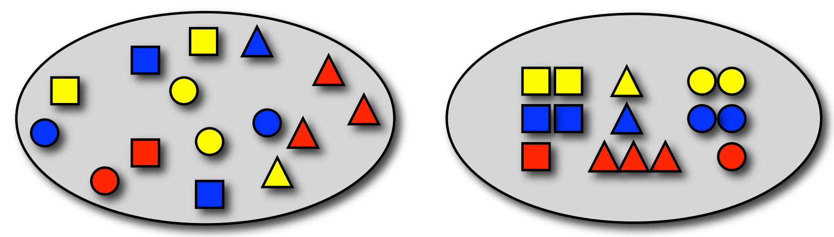
CS440/ECE448: Intro to Artificial Intelligence

Lecture 15: Probability review Bayes Nets intro

Prof. Julia Hockenmaier
juliahmr@illinois.edu

http://cs.illinois.edu/fa11/cs440

What is the probability of…?



P(■) = 2/15
P(blue) = 5/15
P(blue | □) = 2/5

P(■) = 1/15
P(red) = 5/15
P(□) = 5/15

P(■ or ▲) = 2/15
P(△|red) = 3/5

Some terminology…

Trial: e.g. picking a shape

Sample space Ω: the set of all possible outcomes (e.g. all kinds of shapes)

Event ω ⊆ Ω: an actual outcome of a trial (a subset of *Ω*)

Coin tossing

Bernoulli distribution:

Probability of success (*head*) in single yes/no trial

The probability of *head* is *p*.

The probability of *tail* is *1–p*.

Binomial distribution:

Prob. of getting *k* heads in *n* independent yes/no trials

$$P(k \text{ heads},\ n-k \text{ tails}) = \binom{n}{k} p^k (1-p)^{n-k}$$

Rolling a die

Categorical distribution:

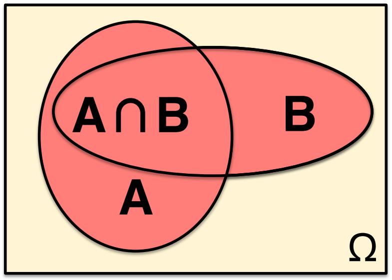
Prob. of getting one of *K* outcomes in a single trial
The probability of outcome c_i is p_i $(\sum p_i = 1)$

Multinomial distribution:

Prob. of observing each possible outcome c_i exactly x_i times in a sequence of *n* yes/no trials

$$P(X_1 = x_i, \ldots, X_N = x_N) = \frac{n!}{x_1! \cdots x_N!} p_1^{x_1} \cdots p_N^{x_N} \quad \text{if} \sum_{i=1}^{N} x_i = n$$

Laws of probability



$P(\Omega) = 1$

$\forall A \subseteq \Omega: \quad 0 \leq P(A) \leq 1$

$\forall A,B \subseteq \Omega: P(A \cap B) \leq P(A)$

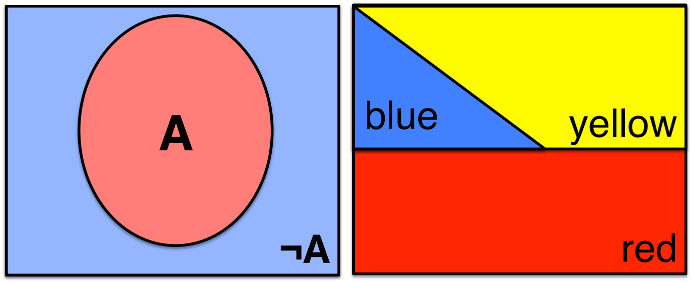
$\forall A,B \subseteq \Omega: P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Random variables

A function which maps every element in the sample space to some value.

Boolean random variables: heads or tails?
Categorical random variables: color, shape
Continuous random variables: size, height,…

Discrete random variables



The possible outcomes of discrete random variables (=atomic events) partition the sample space

What is the probability of…

…. a circle when drawing a red shape?

(# of red circles) / (# of red shapes)

Conditional probability: the probability of one event (*circle*) given another (*red*)

…. drawing a red circle?

(# of red circles) / (# of all shapes in bag)

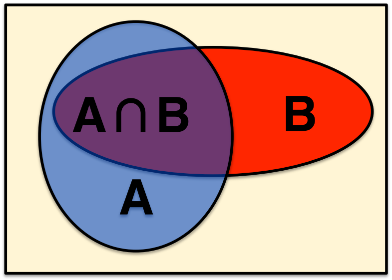
Joint probability: the probability of two events (*red* and *circle*) occurring together

Joint probability P(A,B)

$P(A \cap B) = P(A, B)$

If A and B are boolean variables:

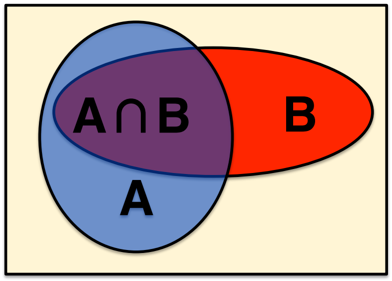
$P(A,B) = P(A \wedge B)$



Conditional probability P(A|B)

Bayes rule:

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$



Product rule

$P(A,B) = P(A \mid B)P(B)$

Conditional probability

Probability of A given B:

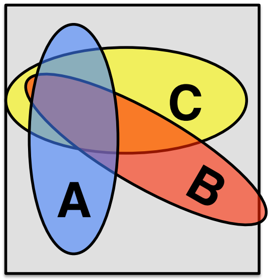
$P(A \mid B)$

Probability of A and B given C:

$P(A,B \mid C)$

Probability of A given B and C:

$P(A \mid B, C)$



Conditional probability

If A and B are events (A: 'red'), (B:'triangle'),
P(A | B): probability of event A given event B.
P(red | triangle)

If A and B are random variables (A:'color', B:'shape'),
P(A | B) is a set of distributions (one per value of B):
P(color = red | shape = 'triangle'),
P(color = blue | shape = 'triangle'),
P(color = yellow | shape = 'triangle'),
P(color = red| shape = 'square')
...

Chain rule

Extends the product rule to multiple variables:

$$P(X_1, \ldots, X_n) = P(X_1) \times P(X_2 \mid X_1) \times P(X_3 \mid X_{1..2}) \times \ldots \times P(X_i \mid X_{1..i-1}) \times \ldots \times P(X_n \mid X_{1..n-1})$$

The parameters of a distribution

How many numbers do we need to specify a distribution?

Bernoulli distribution:
1 parameter (two, but one is implied)

Categorical distribution:
N-1 parameters (N, but one is implied)

Parameters of joint distributions

Joint distribution of *N* Bernoulli RVs:
2^N parameters

How many models does a formula with *N* propositional variables have?
Answer: 2^N

Full joint probability distribution

Weather

Have Fun?	Sunny	Cloudy	Rainy	Snowy
Yes	0.25	0.15	0.05	0.13
No	0.05	0.1	0.25	0.02

Probabilities for each atomic event

Marginal Probabilities: P(Sunny), P(Fun), etc (Written in the margins)

Conditional Probabilities: P(Fun|Sunny)

Weather

Have Fun?	Sunny	Cloudy	Rainy	Snowy
Yes	0.25	0.15	0.05	0.13
No	0.05	0.1	0.25	0.02

P(Rainy)

= 0.05 + 0.25

= 0.3

This is the **marginal probability**

Weather

Have Fun?	Sunny	Cloudy	Rainy	Snowy
Yes	0.25	0.15	0.05	0.13
No	0.05	0.1	0.25	0.02

P(Rainy | ¬Sunny)

$$= \frac{P(Rainy \wedge \neg Sunny)}{P(\neg Sunny)}$$

= 0.3 / 0.7 ≈ 0.43

P(Fun | Sunny)

$$= \frac{P(Fun \wedge Sunny)}{P(Sunny)}$$

= 0.25 / 0.3 ≈ 0.83

Weather

Have Fun?	Sunny	Cloudy	Rainy	Snowy
Yes	0.25	0.15	0.05	0.13
No	0.05	0.1	0.25	0.02

Do I prefer Sun or Snow?

P(Fun | Snowy)

= 0.13 / 0.15 ≈ 0.87 So I prefer snow

Marginal distributions

If we only know the full joint P(X,Y), we can still compute P(X):

$$P(X) = \sum_y P(X, Y = y) = \sum_y P(X \mid Y = y) \times P(Y = y)$$

Joint distributions $P(X_1 \ldots X_i \ldots X_n)$

How many parameters does *P(A, B)* have?
Answer: $K_A \times K_B$ parameters

In general:
The joint distribution of n discrete random variables $X_1 \ldots X_i \ldots X_n$ with K_i possible outcomes each has $K_1 \times \ldots \times K_i \times \ldots \times K_n$ parameters.

Parameters of conditional distributions

How many parameters does *P(A |B)* have?

How many distributions does *P(A |B)* stand for?
Answer: K_B; one for each possible value of *B.*

How many parameters does *P(A |B=b)* have?
Answer: K_A; one for each possible value of *A*
(minus one implied parameter)

So, *P(A |B)* has $K_A \times K_B$ parameters

Independence

Two random variables X and Y are independent if $P(X,Y) = P(X) \times P(Y)$

If X and Y are independent: $P(X|Y) = P(X)$:

$$P(X \mid Y) = \frac{P(X,Y)}{P(Y)} = \frac{P(X) \times P(Y)}{P(Y)} = P(X)$$

X,Y are independent

Conditional Independence

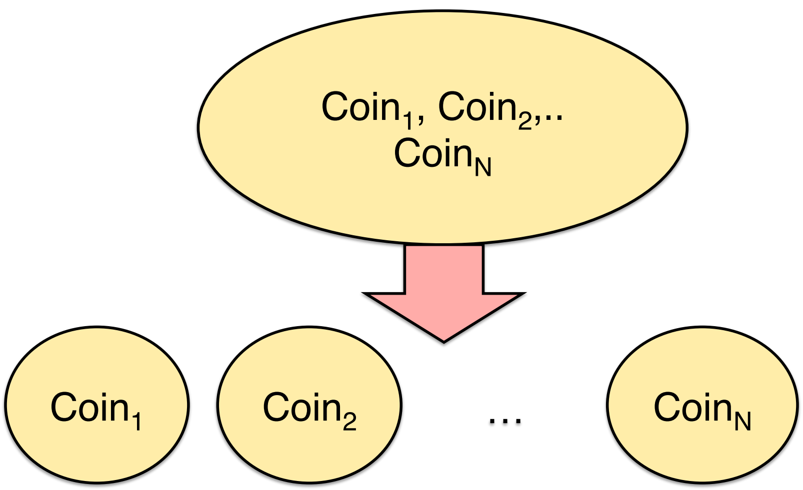
Two random variables X and Y are conditionally independent given Z
if $P(X,Y \mid Z) = P(X \mid Z) \times P(Y \mid Z)$

Independence assumptions

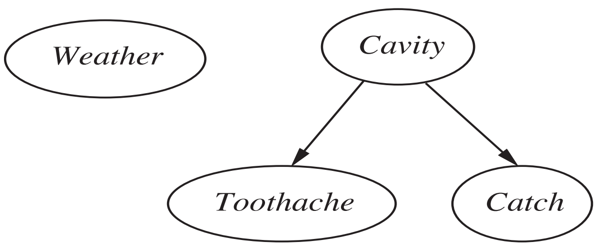
In probabilistic modeling, we often **assume** random variables X, Y, Z are independent

Therefore we can **factor** the distribution:
$P(X,Y,Z) = P(X) \times P(Y) \times P(Z)$

How many parameters do we need to know to specify $P(X,Y,Z)$?
Only $K_X + K_Y + K_Z$



Bayesian networks



We have four random variables
Weather is independent of cavity, toothache and catch
Toothache and catch both depend on cavity.

Bayes Net

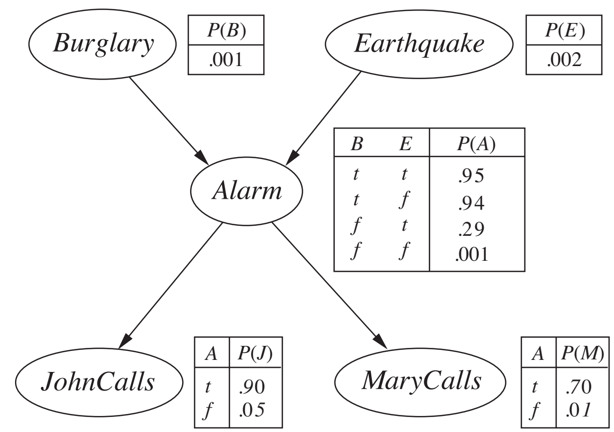
Each random variable is a node.

Each node depends only on its parent.

Each node is conditionally independent of its siblings

Each node specifies a conditional probability table (CPT)

The parameters of a Bayes Net



To conclude…

Today's key concepts

- **CONCEPT:**
  explanation
- **CONCEPT**
  explanation
- **CONCEPT**
  explanation

Your tasks

- **Reading**:
- **Compass quiz**:
- **Assignments**: