 , T_2

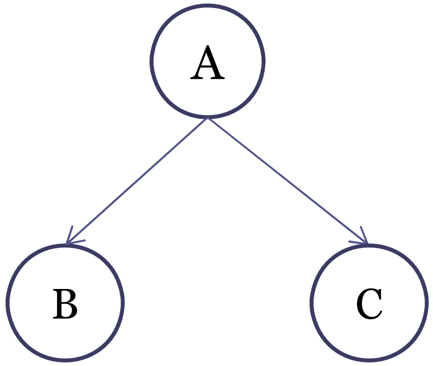


But T_1 , T_2 are independent

- $P(C \mid T_1=+\ \ T_1=-) \rightarrow P(C \mid +\ -)=$??

	P(..)	P(+ | ..)	P(-| ..)	P\`(.. | + -)	P(..|+ -)
C	0.01	0.9	0.1	0.01*0.9*0.1 =0.0009	0.0056
¬C	0.99	0.2	0.8	0.99*0.2*0.8 =0.1584	0.994
				0.1593	

Conditional Independence



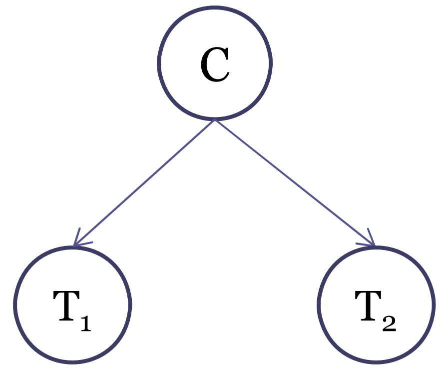
Given A

B⊥C
B and C are conditionally independent

B⊥C |A

- Suppose that we don't know A (not given)
- Is B and C still independent ??
  - Yes
  - No
- E.g. The cancer example, if B is positive then the probability of A is raised and will affect the probability of C.

Conditional Independence cont.1



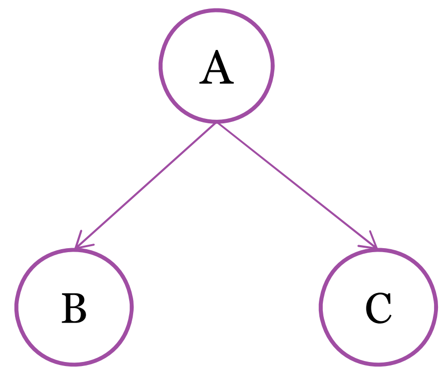
- P(C)=0.01
- P(+|C)=0.9
- P(-|¬C)=0.8

- $P(+_2|C,+_1)$ = ??
  - $\because P(+_2|C, +_1) = P(+_2 \mid +_1)$
  - $\because P(+) = P(C) * P(+|C) + P(\neg C) * P(+|\neg C)$
  - $\therefore P(+_2|+_1) = P(C|+_1) * P(+_2|C) + P(\neg C|+_1) * P(+_2|\neg C)$
  - $= 0.043 * 0.9 + 0.957 * 0.2$
  - $= 0.2301$
- That says, if my first test comes in positive, I expect my second test to be positive with probably 0.2301.

Conditional Independence cont.2

- $B \perp C$ (absolute independence)

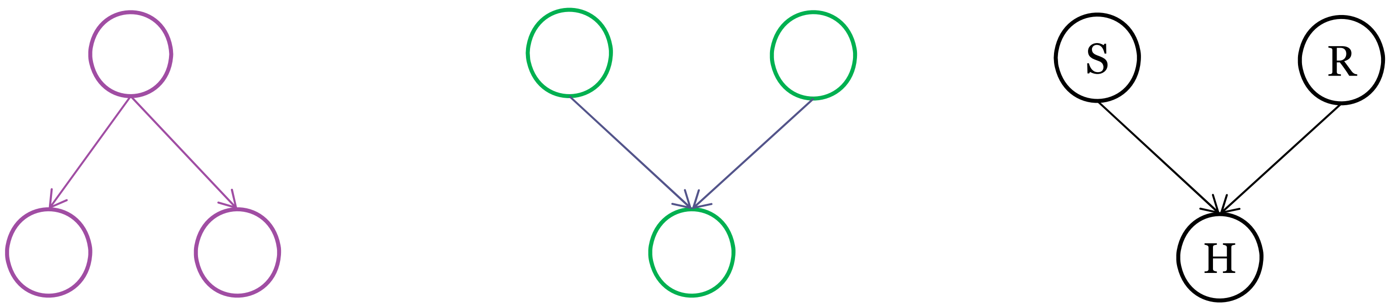
- $B \perp C \mid A$ (conditional independence)



- Does $B \perp C \mid A \overset{implies}{\Longrightarrow} B \perp C$??
  - Yes
  - No
- Does $B \perp C \overset{implies}{\Longrightarrow} B \perp C \mid A$??
  - Yes
  - No

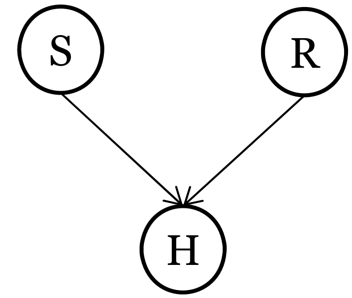
Different Type of Bayes Network

Confounding Cause



- Two independent hidden causes confounded by a single observational variable.
- Suppose I become happy if:
  - the weather is sunny
  - or I get a raise.

Confounding Cause cont.1

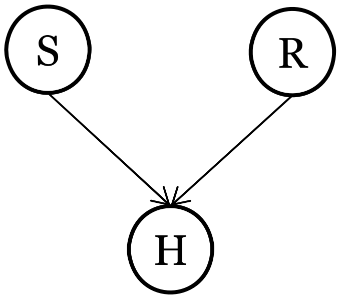


- P(R|S)=??
  - =P(R)=0.01
  - R, S are Independent causes when H is absence
- Explain away Effect
  - Seeing one cause can explain away any other potential causes. E.g. if it rains and I'm happy this explain away that I got a raise.
- P(R|H, S)= ??
  - Remember Bayes rule $\left(\frac{P(R) * P(H|R)}{P(H)}\right)$
  - $= \frac{P(R) * P(H|R,S)}{P(H|S)}$ (BR but with respect to S)
  - Remember the total probability $(P(R) * P(H|R) + P(\neg R) * P(H|\neg R))$
  - $= \frac{P(R) * P(H|R,S)}{P(R)*P(H|R,S)+P(\neg R)*P(H|\neg R,S)}$ (TP but with respect to S)
  - $= \frac{0.01 * 1}{0.01 * 1 + 0.99 * 0.7} = 0.0142$

- P(S)=0.7
- P(R)=0.01

- P(H|S,R)=1
- P(H|¬S,R)=0.9
- P(H|S,¬ R)=0.7
- P(H| ¬ S,¬ R)=0.1

Confounding Cause cont.2



- P(R|H)= ??
  - $\frac{P(R)*P(H|R)}{P(H)}$ How to calculate P(H)
  - Remember the total probability $(P(R)*P(H|R)+P(\neg R)*P(H|\neg R))$
  - But we have another cause S and it's unknown (may be R)!
  - ∴ P(H) = P(R,S)*P(H|R,S) + P(¬R,S) * P(H| ¬ R,S) + P(R, ¬ S) *P(H|R, ¬ S) + P(¬ R,¬ S) * P(H| ¬ R, ¬ S)
  - P(H) = 0.5245
  - P(R|H) = $\frac{P(R)*P(H|R)}{P(H)}$ (using TB for P(H|R))
  - $= \frac{P(R)*(P(S)*P(H|R,S)+P(\neg S)*P(H|\neg S,R))}{P(H)}$
  - $= \frac{P(R)*0.97}{P(H)}$
  - =0.0185
- Compare this to P(R|H, S)=0.0142
  - Because now we don't know about the weather the chances are higher to have a raise (explain away effect).

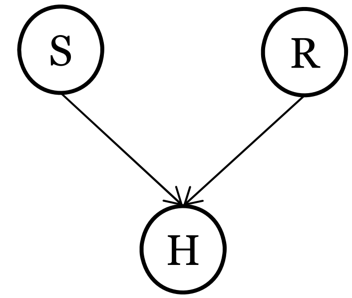
- P(S)=0.7
- P(R)=0.01

- P(H|S,R)=1
- P(H|¬S,R)=0.9
- P(H|S,¬ R)=0.7
- P(H| ¬ S,¬ R)=0.1

Confounding Cause cont.3

- P(R|H, ¬S)= ??
  - $= \frac{P(R) * P(H|R,\neg S)}{P(H|\neg S)}$ (BR but with respect to ¬S)
  - Remember the total probability $(P(R) * P(H|R) + P(\neg R) * P(H|\neg R))$
  - $= \frac{P(R) * P(H|R,\neg S)}{P(R)*P(H|R,\neg S)+P(\neg R)*P(H|\neg R,\neg S)}$ (TP but with respect to ¬S)
  - $= \frac{0.01 * 0.9}{0.01 * 0.9 + 0.99 * 0.1} = 0.0142$
  - = 0.0833

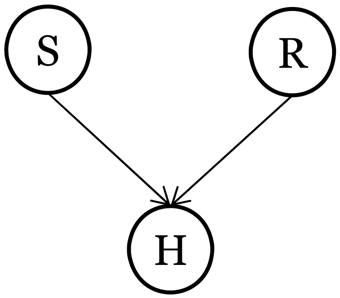
- This to say, I'm happy but it's not sunny. So what the probability for having a raise causing me happy and not any other unknown random reason?



- P(S)=0.7
- P(R)=0.01

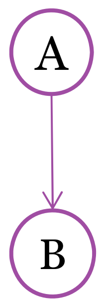
- P(H|S,R)=1
- P(H|¬S,R)=0.9
- P(H|S,¬ R)=0.7
- P(H| ¬ S,¬ R)=0.1

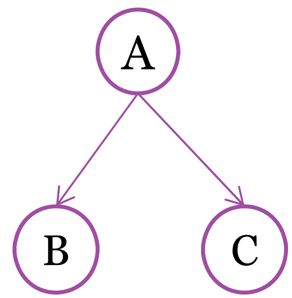
Conditional dependence

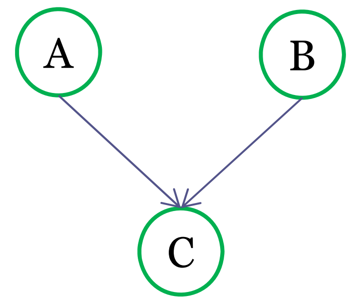


- P(R|H,S) = 0.0142 ≠ P(R|H)=0.0185
- P(R|S)= 0.01 = P(R)= 0.01
- So, if H is given then S and R are conditional dependent on H.

Bayes network different forms



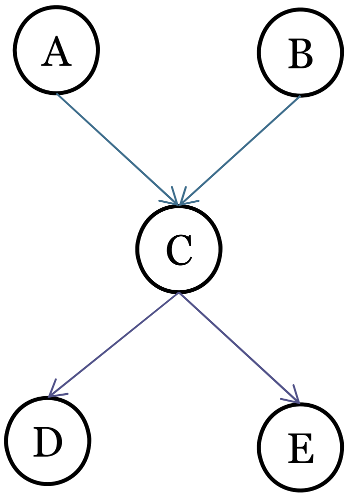




Information:	P(A) P(B|A) P(B| ¬A)	P(A) P(B| A) P(B| ¬A) P(C| A) P(C| ¬A)	P(A) P(B) P(C| A) P(C| ¬A) P(C| B) P(C| ¬B)	
Query:	P(B)	P(B) P(C)	P(C)	Total Probability
	P(A|B)	P(A| B) P(A| C)	P(A| B) P(A| C)	Bayes rule

General Bayes Network

Bayes Network



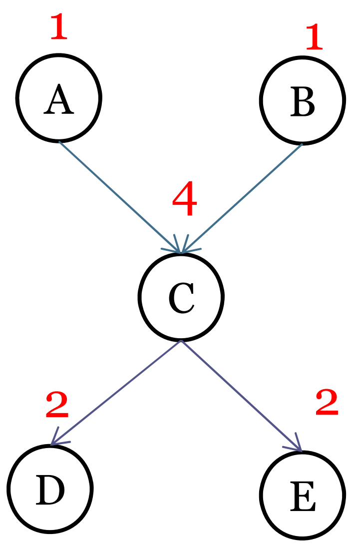
- $P(A), P(B)$
- $P(C|A,B), P(C|\neg A,B), P(C|A,\neg B), P(C|\neg A,\neg B)$
- $P(D|C), P(D|\neg C)$
- $P(E|C), P(E|\neg C)$

- Bayes network is a directed graph of random variables.
- It helps to define the distribution over these random variable.
- The Advantage of Bayes network is:
  - Instead of enumerating all possibilities of combination of these random variables. The Bayes network is defined by probability distribution that are inherent to each individual node.
- For Example to calculate P(A,B,C,D,E) we have two methods:
  - Using joint probability

    $=1-(P(A,B,C,D,\neg E)+\ldots+P(\neg A,\neg B,\neg C,\neg D,\neg E))$ (2^5 -1 parameters)
  - Or by Bayes network which requires only 10 parameters as shown above

    $= P(A)*P(B)*P(C|A,B)*P(D|C)*P(E|C)$ (10 estimated parameters)

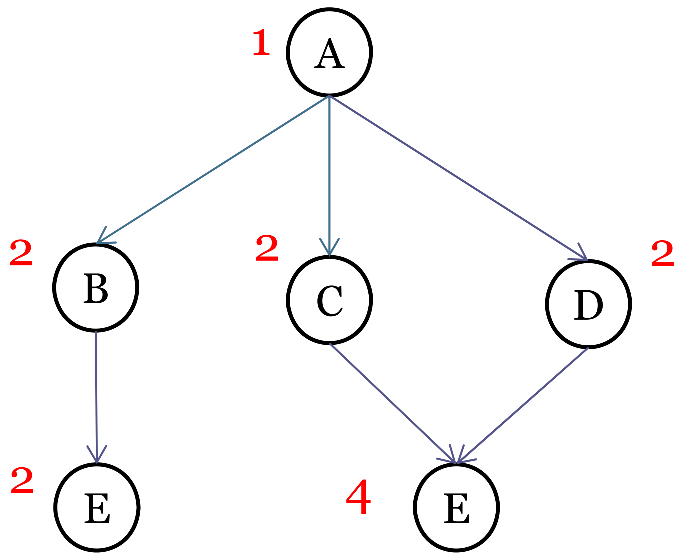
Bayes Network parameters estimation



- P(A), P(B)
- P(C|A,B), P(C|¬A,B), P(C|A,¬ B), P(C| ¬ A,¬ B)
- P(D|C), P(D| ¬ C)
- P(E|C), P(E| ¬ C)

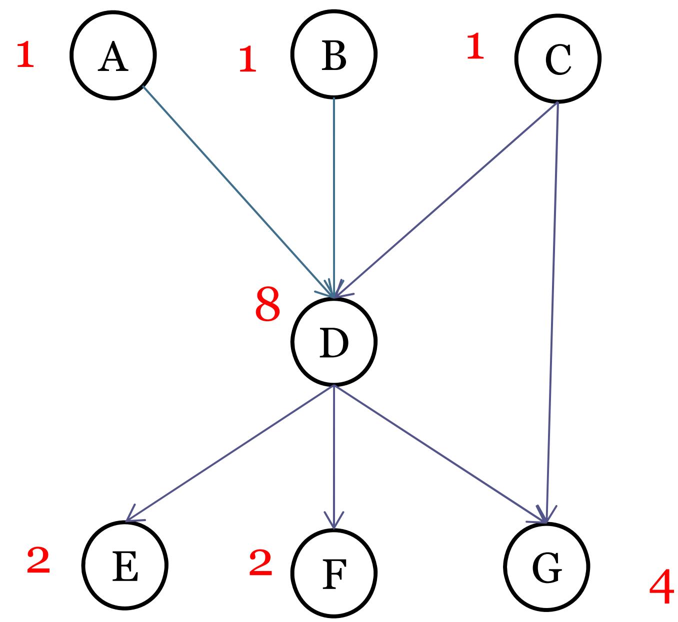
- Each node is conditioned on the incoming arcs.
- So No. of estimated parameters for node $X_i = 2^{(\text{in arcs})}$

Quiz



- How many probability values required to specify this Bayes network ?
  - =13

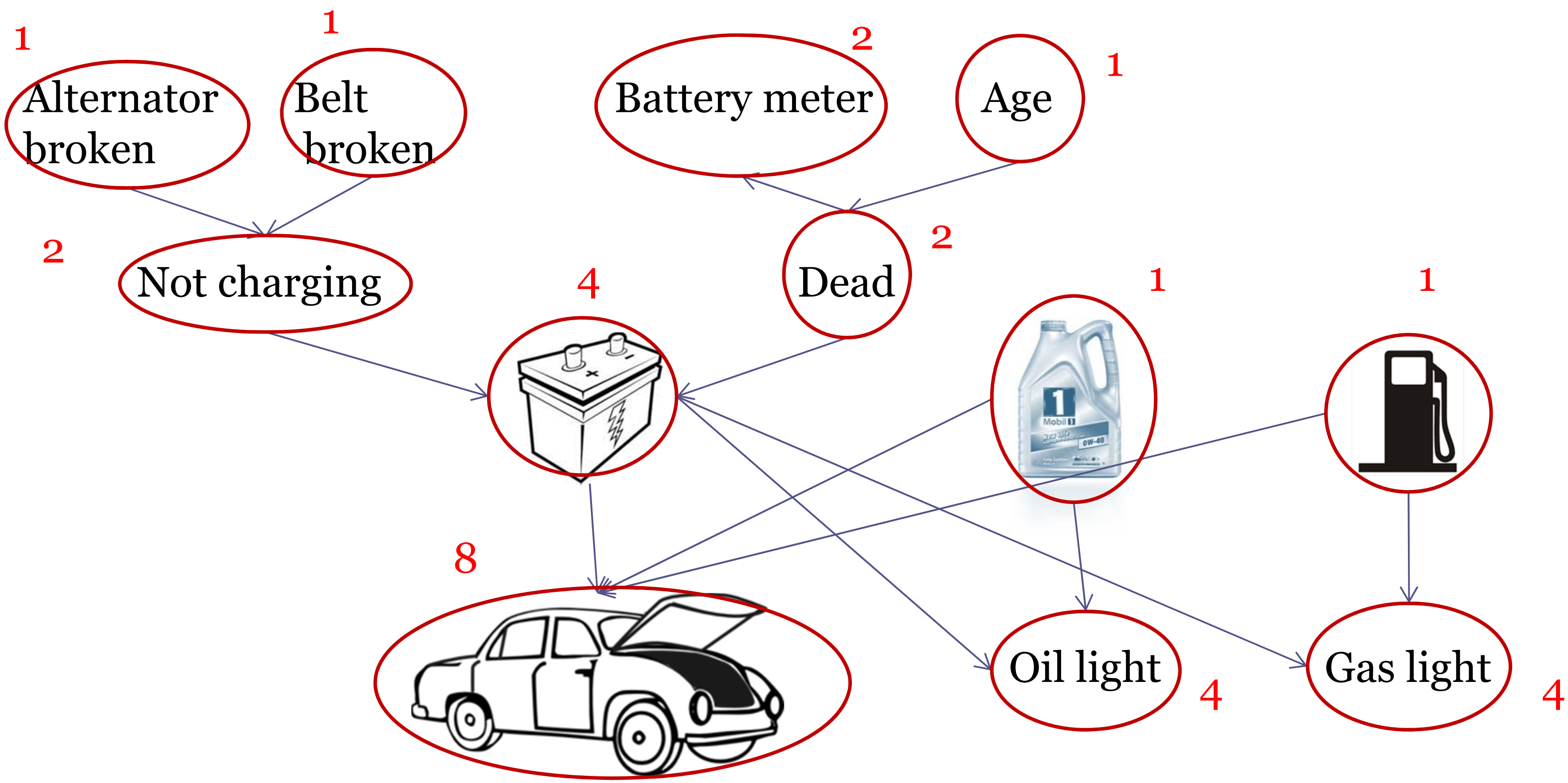
Quiz



- How many probability values required to calculate the full joint probability with Bayes network ?
  - =19

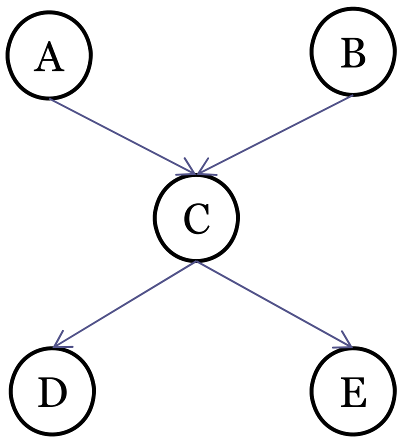
Harvest

Back to car problem



- Using naïve probability requires:
  - 2^{12} =4096-1
- Using Bayes network requires:
  - 31

D-Separation



Independence Check

	Yes	No
$A \perp E$	☐	☑
$A \perp E \mid B$	☐	☑
$A \perp E \mid C$	☑	☐
$D \perp E$	☐	☑
$D \perp E \mid B$	☑	☐
$A \perp B$	☑	☐
$A \perp B \mid C$	☐	☑

- 2 variables are independent if there is a third variable separate them and it's given