Artificial Intelligence

Steve James

Bayesian Networks

Recall

- Joint distributions:
  - $\mathrm{P}(X_1, \dots, X_n)$
  - Use to infer $P(X_1)$, $P(X_1, X_s)$, etc

Raining	Cold	Prob
True	True	0.3
True	False	0.1
False	True	0.4
False	False	0.2

Joint distributions are useful

- All you statistically need to know about $X_1, \dots, X_n$

- Classification
  - $P(X_1 | X_2, \dots, X_n)$

things you know

thing you want to know

- Co-occurrence
  - $P(X_a, X_b)$

how likely are these two things together?

- Rare event detection
  - $P(X_1, \dots, X_n)$

surprising event?

Independence

- If independent, can break JPD into separate tables

Raining	Prob
True	0.6
False	0.4

Cold	Prob
True	0.75
False	0.25

Raining	Cold	Prob
True	True	0.45
True	False	0.15
False	True	0.3
False	False	0.1

Conditional independence

- A and B are conditionally independent given C if:
  - $P(A|B, C) = P(A|C)$
  - $P(A, B|C) = P(A|C)P(B|C)$

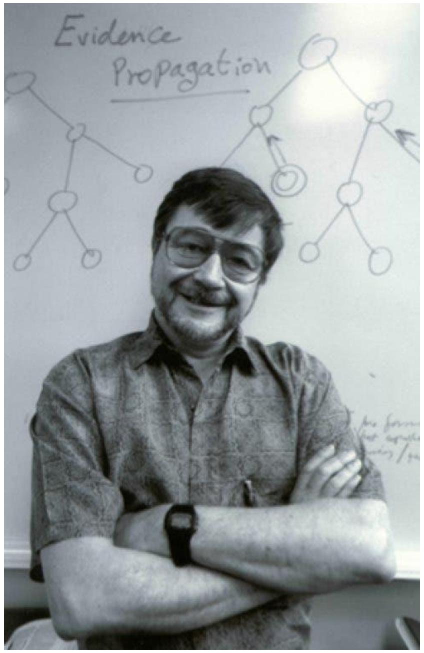
- If we know C, we can treat A and B as if they were independent
  - A and B might not be independent otherwise

Example

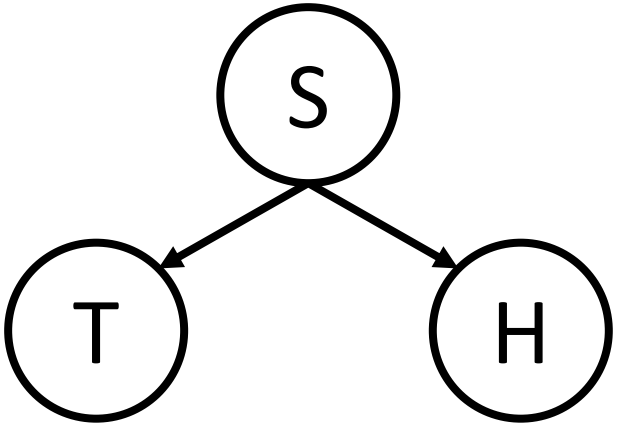
- Consider 3 random variables:
  - Temperature
  - Humidity
  - Season
- Temperature and humidity are not independent

- But they might be given the season
  - Season explains both and they become independent of each other

Bayes Net

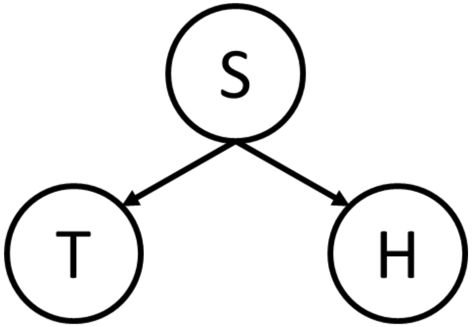


- A particular type of graphical model
  – A directed, acyclic graph
  – Node for each RV



- Given parents, each RV independent of non-descendants

Bayes net

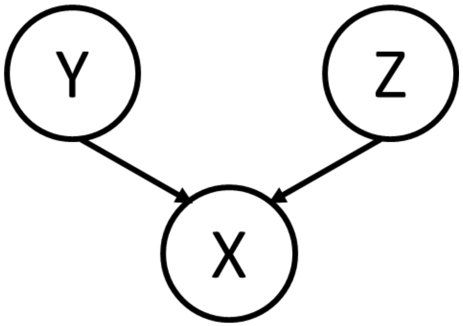


- JPD decomposes

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | parents(x_i))$$

- So for each node, store conditional probability table (CPT): $P(x_i | parents(x_i))$

CPTs



- Conditional probability table
  - Probability distribution over variable given parents
  - One distribution per setting of parents

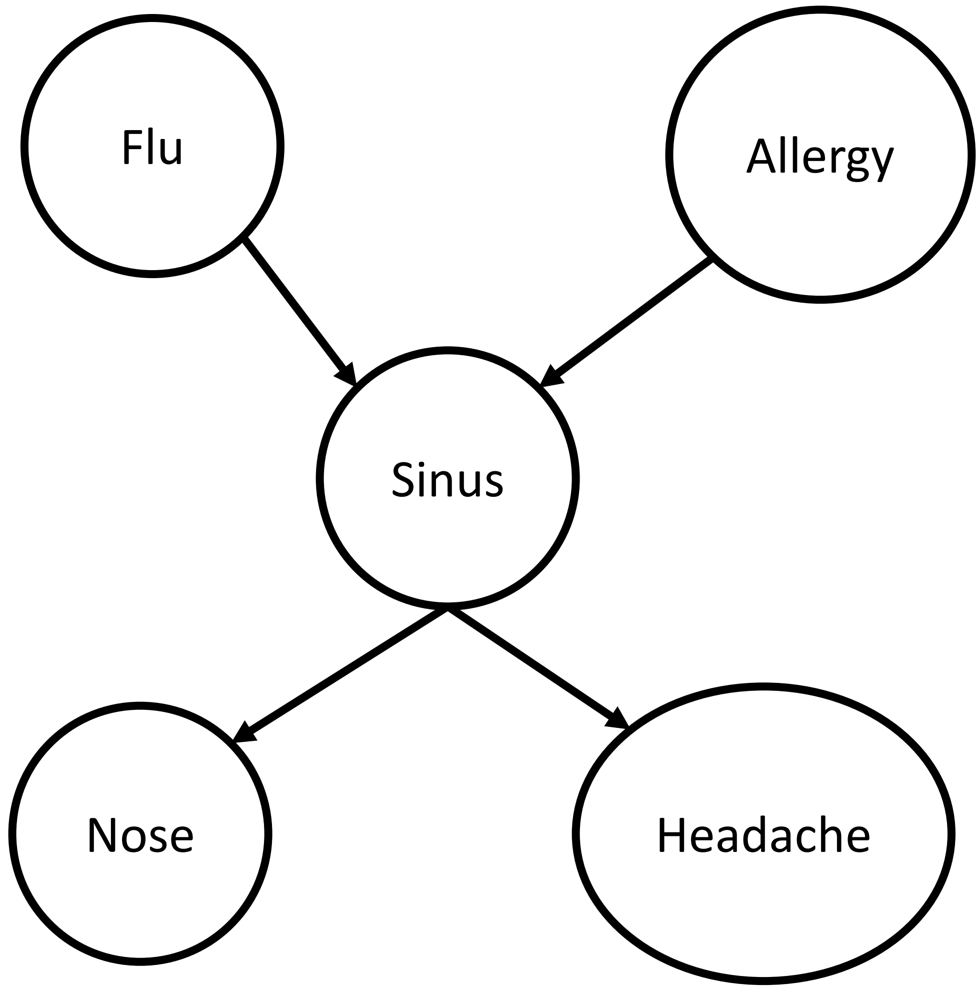
X	Y	Z	P
True	True	True	0.7
False	True	True	0.3
True	True	False	0.2
False	True	False	0.8
True	False	True	0.5
False	False	True	0.5
True	False	False	0.4
False	False	False	0.6

X = variable of interest
Y,Z = conditioning variables

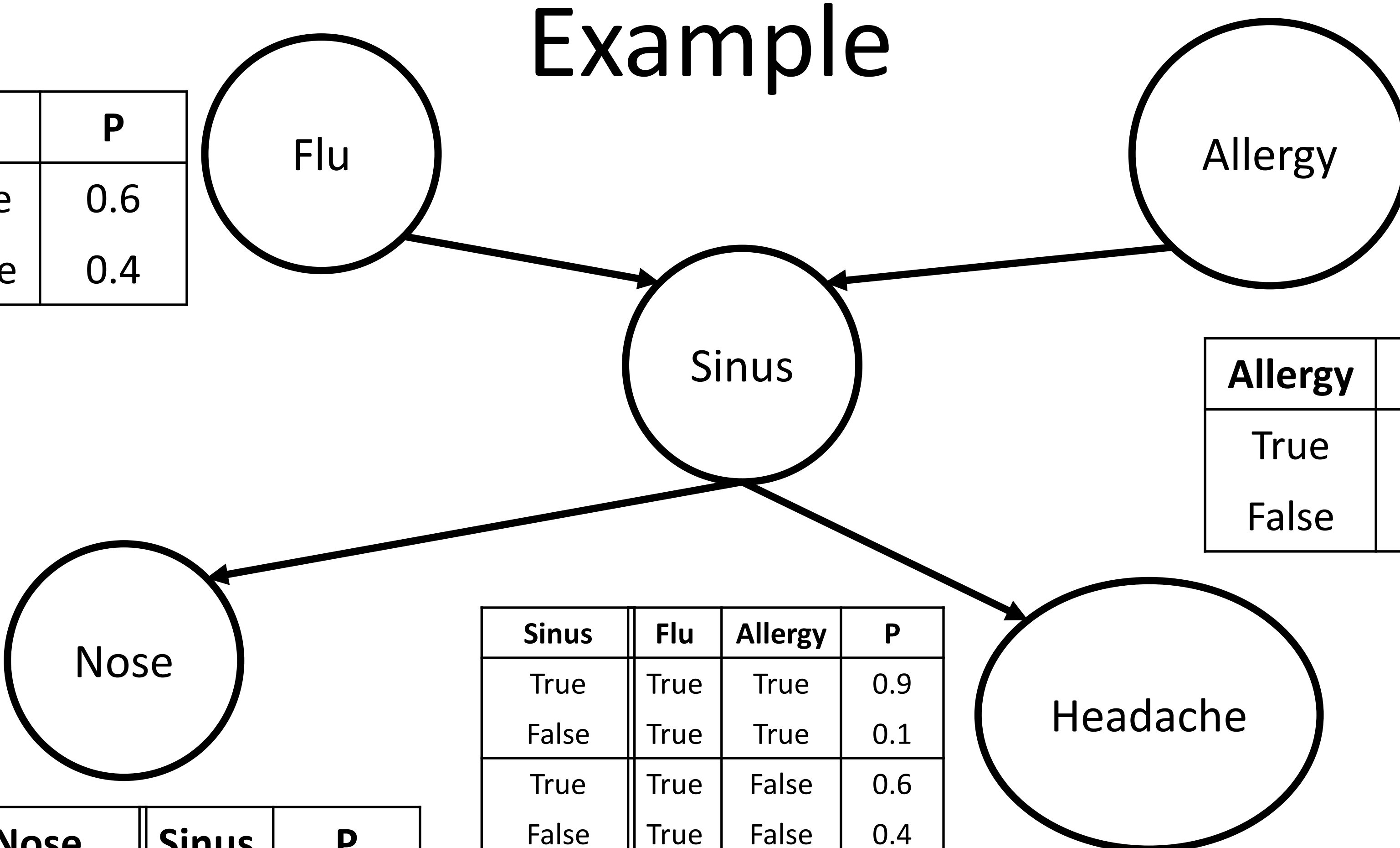
Example

- Suppose we know
  - The flu causes sinus inflammation
  - Allergies cause sinus inflammation
  - Sinus inflammation causes a runny nose
  - Sinus inflammation causes headaches

Example



Example



Flu	P
True	0.6
False	0.4

Allergy	P
True	0.2
False	0.8

Sinus	Flu	Allergy	P
True	True	True	0.9
False	True	True	0.1
True	True	False	0.6
False	True	False	0.4
True	False	True	0.2
False	False	True	0.8
True	False	False	0.4
False	False	False	0.6

Nose	Sinus	P
True	True	0.8
False	True	0.2
True	False	0.3
False	False	0.7

Headache	Sinus	P
True	True	0.6
False	True	0.4
True	False	0.5
False	False	0.5

Joint: 32 (31) entries

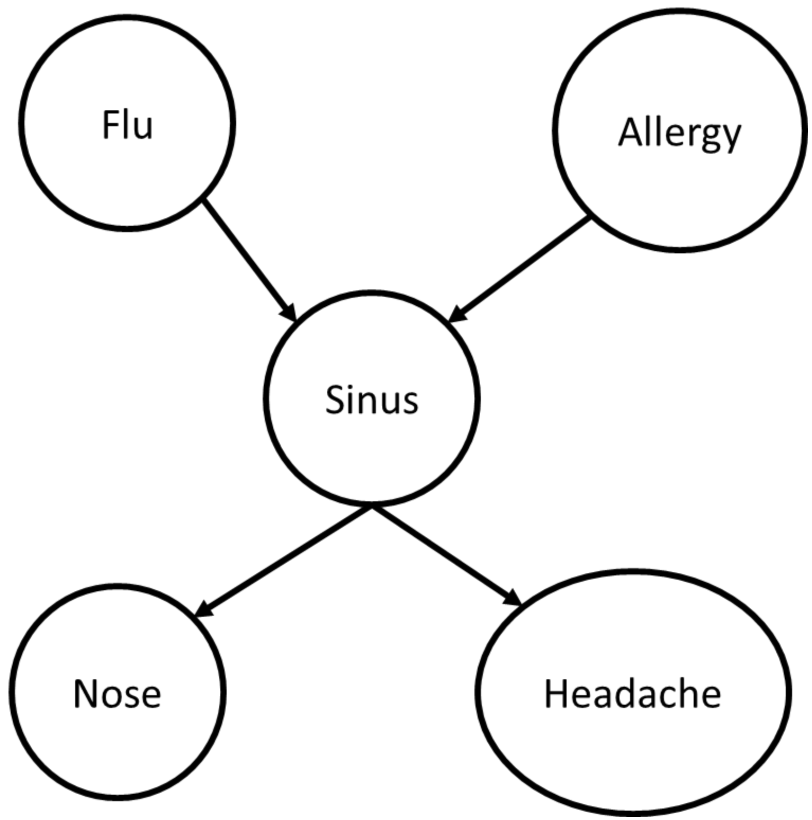
Uses

- Things you can do with a Bayes net
  - Inference: given some variables, posterior?
    - Might be intractable (NP-hard)
  - Learning (fill in CPTs)
  - Structure learning (fill in edges)

- Generally friendly
  - Often few parents
  - Inference cost reasonable
  - Can include domain knowledge

Inference

- What is $P(flu = True|headache = True)$?



Inference

$$P(f|h) = \frac{P(f,h)}{P(h)}$$

- Marginalisation:

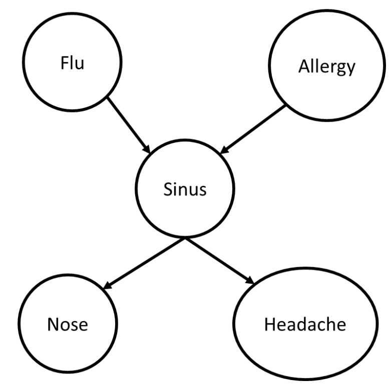
$$P(a) = \sum_{B=T,F} P(a,B)$$

$$P(a) = \sum_{B=T,F} \sum_{C=T,F} P(a,B,C)$$

Inference

$$P(f|h) = \frac{P(f,h)}{P(h)} = \frac{\sum_{SAN} P(f,h,S,A,N)}{\sum_{SANF} P(F,h,S,A,N)}$$

- We know that:



$$P(h) = \sum_{SANF} P(F,h,S,A,N)$$

$$P(h) = \sum_{SANF} P(h|S)P(N|S)P(S|A,F)P(F)P(A)$$

Variable elimination

$$P(h) = \sum_{SANF} P(h|S)P(N|S)P(S|A,F)P(F)P(A)$$

- We can eliminate variables one at a time (distributive law)

$$P(h) = \sum_{SN} P(h|S)P(N|S) \sum_{AF} P(S|A,F)P(F)P(A)$$

$$P(h) = \sum_{S} P(h|S) \sum_{N} P(N|S) \sum_{AF} P(S|A,F)P(F)P(A)$$

Variable elimination

$$P(h) = \sum_{S} P(h|S) \sum_{N} P(N|S) \sum_{AF} P(S|A,F)P(F)P(A)$$

Sinus true

$$0.6 \times \sum_{N} P(N|S = True) \sum_{AF} P(S = True|A,F)P(F)P(A)$$

$$+ 0.5$$ Sinus false

$$\times \sum_{N} P(N|S = False) \sum_{AF} P(S = False|A,F)P(F)P(A)$$

Headache	**Sinus**	**P**
True	True	0.6
False	True	0.4
True	False	0.5
False	False	0.5

Variable elimination

$$P(h) = \sum_S P(h|S) \sum_N P(N|S) \sum_{AF} P(S|A,F)P(F)P(A)$$

$$0.6 \times \left[0.8 \times \sum_{AF} P(S = True|A,F)P(F)P(A) + 0.2 \times \sum_{AF} P(S = True|A,F)P(F)P(A)\right] + 0.5 \times \left[0.3 \times \sum_{AF} P(S = False|A,F)P(F)P(A) + 0.7 \times \sum_{AF} P(S = False|A,F)P(F)P(A)\right]$$

Nose	**Sinus**	**P**
True	True	0.8
False	True	0.2
True	False	0.3
False	False	0.7

Variable elimination

- Cons:
  - How to simplify (hard in general)?
  - Computational complexity
  - Hard to parallelise

- Alternative?
  - Sampling approaches!
    - Based on drawing random numbers
    - Computationally expensive but easy to code
    - Easy to parallelise

Sampling

- From a CPT

- General assumption:
  - Computer can generate (pseudo) random number between 0 and 1

- Generate $x \in [0, 1]$
  - If $x < 0.6$, return True
  - else return False

Flu	P
True	0.6
False	0.4

Generative models

- How do we sample from a Bayes net?
- Bayes net is known as a generative model
- Describes generative process for the data
  - Each variable is generated by a distribution
  - Describes the structure of that generation
  - Can generate more data
- Natural way to include domain knowledge via causality

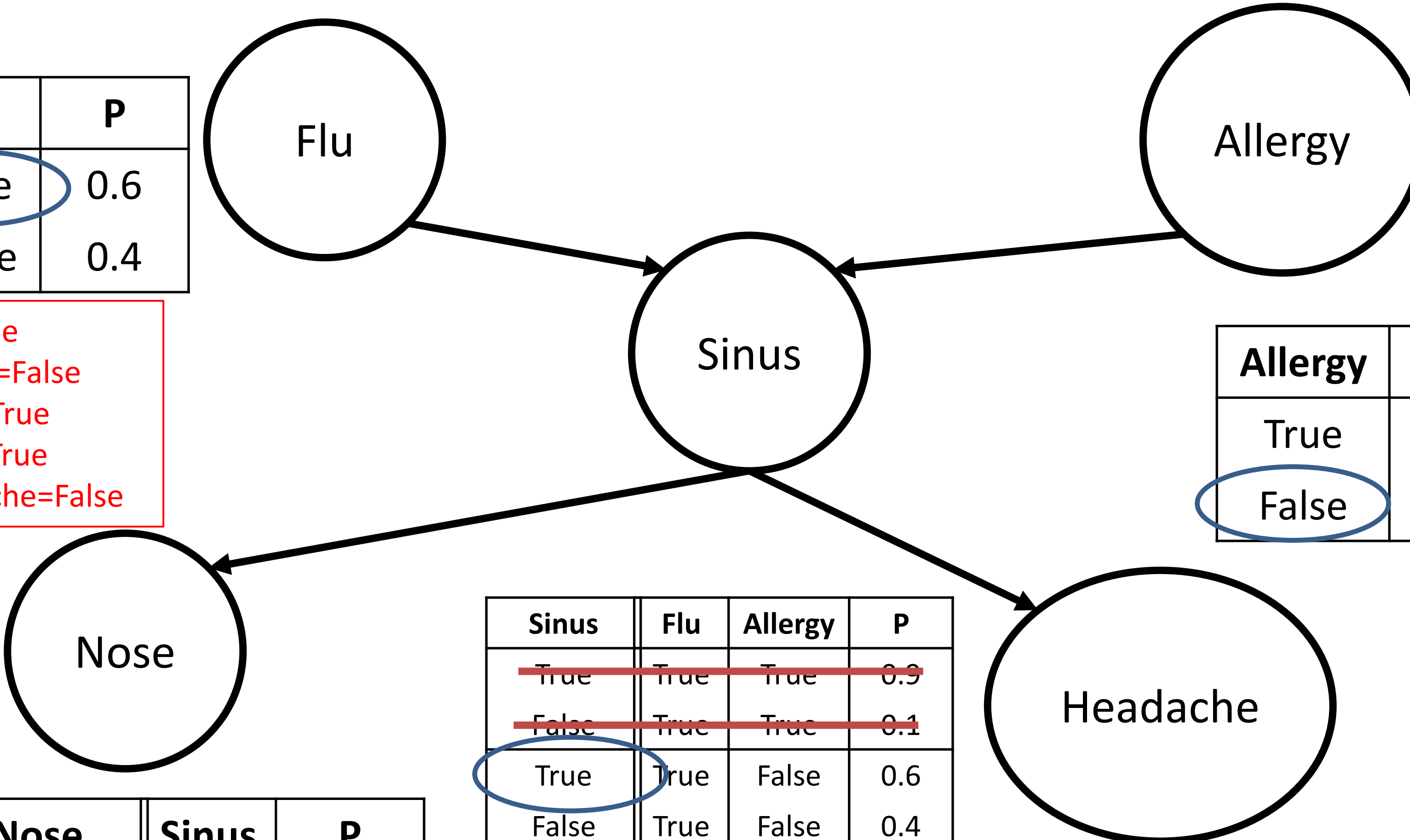
Sampling the joint

- Algorithm for generating samples drawn from the joint distribution
- For each node with no parents:
  - Draw sample from marginal distribution
  - Condition children on choice (removes edge)
  - Repeat
- Results in artificial dataset
- Probability values?
  - Literally just count!

Generative models

Flu	P
True	0.6
False	0.4

Flu=True
Allergy=False
Sinus=True
Nose=True
Headache=False



Allergy	P
True	0.2
False	0.8

Nose	Sinus	P
True	True	0.8
False	True	0.2
~~True~~	~~False~~	~~0.3~~
~~False~~	~~False~~	~~0.7~~

Sinus	Flu	Allergy	P
~~True~~	~~True~~	~~True~~	~~0.9~~
~~False~~	~~True~~	~~True~~	~~0.1~~
True	True	False	0.6
False	True	False	0.4
~~True~~	~~False~~	~~True~~	~~0.2~~
~~False~~	~~False~~	~~True~~	~~0.8~~
~~True~~	~~False~~	~~False~~	~~0.4~~
~~False~~	~~False~~	~~False~~	~~0.6~~

Headache	Sinus	P
True	True	0.6
False	True	0.4
~~True~~	~~False~~	~~0.5~~
~~False~~	~~False~~	~~0.5~~

Sampling the conditional

- What if we want to know $P(A|B)$?
- We could use previous procedure and just divide data up based on B

- What if we want $P(A|b)$?
  - Could do the same, just use data with $B = b$
  - Throw away rest of the data
  - Rejection sampling

Rejection sampling

- Essentially, sample from Q instead of P and then keep if sample was likely to come from P

- But lots of rejecting in high dimensions:
  - If P, Q are Gaussian but P has std dev 1% more
  - Then in 1000 dimensions…
    - Will reject 19999/20000 samples!

Sampling the conditional

- What if b is uncommon?
- What if b involves many variables?

- Importance sampling:
  - Bias the sampling process to get more "hits"
    - New distribution Q
  - Use reweighting trick to unbiased probabilities
    - Multiply by P/Q to get probability of sample

Sampling

- Properties of sampling:
  - Slow: convergence $O(\frac{1}{\sqrt{n}})$
    - But independent of dimension!
  - Always works
  - Always applicable
  - Easy to parallelise
  - Computers are getting faster

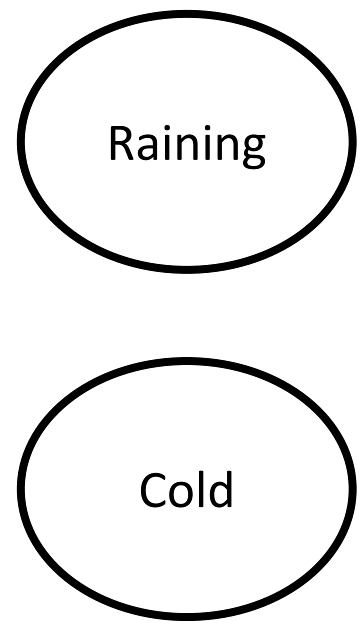
Independence

- What does it look like with a Bayes net?

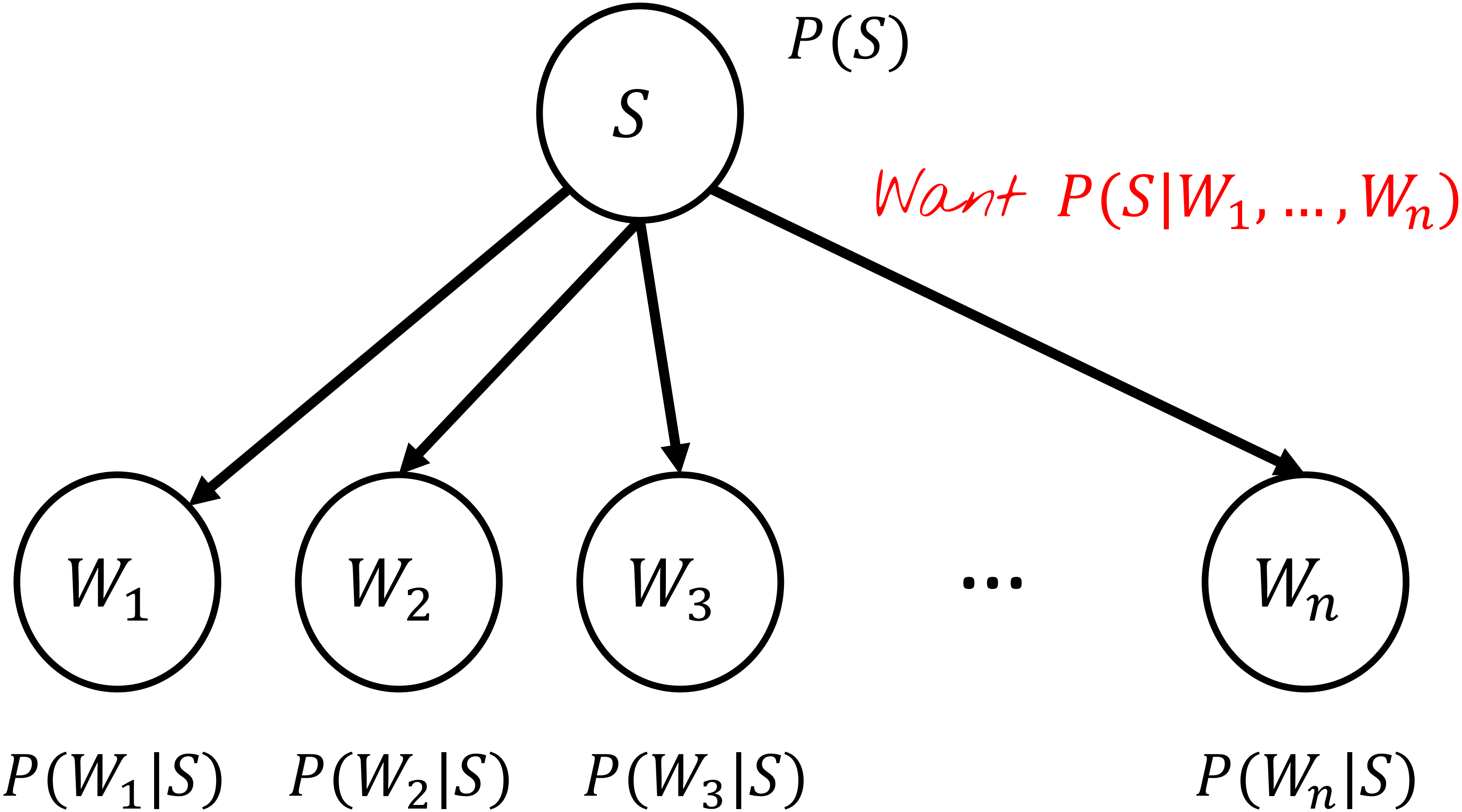
Raining	Prob
True	0.6
False	0.4

Cold	Prob
True	0.75
False	0.25

Raining	Cold	Prob
True	True	0.45
True	False	0.15
False	True	0.3
False	False	0.1



Naïve Bayes



Naïve Bayes

Given

$$P(S|W_1,\ldots,W_n) = \frac{P(W_1,\ldots,W_n|S)P(S)}{\cancel{P(W_1,\ldots,W_n)}}$$

$$P(W_1,\ldots,W_n|S) = \prod_i P(W_i|S)$$

From Bayes net

https://www.infoworld.com/article/2674702/techology-business-paul-graham-provides-stunning-answer-to-spam-e-mails.html

Bayes nets

- Bayes nets are a type of representation
- Multiple inference algorithms; you can choose!
  - AI researchers talk about models more than algorithms

- Potentially very compressed but exact
  - Requires careful construction!

VS

- Approximate representation
  - Hope you're not too wrong!