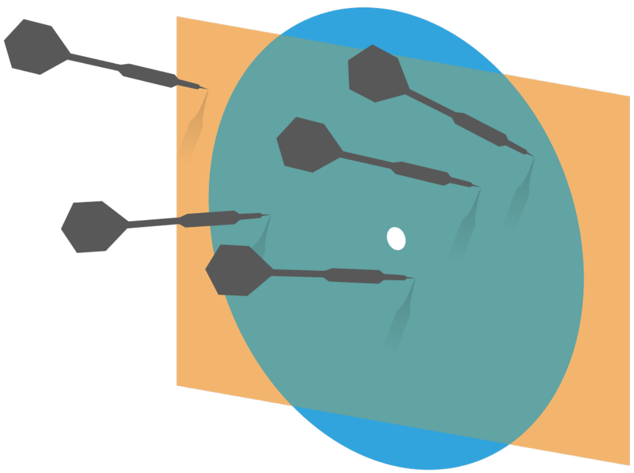
Probability and Statistics for Computer Science



Credit: wikipedia

"A major use of probability in statistical inference is the updating of probabilities when certain events are observed" – Prof. M.H. DeGroot

Hongye Liu, Teaching Assistant Prof, CS361, UIUC, 9.8.2020

Laws of Sets

Commutative Laws

$A \cap B = B \cap A$
$A \cup B = B \cup A$

Associative Laws

$(A \cap B) \cap C = A \cap (B \cap C)$
$(A \cup B) \cup C = A \cup (B \cup C)$

Distributive Laws

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Laws of Sets

Idempotent Laws

$A \cap A = A$
$A \cup A = A$

Identity Laws

$A \cup \emptyset = A$
$A \cap U = A$
$A \cup U = U$
$A \cap \emptyset = \emptyset$

Involution Law $(A^c)^c = A$

Complement Laws

$A \cup A^c = U$
$A \cap A^c = \emptyset$
$U^c = \emptyset$
$\emptyset^c = U$

De Morgan's Laws

$(A \cap B)^c = A^c \cup B^c$
$(A \cup B)^c = A^c \cap B^c$

U is the complete set

Last time

- Probability a first look
  - Outcome and Sample Space
  - Event
  - Probability

    Probability axioms & Properties
  - Calculating probability

Content

- Probability
  - More probability calculation
  - Conditional Probability
  - Independence

Senate Committee problem

The United States Senate contains **two** senators from each of the **50** states. If a committee of eight senators is selected at random, what is the probability that it will contain at least one of the two senators from **IL**?

Probability: Birthday problem

- Among 30 people, what is the probability that at least 2 of them celebrate their birthday on the same day? Assume that there is no February 29 and each day of the year is equally likely to be a birthday.

Conditional Probability

- Motivation of conditional probability

Conditional Probability

- Example:

An insurance company knows in a population of 100 thousands females, 89.835% expect to live to age 60, while 57.062% can expect to live to 80. Given a woman at the age of 60, what is the probability that she lives to 80?

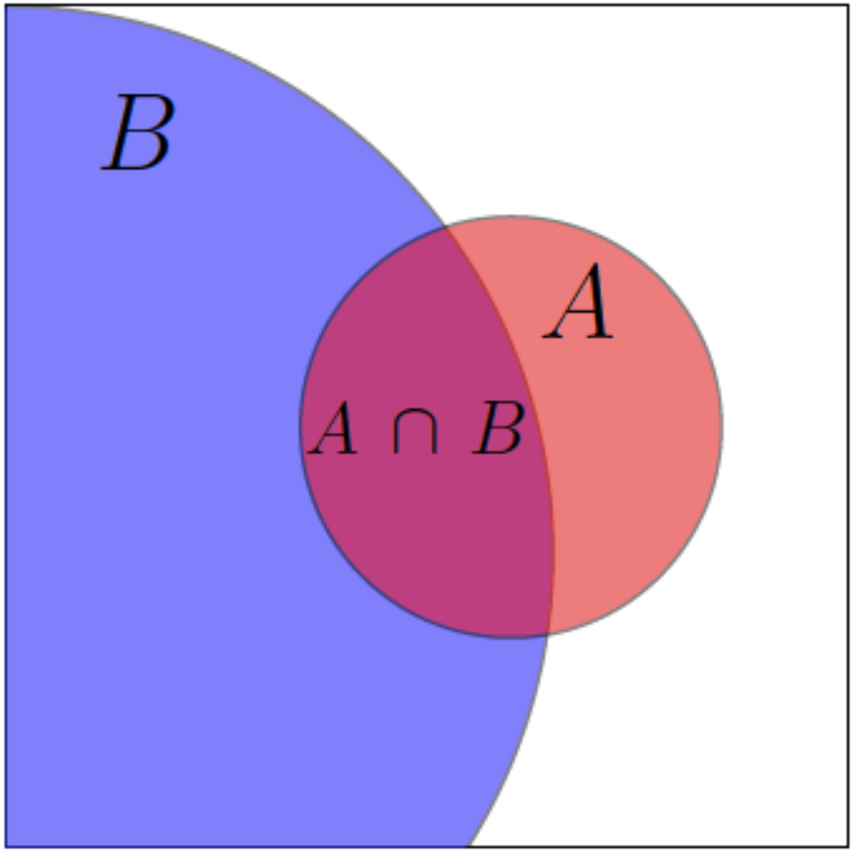
Conditional Probability

- Given the condition she is 60 already, the size of the sample space for the outcomes has changed to

  89,835 instead of 100,000

Conditional Probability

- The probability of ***A*** given ***B***



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

$$P(B) \neq 0$$

The "Size" analogy

Credit: Prof. Jeremy Orloff & Jonathan Bloom

Conditional Probability

A : a woman lives to 80

$$P(A|B) = \frac{57,062}{89,835} = 0.6352$$

B : a woman is at 60 now

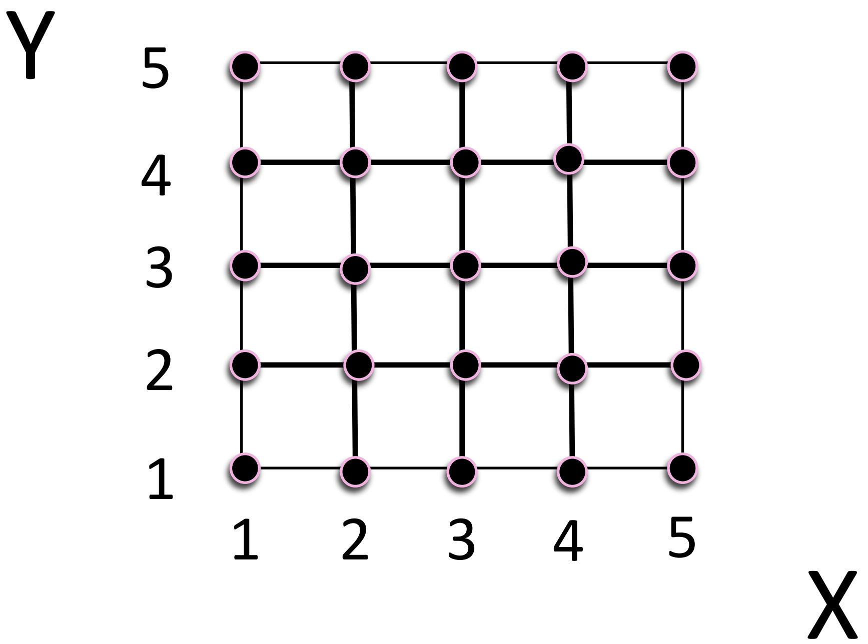
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

While $P(A) = \frac{57,062}{100,000} = 0.57062$

Conditional Probability: die example

Throw 5-sided fair die twice.

$$A : max(X,Y) = 4$$
$$B : min(X,Y) = 2$$



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

Conditional probability

- Now we will see how this formula morphs into many interesting or important formulas

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad P(B) \neq 0$$

Multiplication rule using conditional probability

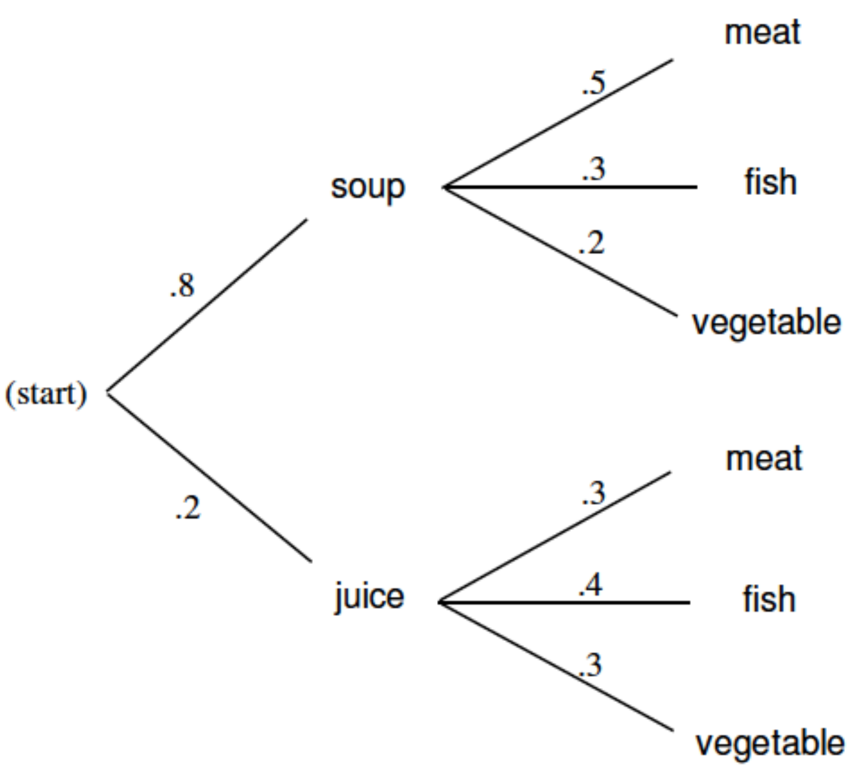
- Joint event

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad P(B) \neq 0$$

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

Multiplication using conditional probability

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



$$P(soup \cap meat) =$$
$$P(meat|soup)P(soup)$$
$$= 0.5 \times 0.8 = 0.4$$

Symmetry of joint event in terms of conditional prob.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad P(B) \neq 0$$

$$\Rightarrow P(A \cap B) = P(A|B)P(B)$$
$$\Rightarrow P(B \cap A) = P(B|A)P(A)$$

Symmetry of joint event in terms of conditional prob.

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

$$\Rightarrow$$

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

The famous **Bayes** rule

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

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

Thomas Bayes (1701-1761)

Bayes rule: lemon cars

There are two car factories, **A** and **B**, that supply the same dealer. Factory **A** produced **1000** cars, of which **10** were lemons. Factory **B** produced **2** cars and both were lemons. You bought a car that turned out to be a lemon. What is the probability that it came from factory **B**?

Bayes rule: lemon cars

There are two car factories, A and B, that supply the same dealer. Factory A produced 1000 cars, of which 10 were lemons. Factory B produced 2 cars and both were lemons. You bought a car that turned out to be a lemon. What is the probability that it came from factory B?

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

Bayes rule: lemon cars

Given the above information, what is the probability that it came from factory A?

$$P(A|L) = ?$$

Bayes rule: lemon cars

Given the above information, what is the probability that it came from factory A?

$$P(A|L) = ?$$

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

Or in this case

$$P(A|L) = 1 - P(B|L)$$

Bayes rule: lemon cars

Given the above information, what is the probability that it came from factory A?

$$P(A|L) = ?$$

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

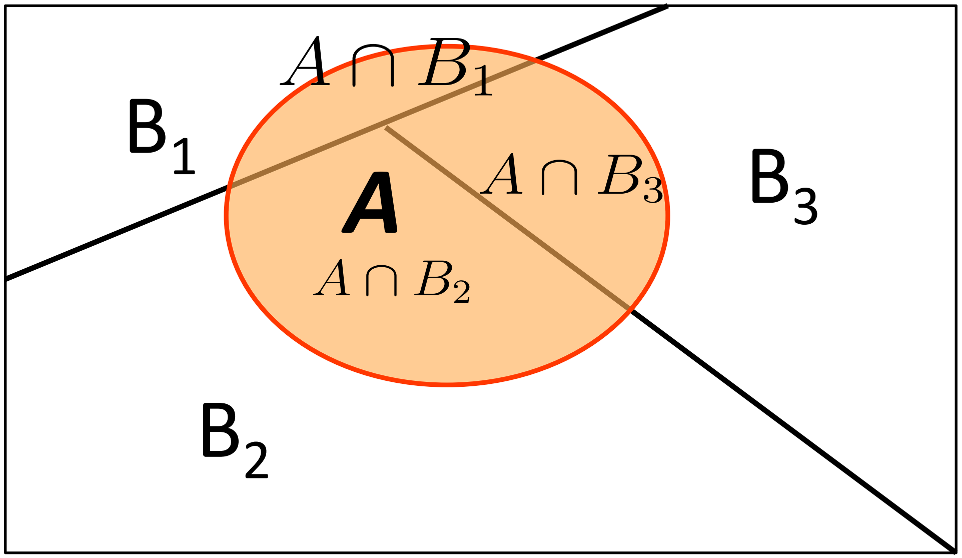
Or in this case

$$P(A|L) = 1 - P(B|L)$$

=

Total probability

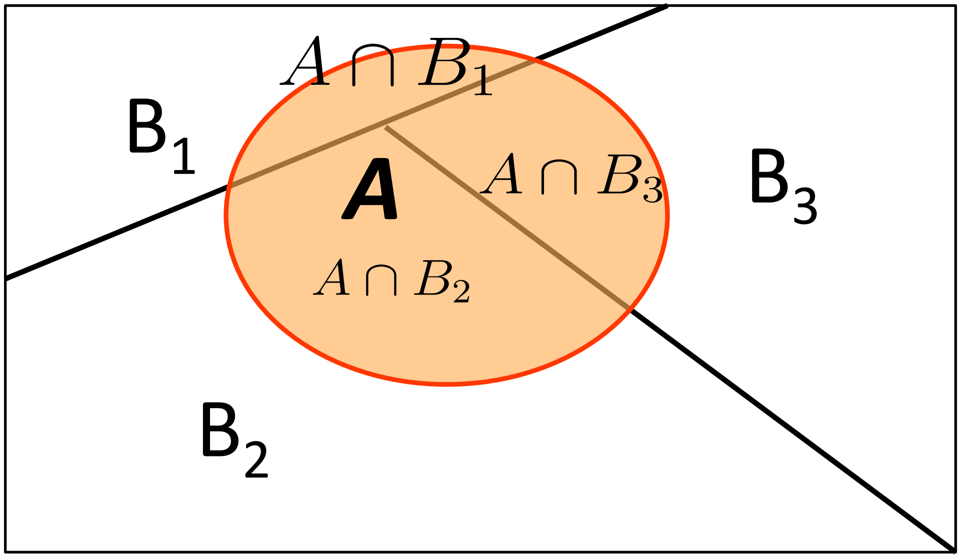
$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$
$$= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$$



Total probability general form

$$P(A) = \sum_j (P(A|B_j)P(B_j))$$

$$if\ B_i \cap B_j = \emptyset\ for\ all\ i \neq j$$



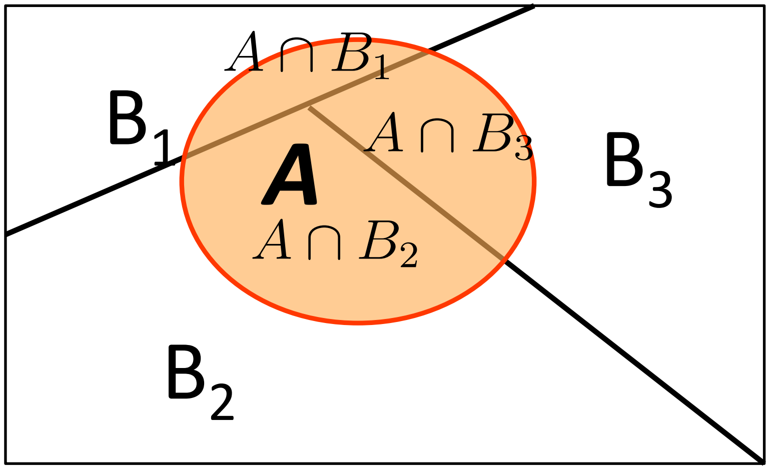
Total probability: candy example

Two boxes contain large numbers of pieces of hard candy in three flavors:
lemon, watermelon and mint. The fractions are as follows:
Box1: 0.3 lemon, 0.4 watermelon, 0.3 mint
Box2: 0.4 lemon, 0.5 watermelon, 0.1 mint
A box is chosen at random with equal probability, then two pieces of candy are chosen from that box randomly. Assume the number of pieces is large enough so that the choice of the first piece does not affect the choice of the second piece. What's the probability that two watermelon pieces are chosen?

Bayes rule using total prob.

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

$$= \frac{P(A|B_j)P(B_j)}{\sum_j P(A|B_j)P(B_j)}$$



Bayes rule: rare disease test

There is a blood test for a rare disease. The frequency of the disease is 1/100,000. If one has it, the test confirms it with probability 0.95. If one doesn't have, the test gives false positive with probability 0.001. What is $P(D|T)$, the probability of having disease given a positive test result?

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)}$$

Using total prob.

$$= \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)}$$

Bayes rule: rare disease test

There is a blood test for a rare disease. The frequency of the disease is **1/100,000**. If one has it, the test confirms it with probability 0.95. If one doesn't have, the test gives false positive with probability **0.001**. What is $P(D|T)$, the probability of having disease given a positive test result?

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)}$$

Independence

- One definition:

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

Whether A happened doesn't change the probability of B and vice versa

Independence: example

- Suppose that we have a fair coin and it is tossed twice. let A be the event "the first toss is a head" and B the event "the two outcomes are the same."

- These two events are independent!

Independence

- Alternative definition

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

$$\Rightarrow \frac{P(A \cap B)}{P(B)} = P(A)$$

$$\Rightarrow \boxed{P(A \cap B) = P(A)P(B)}$$

Testing Independence:

- Suppose you draw one card from a standard deck of cards. E_1 is the event that the card is a King, Queen or Jack. E_2 is the event the card is a Heart. Are E_1 and E_2 independent?

Simulation of Conditional Probability

http://www.randomservices.org/random/apps/ConditionalProbabilityExperiment.html

Additional References

- Charles M. Grinstead and J. Laurie Snell "Introduction to Probability”
- Morris H. Degroot and Mark J. Schervish "Probability and Statistics”

Assignments

- Reading Chapter 3 of the textbook
- Next time: More on independence and conditional probability

Addition material on Counting

Addition principle

- Suppose there are ***n*** disjoint events, the number of outcomes for the union of these events will be the sum of the outcomes of these events.

Multiplication principle

- Suppose that a choice is made in two consecutive stages
  - Stage 1 has ***m*** choices
  - Stage 2 has ***n*** choices
- Then the total number of choices is ***mn***

Multiplication: example

- How many ways are there to draw two cards of the same suit from a standard deck of 52 cards? The draw is without replacement.

Permutations (order matters)

- From 10 digits (0,…9) pick 3 numbers for a CS course number (no repetition), how many possible numbers are there?

Combinations (order not important)

- A graph has N vertices, how many edges could there exist at most? Edges are un-directional.

$$C(n,r) = \frac{n!}{(n-r)!r!} = \frac{P(n,r)}{r!} = C(n, n-r)$$

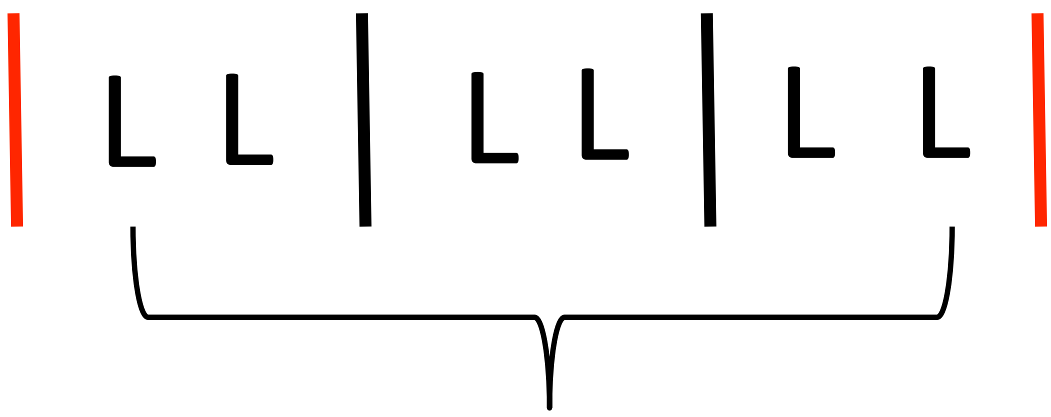
Partition

- How many ways are there to rearrange ILLINOIS?

- General form

Allocation

- Putting 6 identical letters into 3 mailboxs (empty allowed)



Counting: How many think pairs could there be?

- Q. Estimate for # of pairs from different groups. There are **4** even sized groups in a class of 200

Random experiment

- Q: Is the following experiment a random experiment for probabilistic study?

$$2H_{2(g)} + O_{2(g)} \leftrightarrows 2H_2O_{(l)}$$

A. Yes

B. No

Size of sample space

- Q: What is the size of the sample space of this experiment? Deal 5 different cards out of a fairly shuffled deck of standard poker (order matters).

A. C(52,5) **B**. P(52,5) **C**. 52

Event

- Roll a 4-sided die twice

  The event “max is 4” and “sum is 4” are disjoint.

  A. True

  B. False

Probability

- Q: A deck of ordinary cards is shuffled and 13 cards are dealt. What is the probability that the last card dealt is an ace?

  A. 4*P(51,12)/P(52,13) **B**. 4/13

  C. 4*C(51,12)/C(52,13)

Allocation: beads

- Putting 3000 beads randomly into 20 bins (empty allowed)

See you next time

See You!