Uncertainty – probability

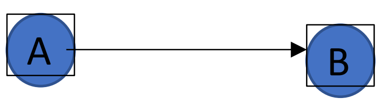
P(A) – marginal probability, formally P(A = a)
P(A, B) - joint probability, formally P(A = a, B = b), P(A) >= P(A, B)
P(A|B) – conditional probability, given B, what is the probability of A,
P(A|B) >= P(A, B), but we don't know P(A), P(B), P(A|B)

P(A, B) = P(A|B) P(B) = P(B|A) P(A)   Bayes Rule

Bayesian Network

Why? A way to simplify the storage or computation (training) when we have lots of variables (events) involved through graph models. Assuming each variable has 3 possible values, there are P(A, …, I) has power(3, 9) combinations.

A	B	C	D	E	F	G	H	I
0	0	0						
1	0							
2	0							



The arrow edge indicates the causal relationship

Both of the two terms P(A|B) and P(B|A) are valid!

During training stage, most of the time P(B|A) is more convenient to obtain.

For example, A is a season variables {0, 1, 2, 3}, B is a weather variable {0, 1, 2}

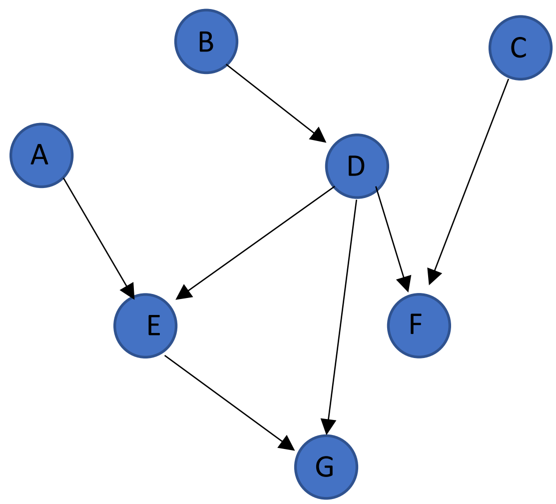
P(B|A) is easier to be trained, e.g. when A = 0, which is Spring season, then you can train how many days of rain, sun shine, and wind. But P(A|B) can be still valid, especially during testing stage. P(B|A)*P(A) = P(A|B) * P(B)

If variables B and E are independent from each other, their joint probability

$P(B, E) = P(B)\,P(E)$ - factorization (decompose the large joint table into multiple products format)

If variables B and E are not independent: $P(B, E) = P(B)P(E|B) = P(E)P(B|E)$

$P(X1, X2, \ldots, Xn) = \text{Product}\{ P(Xi \mid \text{parents}(Xi)) \}_{i = 1, \ldots, n}$



1. P(A, B, C, …, G) what is the joint probability simplified format?
   $P(A, \ldots, G) = P(A)*P(B)*P(C)*P(D|B)*P(E|A, D)*P(F|D, C)* P(G|D, E)$
2. How many sub-tables do you need to build to store the probabilities?
   7 small tables
3. If each variable has 3 possible values, how many combinations totally from the multiple sub-tables?
   Without Bayesian: $3 ^ 7$
   With this Bayesian structure: $3 + 3 + 3 + 3 * 3 + 3 * 3 * 3 + 3 * 3 * 3 + 3 * 3 * 3$

4. $P(D = d1) = P(D = d1|B = b1)$?

   $P(D = d1) = \text{sum}\{P(D = d1 \mid B = bi)\}_{bi} = P(D|B = b1) + P(D|B = b2) + P(D|B = b3)$
   No $P(D= d1) >= P(D=d1|B=b1)$

5. $P(F|D) = P(F|D = d1, C = c2)$?

   $P(F|D) = P(F|D, C = c1) + P(F|D, C = c2) + P(F|D, C= c3)$
   No $P(F|D) >= P(F|D, C)$
6. $P(G|E, D) = P(G|E, D, A)$? Yes
7. $P(G|D) = P(G|F, C, E)$?

P(G|F, C, E) = P(G|E)
P(G|E) = P(G|D) ? It is hard to say (No)

8. P(F|D, C) = P(F|C, D, A, B)? Yes

(Put your answers in a word or .txt file and submit it to the Isidore)

What is the role of Bayesian Network?
It aims to simplify the joint probability, e.g. P(A, B, C, D, E) = P(A)P(B)P(C|A, B)P(D|C)P(E|C, D)

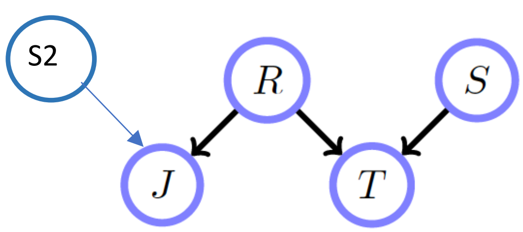
What is the motivation of doing this simplification?
It help to make the marginal or joint probability easier.

$$\mathrm{P(A)} = \sum_{B,C,D,E} \mathrm{P(A, B, C, D, E)} = \sum_{B,C,D,E} \mathrm{P(A)} P(B)P(C|A,B)P(D|C)P(E|C,D)$$

$$= \mathrm{P(A)} \sum_{B} \{\mathrm{P(B)} \sum_{C} \mathrm{P(C|A, B)} \sum_{D} \mathrm{P(D|C)}\ 1\}$$

One morning Tracey leaves her house and realizes that her grass is wet. Is it due to overnight rain or did she forget to turn off the sprinkler last night? Next she notices that the grass of her neighbor, Jack, is also wet. This explains away to some extent the possibility that her sprinkler was left on, and she concludes therefore that it has probably been raining.

T - Tracey's grass (0 – dry; 1 – wet)
J - John's grass (0 – dry; 1 – wet)
S - Tracey's Sprinkler (0 – off; 1 – on)
R - Raining (0 – no rain; 1 – rains)
S2 - John's Sprinkler (0 – off; 1 – on) (Optional)



Case 1:
P(J, R, S, T) = P(R) P(S) P(J|R) P(T|R, S)
Case 2:
P(J, R, S, S2, T) = P(S2) P(R)P(S)P(J|S2, R)P(T|R, S)

(1) Define the remaining possible variables in this problem following the above example.
(2) Draw a Bayesian network model of these variables
(3) Write down your simplified joint probability according to your Bayesian Network
(4) What is the probability that Tracey's sprinkler was on last night? Write a probability equation.

$P(S = 1|T = 1)$
$= P(S = 1, T = 1) / P(T=1)$

Why do we need to convert the conditional probability in such way?
Answer: This is depends on how much probability information do we know after training. If we can only get the joint probability directly based on some background knowledge, then we have do this conversion based on Bayes rule.

$p(R = 1) = 0.2 \quad p(R=0) = 0.8$
$p(S = 1) = 0.1$
$p(J = 1|R = 1) = 1$, $p(J = 1|R = 0) = 0.2$ (because some other factors can cause John's grass wet)
$p(T = 1|R = 1, S = 0) = 1$
$p(T = 1|R = 1, S = 1) = 1$
$p(T = 1|R = 0, S = 1) = 0.9$
$p(T = 1|R = 0, S = 0) = 0$

$= \sum_{J,R} P(S = 1, T = 1, J, R) / \sum_{J,R,S} P(S, T = 1, J, R)$
$= \sum_{J,R} P(R) P(S=1) P(J|R) P(T=1|R, S=1)$
$/ \sum_{J,R,S} P(R) P(S) P(J|R) P(T=1|R, S)$

$\sum_{J,R} P(R) P(S=1) P(J|R) P(T=1|R, S)$
$= \sum_{R} P(R) P(S=1) P(T=1|R, S) \cancel{\sum_{J} P(J|R)}$
$= P(S=1) \sum_{R} P(R) P(T=1|R, S)$
$= 0.1 * \{P(R=0)P(T=1|R=0, S=1) + P(R=1)P(T=1|R=1, S=1)\}$
$= 0.1 * \{0.8 * 0.9 + 0.2 * 1\}$
$= 0.1 * 0.92$
$= 0.092$

$\sum_{J,R,S} P(R) P(S) P(J|R) P(T=1|R, S)$
$=\sum_{R,S} P(R) P(S) P(T=1|R, S)$
$= P(R=0)P(S=0)P(T=1|R = 0, S=0)$
$+ P(R=0)P(S=1)P(T=1|R=0, S=1)$
$+ P(R=1)P(S=0)P(T=1|R=1, S=0)$
$+ P(R=1)P(S=1)P(T=1|R=1, S=1)$

The answer: $P(S=1|T=1) = 0.3382$

$\sum_{J} P(J|R) = P(J=0|R) + P(J=1|R)$

$=(P(J=0|R=0) + P(J=1|R=0))P(R=0)$
$+(P(J=0|R=1) + P(J=1|R=1))P(R=1)$
$= 1.0 * 0.3 + 1.0 * 0.7$
$= 1.0$

$P(R=1) = 0.3$
$P(R=0) = 0.7$

Question: based on the above equations and probability table. Calculate what is the result of $P(S=1|T=1)$.

Marginalization:

$\sum_{A,B,C} P(A,B)P(B|C)P(C)P(B)$
$= \sum_{A,B} P(A,B)P(B) \sum_{c} P(B|C)P(C)$
$= \sum_{B,C} P(B|C)P(C)\, P(B) \sum_{A} P(A,B)$

A –> B <- C -> D
C->E

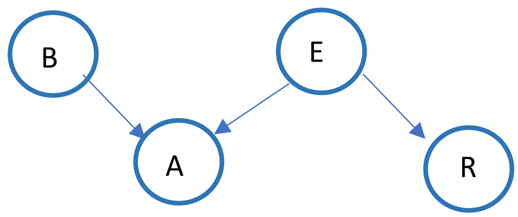
Quiz:
Sally comes home to find that the burglar alarm is sounding (A = 1). Has she been burgled (B = 1), or was the alarm triggered by an earthquake (E = 1)? She turns the car radio on for news of earthquakes and finds that the radio broadcasts an earthquake alert (R = 1).

Alarm = 1	Burglar	Earthquake
0.9999	1	1
0.99	1	0
0.99	0	1
0.0001	0	0

Radio = 1	Earthquake
1	1
0	0

$p(B = 1) = 0.01$ and $p(E = 1) = 0.000001$

$P(B=1|A=1, R=1)$?
(1) Draw a Bayesian graphic model.
(2) Convert this equation into the form made of joint and marginal probability probabilities.
(3) Simplify the equation further using marginalization.
(4) Calculate the final result.



P(A, B, E, R) = P(B)P(A|B, E) P(E)P(R|E)

P(B=1|A=1, R=1) = P(B=1, A=1, R=1) / P(A=1, R=1)

P(B=1, A=1, R=1) = $\sum_E$ P(A, B, E, R)
= $\sum_E$ P(B=1)P(A=1|B=1, E) P(E)P(R=1|E)
= P(B=1) $\sum_E$ P(A=1|B=1, E) P(E)P(R=1|E)
= 0.01* {P(A=1|B=1, E=0)P(E=0)P(R=1|E=0) + P(A=1|B=1, E=1)P(E=1)P(R=1|E=1)}
= 0.01 * {0.99*0.999999*0 + 0.9999 * 0.000001 * 1}

P(A=1, R=1) = $\sum_{E,B}$ P(A=1, B, E, R=1)
= $\sum_{E,B}$ P(B)P(A=1|B, E) P(E)P(R=1|E)
= (E=0, B=0) + (E=0, B=1) + (E=1, B=0) + (E=1, B=1)

(E=0, B=0)
= P(B= 0)P(A=1|B=0, E= 0) P(E=0)P(R=1|E=0) = 0

$p(B = 1) = 0.01$ and $p(E = 1) = 0.000001$

Alarm = 1	Burglar	Earthquake
0.9999	1	1
0.99	1	0
0.99	0	1
0.0001	0	0

Radio = 1	Earthquake
1	1
0	0