Probability

What are the chances? Are you sure?

Matt Eicholtz
Department of Mechanical Engineering
Carnegie Mellon University
meicholt@andrew.cmu.edu

What is the probability of getting in this course?

There are many variables to consider:

- Are you currently on the waitlist?
- What position are you on the waitlist?
- Are you graduating soon?
- Do you need this class to graduate?
- Are we able to move to a bigger classroom?
- Has anyone dropped the course?

Random variables

A **random variable** is an event for which there is uncertainty about whether it occurs

The **domain** is the set of values a random variable can take

Random variables can be **discrete** (boolean, multi-valued) or **continuous**

Examples:

A → you getting in the class (boolean)

→ rolling dice (multi-valued)

→ cards from a deck (multi-valued)

→ distance between points (continuous)

→ temperature (continuous)

What is a probability?

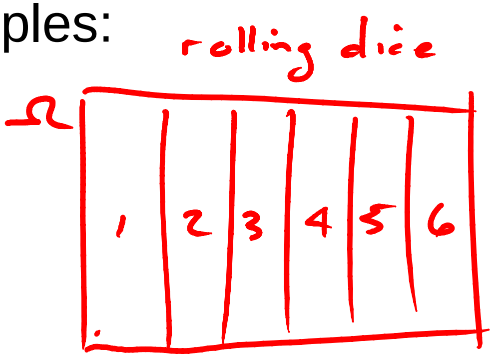
A **possible world** is an assignment of values to the set of relevant random variables

The **search space** is the *exhaustive* set of *mutually exclusive* possible worlds

The **probability** of an event is the fraction of all possible worlds in which the event occurs

A fully specified **probability model** is one in which all possible worlds have an associated numerical probability

Examples:

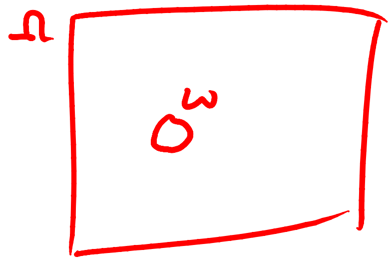


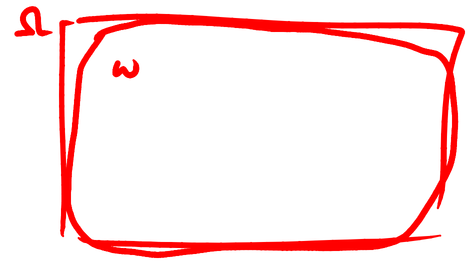
Axioms of Probability

$$0 \le P(\omega) \le 1$$

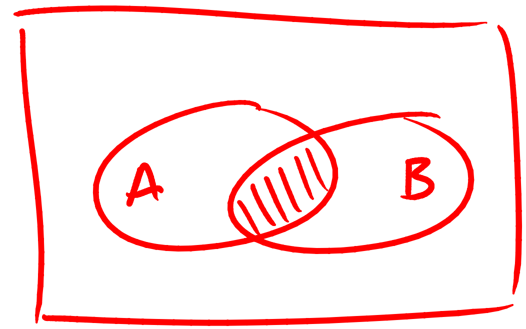
$$\sum_{\omega \in \Omega} P(\omega) = 1$$

$$P(\phi) = \sum_{\omega \in \phi} P(\omega)$$





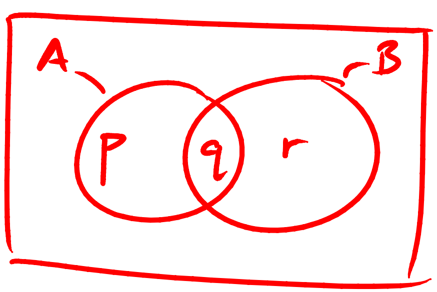
$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$



We can then derive the following relations:

$$P(\neg A) = 1 - P(A)$$

$$\overset{p+q}{P(A)} = P(\overset{q}{A \wedge B}) + P(\overset{p}{A \wedge \neg B})$$



A note on notation

$P(\sim) \longrightarrow$ probability

AND $\Longrightarrow$ $\wedge$, $\cap$,

OR $\Longrightarrow$ $\vee$ $\cup$

NOT $\Longrightarrow$ $\sim$ $\neg$

$|$ $\longrightarrow$ conditional probability

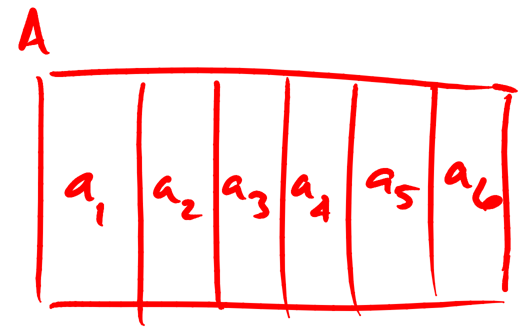
$P(A) = P(A=1)$

$P(\sim A) = P(A=0)$

$P(X=x_i) = P(x_i)$

Multi-valued discrete random variables

A discrete random variable A has **arity** k if it can take 1 of k values $\{a_1, a_2, \ldots, a_k\}$.



Then,

$$P(A = a_i \wedge A = a_j) = 0 \quad \text{if } i \neq j$$

$$P(A = a_1 \vee A = a_2 \vee \cdots \vee A = a_k) = 1$$

Also,

$$P(A = a_1 \vee A = a_2 \vee \cdots \vee A = a_k) = \sum_{j=1}^{k} P(A = a_j)$$

Therefore,

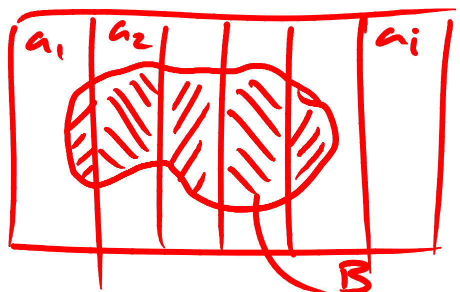
$$\sum_{j=1}^{k} P(A = a_j) = 1$$

Multi-valued discrete random variables

We can also show that

$$P(B \wedge [A = a_1 \vee A = a_2 \vee \cdots \vee A = a_i]) = \sum_{j=1}^{i} P(B \wedge A = a_j)$$

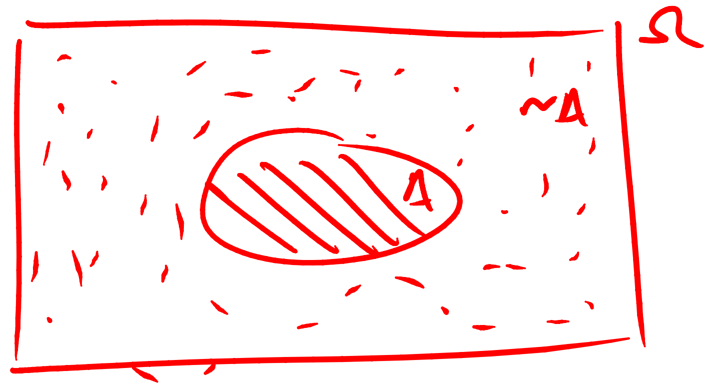
$$= P(B, A=a_1) + P(B, A=a_2) + \ldots + P(B, A=a_i)$$



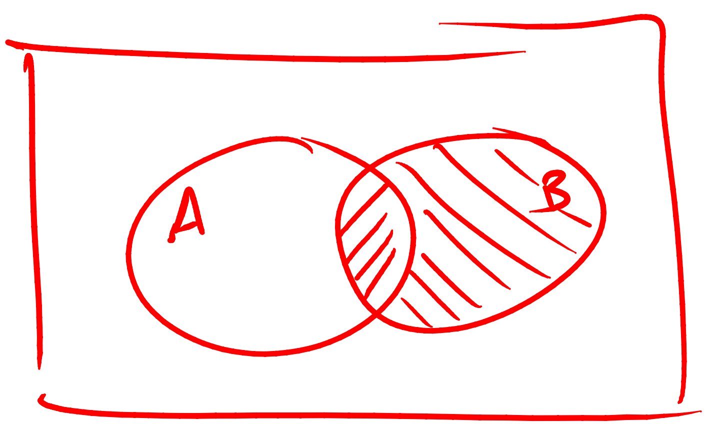
Therefore, for a multi-valued random variable with arity k,

$$\underline{P(B)} = \sum_{j=1}^{k} P(B \wedge A = a_j)$$

$$P(A) + P(\neg A) = 1$$

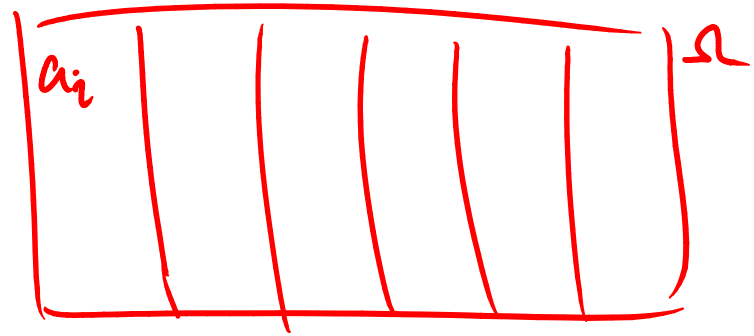


$$P(B) = P(A \wedge B) + P(\neg A \wedge B)$$

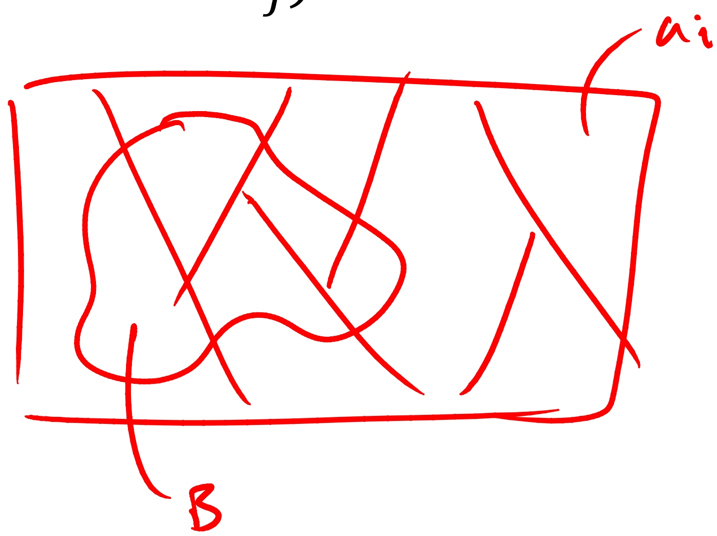


Visualizing Probabilistic Relations

$\sum_{j=1}^{k} P\left(A = a_j\right) = 1$



$P(B) = \sum_{j=1}^{k} P(B \wedge A = a_j)$



Conditional Probability

The probability of A given B, $P(A|B)$, refers to the fraction of possible worlds in which B is true that also have A true

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

Corollary: the chain rule

$$\boxed{P(A,B) = P(A|B)P(B)}$$

Bayes Rule

(1)

$$P(A|B) = \frac{P(A,B)}{P(B)} \quad \Rightarrow \quad P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

(3)

$P(A,B) = P(A|B)P(B)$

(2) $P(A,B) = P(B,A)$

$P(B,A) = P(B|A)P(A)$

$P(A|B)P(B) = P(B|A)P(A)$

$$\underbrace{P(B|A)}_{\text{posterior}} = \frac{P(A|B)P(B)}{\underbrace{P(A)}_{\text{prior}}}$$

More general forms

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(B) = P(B|A)P(A) + P(B|\neg A)P(\neg A)$$

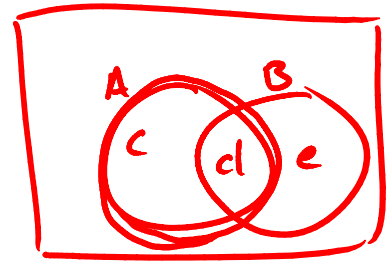
$$= P(A,B) + P(\sim A, B)$$

$$= P(B|A)P(A) + P(B|\neg A)P(\neg A)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}$$

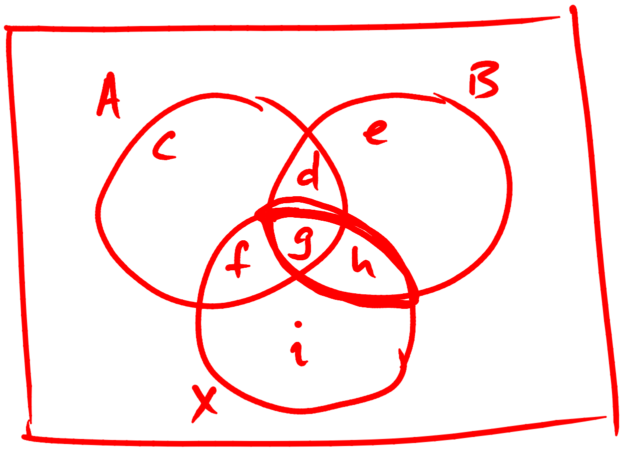
More general forms

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



$P(A|B) = \frac{d}{d+e}$

$P(B|A) = \frac{d}{d+c}$



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

$P(A,B,X) \to P(B|A,X)P(A,X)$

$[A|(B,X)]$

$P(A|B,X) = \frac{g}{g+h}$

$$\boxed{P(A|B,X) = \frac{P(B|A,X)P(A,X)}{\underline{P(B,X)}}}$$

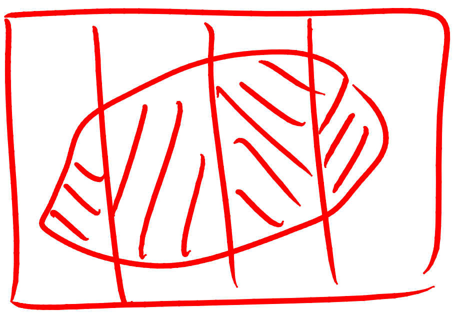
More general forms

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$A \rightarrow$ 3 values

$B \rightarrow$ 2 values

$\underbrace{P(A|B)}_{\text{6 terms}} \Longleftrightarrow \underbrace{P(A=a_i|B)}_{\text{2 terms}} \Longleftrightarrow \underbrace{P(A=a_i|B=b_j)}_{\text{1 term}}$



$$\sum_{j=1}^{k} P(B, A=a_j)$$

$= P(B)$

$$P(A = a_i|B) = \frac{P(B|A = a_i)P(A = a_i)}{\sum_{j=1}^{k} P\big(B\big|A = a_j\big)P(A = a_j)}$$

How do we use all this information?

Probabilistic inference is the process of computing posterior probabilities given observed evidence about the joint distribution.

"Headaches are rare and the flu is rarer, but if you're coming down with the flu, there is a 50-50 chance you'll have a headache."

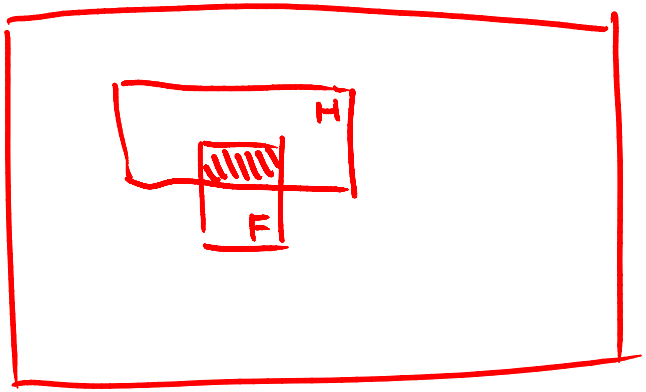
headache

$P(H) = 0.1$ $\left(\frac{1}{10}\right)$

flu

$P(F) = 0.025$ $\left(\frac{1}{40}\right)$

$P(H|F) = 0.5$



$P(H|F) \rightarrow$ fraction of possible worlds in which you have flu that also have a headache.

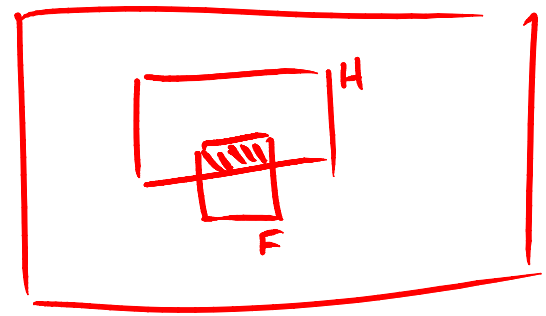
$= \frac{\text{Area of F AND H}}{\text{Area of F}}$

$= 0.5$

Example: Headaches and the Flu

Now suppose you wake up one day with a headache. You think to yourself, "Drat! 50% of flus are associated with headaches so I must have a 50-50 chance of coming down with the flu."

Is this reasoning good? Why or why not?



$$P(H) = 0.1$$
$$P(F) = 0.025$$
$$P(H|F) = 0.5$$

$$P(F|H) \neq P(H|F)$$

$$= \frac{P(F,H)}{P(H)} = \frac{P(H|F)\,P(F)}{P(H)} = \frac{\left(\frac{1}{2}\right)\left(\frac{1}{40}\right)}{\frac{1}{10}} = \frac{1}{8}$$

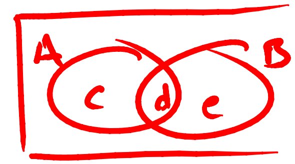
ANS.

Let's try some easy proofs

$P(A|B) + P(\neg A|B) = 1$

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

$$= \frac{P(A,B) + P(\neg A,B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

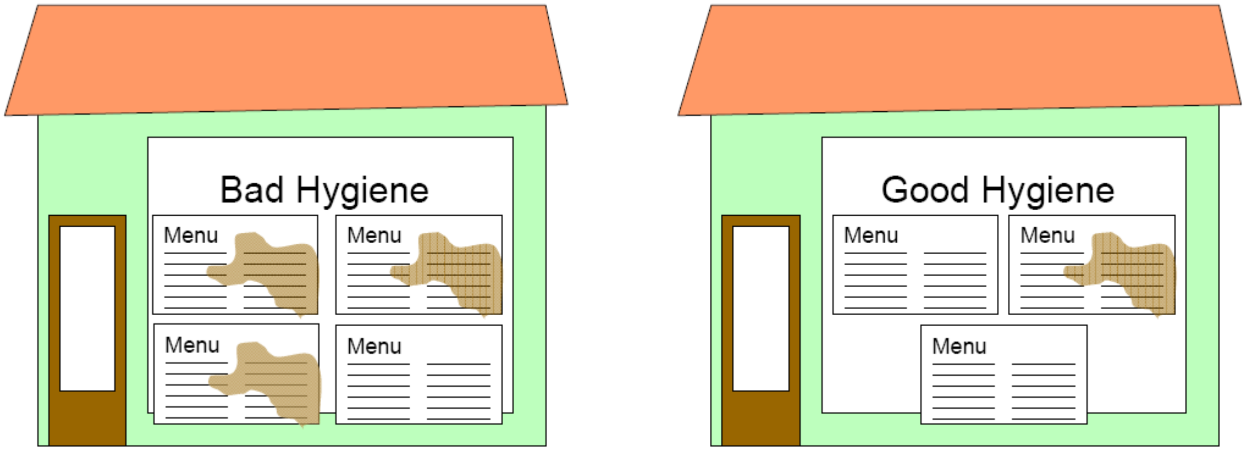


$$\frac{d}{d+e} + \frac{e}{d+e} = 1$$

$\sum_{j=1}^{k} P(A = a_j | B) = 1$

$$\sum_{j=1}^{k} \frac{P(A=a_j, B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

Example: using Bayes Rule to gamble



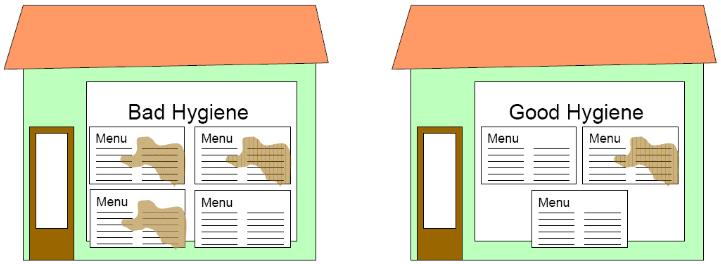
You are a health official deciding whether to investigate a restaurant.

- If you are right, you earn $$$. If you are wrong, you lose $$$.
- Half of all restaurants have bad hygiene. $P(B) = 0.5$
- In a bad restaurant, 3/4 of the menus are smudged. $P(S|B) = 0.75$
- In a good restaurant, 1/3 of the menus are smudged. $P(S|\sim B) = 0.33$
- You are allowed to see one randomly chosen menu.

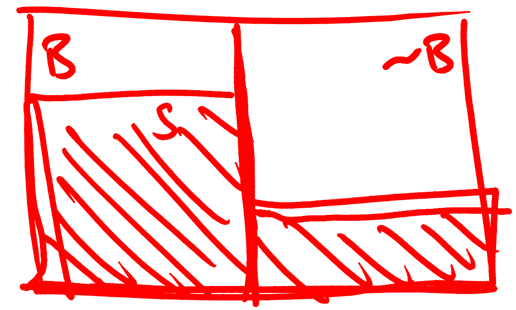
If it is smudged, what is the probability it came from a bad restaurant?

$P(B|S) = ?$

Example: using Bayes Rule to gamble



$$P(B) = 0.5$$
$$P(S|B) = 0.75$$
$$P(S|\sim B) = 0.33$$
$$P(B|S) = ?$$



$$P(B|S) = \frac{P(B,S)}{P(S)} = \frac{P(S|B)P(B)}{P(S,B) + P(S,\sim B)}$$

$$= \frac{P(S|B)P(B)}{P(S|B)P(B) + P(S|\sim B)(1-P(B))}$$

$$= \frac{\left(\frac{3}{4}\right)\left(\frac{1}{2}\right)}{\left(\frac{3}{4}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right)} = \frac{\frac{3}{8}}{\frac{3}{8} + \frac{1}{6}} = \boxed{\frac{9}{13}}$$

NOTE: $P(B|S) + P(\sim B|S) = 1$

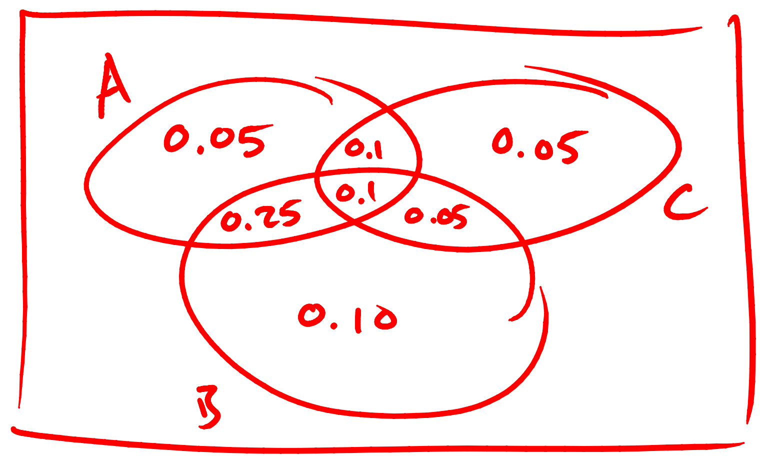
$P(B|S) + P(B|\sim S) \neq 1$

Leave as exercise $\rightarrow$ $P(B|\sim S) = \frac{3}{11}$

What is a joint distribution? Where does it come from? Why is it useful?

1. Make a truth table of all combinations of values of variables → M boolean variables $= 2^M$ probabilities
2. For each combination of values, say how probable it is.
3. If you subscribe to the axioms of probability, the probabilities must sum to 1.

A	B	C	$P(a_i, b_j, c_k)$
0	0	0	0.3
0	0	1	0.05
0	1	0	0.1
0	1	1	0.05
1	0	0	0.05
1	0	1	0.1
1	1	0	0.25
1	1	1	0.1



$$P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

Why is it useful?

$P(A, B) =$ 0.35

$P(B) =$ 0.5

$P(A, C) =$ 0.2

$P(A|B) =$ 0.7

$P(B|\neg A, C) =$ 0.5

Independence

Two random variables A and B are **independent** if

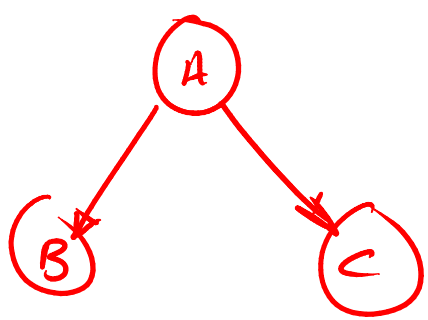
$$P(A|B) = P(A) \qquad P(B|A) = P(B) \qquad P(A,B) = P(A)P(B)$$

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

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

Two random variables A and B are **conditionally independent** given C if

$$P(A|B,C) = P(A|C) \qquad P(B|A,C) = P(B|C)$$



References

Some slides borrowed from Andrew Moore (http://www.autonlab.org/tutorials/) and Levent Burak Kara

Russell and Norvig – Chapter 13; Appendix A

Bishop – Section 1.2

MATLAB Tips and Tricks

Self-assessment

“Helpful” functions to know: *help* and *doc*

Announcements

No lecture on Monday, January 20 (Martin Luther King, Jr. Day)

In-class activity for Wednesday, January 22

- Find a recent (post-2000) media article about AI/ML application
- Print and bring a hardcopy to class