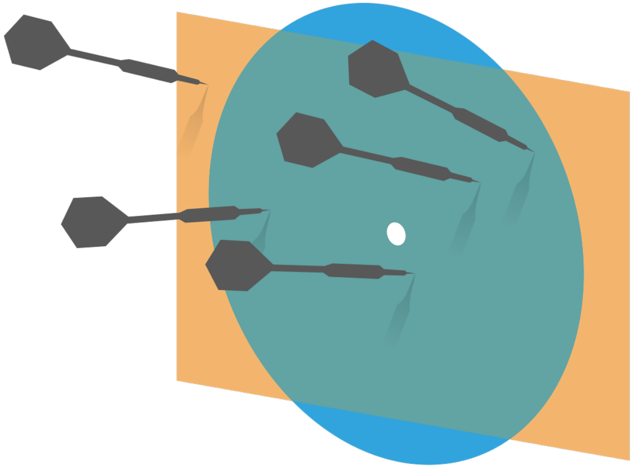
Probability and Statistics for Computer Science



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

Fixed team review Opt out

deadline is today 9/8 @

7 pm central

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 $(A^c)^c = A$

U is the complete set

Complement Laws

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

De Morgan's Laws

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

$P(A \cup B)$

$= P(A) + P(B) - P(A \cap B)$

Warm up

1) Ways of forming a queue with 10 students randomly.

perm

$10!$ $\quad \underline{10}\ \underline{9}\ \underline{8} \cdots \underline{1}$

2) Ways of forming a queue of 5 students randomly from 10 students

K-Perm

$k = 5$

$\frac{10!}{5!}$ $\quad$ A B C D E $\quad$ A C B E D ...

$\underline{10}\ \underline{9}\ \underline{8}\ \underline{7}\ \underline{6}$

$\frac{10!}{5!} = \frac{n!}{(n-k)!}$

3) ways of forming committees of 5 randomly from 10 students

comb.

$\binom{10}{5}$

Which is larger?

1) $\binom{93}{30}$ 2) $\binom{93}{63}$

A. 1) B. 2) C. None

$$\binom{N}{K} = \binom{N}{N-K}$$

Last time

Probability: a first look

Definitions

Random Experiment.

Outcome, Sample Space, Event

Probability – three axioms

Properties of probability

Calculating probability

Objectives

Probability

More probability calculation

- Conditional probability
    * Bayes rule
    * 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**?

100 8

1 − P(none of IL senators are chosen) 98

$$= 1 - \frac{\binom{98}{8}}{\binom{100}{8}}$$

freq. $\frac{\# |E|}{\# |\Omega|}$ $|\Omega|$

1 IL

$$\frac{\binom{2}{1}\cdot\binom{98}{7}}{|\Omega| = \binom{100}{8}} + \frac{\binom{2}{2}\binom{98}{6}}{\binom{100}{8}}$$

2 IL Senators

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.

1 − Prob {none of the people share B-day}

order matters

Hongze Jinglin

$$|\Omega|$$

Hongze	Jinglin
1/1	1/2
1/2	1/1

$$P = 1 - \frac{|E|}{|\Omega|} = 1 - \frac{\frac{365!}{335!}}{365^{30}}$$

30

$$|\Omega| = \underset{s_1}{\underline{365}} \times \underset{s_2}{\underline{365}} \times \underset{s_3}{\underline{365}} \ \_\_ \ \_\_ \ \_\_ \ \_\_$$

$$365^{30}$$

$\{1, 2, \ldots 365\}$

$$|\Omega| = 365^{30}$$

$$|E| = \underline{365} \times \underline{364} \times \underline{363} \ \_\_ \ldots \_\_ \ \_\_$$

all are different

$$\frac{365!}{335!} \rightarrow {}^{365}P_{30}$$

How does it change with # of people

$p = 0.706$

$k = 30$

Table 1.1 The probability p that at least two people in a group of k people will have the same birthday

k	p	k	p
5	0.027	25	0.569
10	0.117	30	0.706
15	0.253	40	0.891
20	0.411	50	0.970
22	0.476	60	0.994
23	0.507		

DeGroot et. al.

What are the differences between these two examples?

Senate Committee, Birthday.

order doesn't matter

matters

Conditional Probability

- Motivation of conditional probability

if test is

+

Disease

?

having
p(Disease given test is positive)

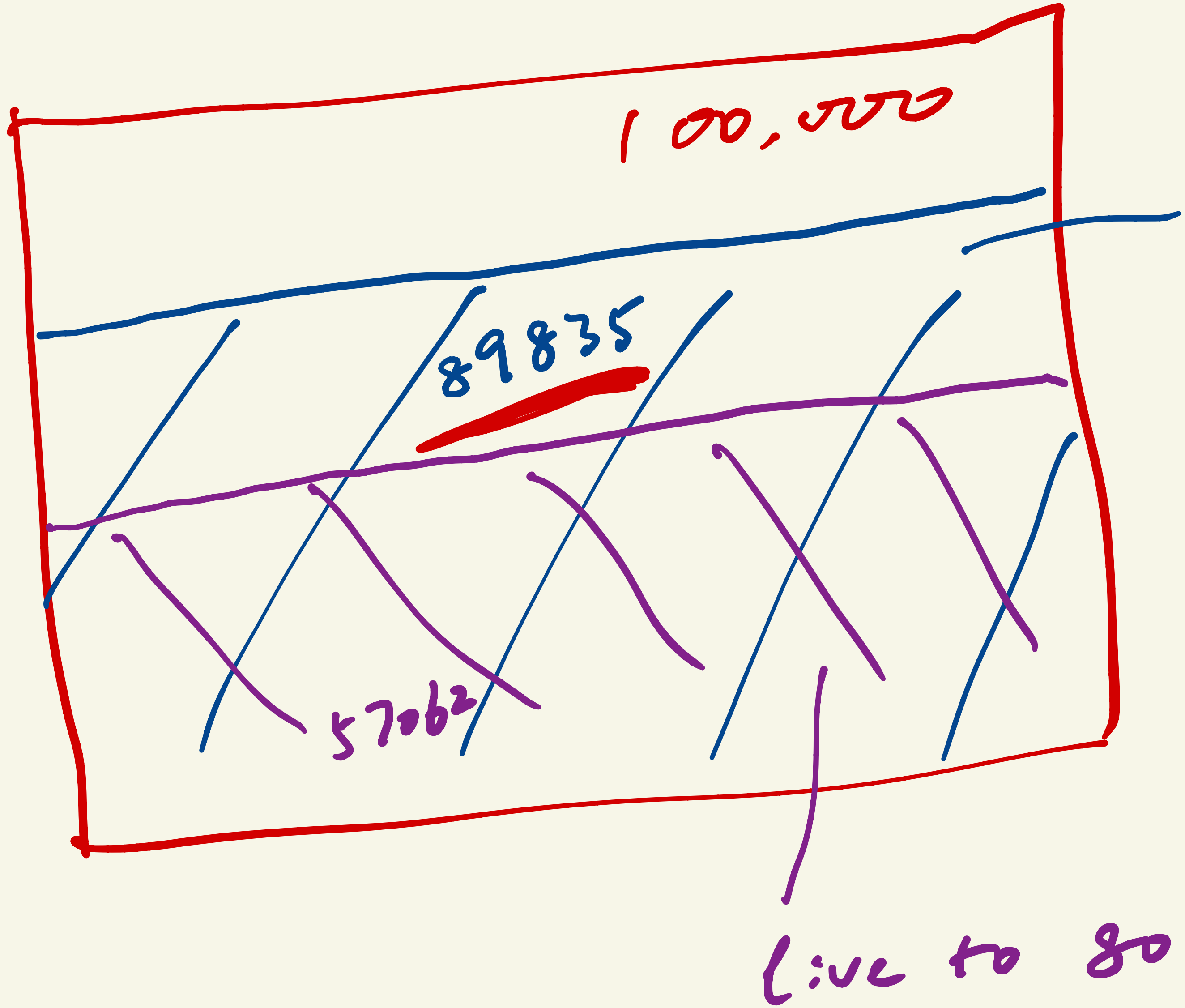
Data

Data driven tool

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?



60

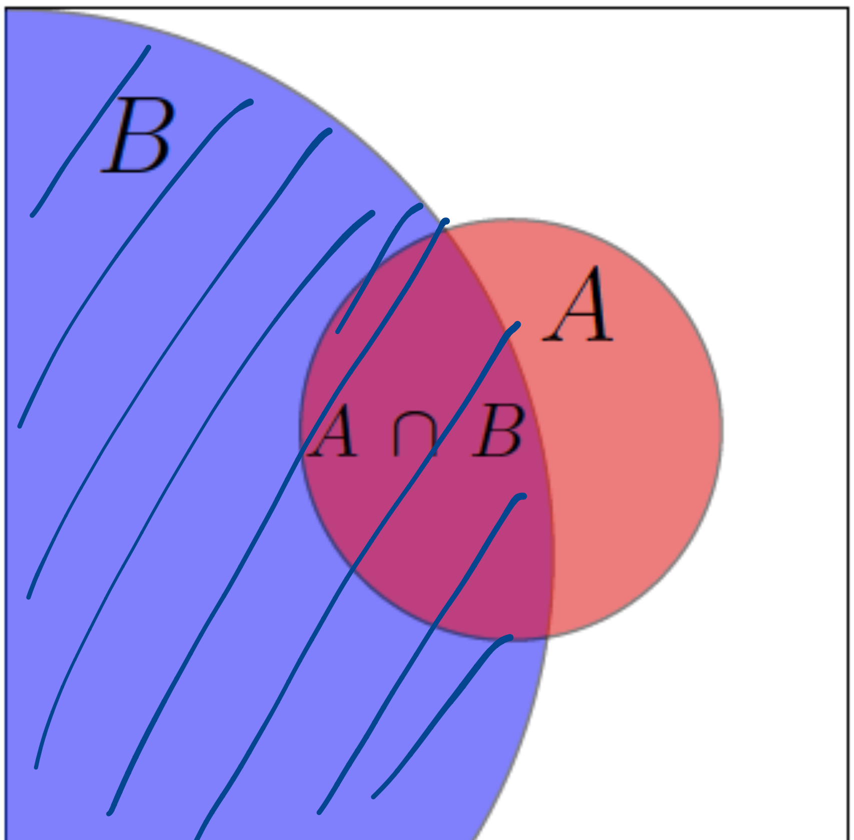
new sample space

1521

= 89,835

Conditional Probability

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



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

assume B happened

$$P(B) \neq 0$$

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

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

$$= \frac{57062/100000}{89835/100000}$$

$$= 0.6352$$

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

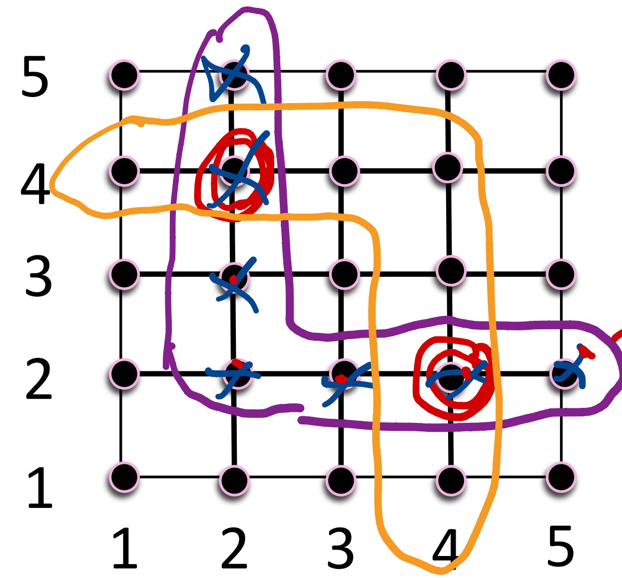
Conditional Probability: die example

Throw 5-sided fair die twice.

$$\underline{A : max(X, Y) = 4}$$

$$\underline{B : min(X, Y) = 2}$$

Y



B

5×5

25

X

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

$\frac{2}{7}$

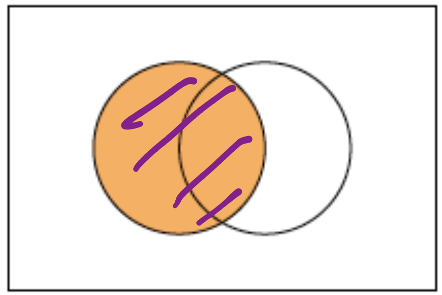
$$\frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{25}}{\frac{7}{25}}$$

Conditional probability, that is?

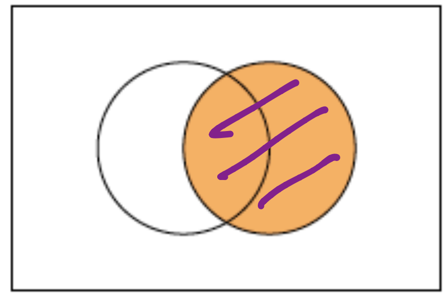
$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad P(B) \neq 0$$

Venn Diagrams of events as sets

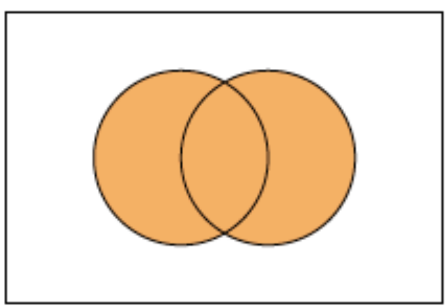
Ω



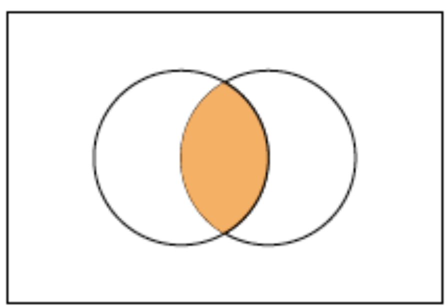
E_1



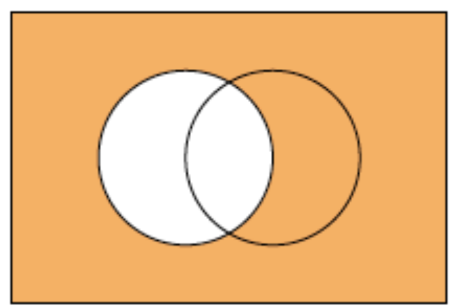
E_2



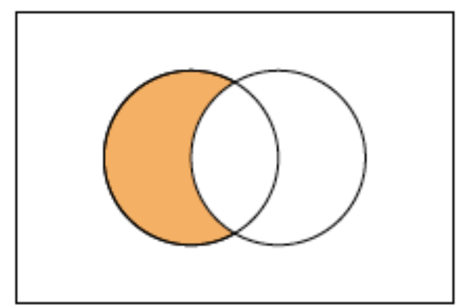
$E_1 \cup E_2$



$E_1 \cap E_2$



${E_1}^c$



$E_1 - E_2$

Multiplication rule using conditional probability

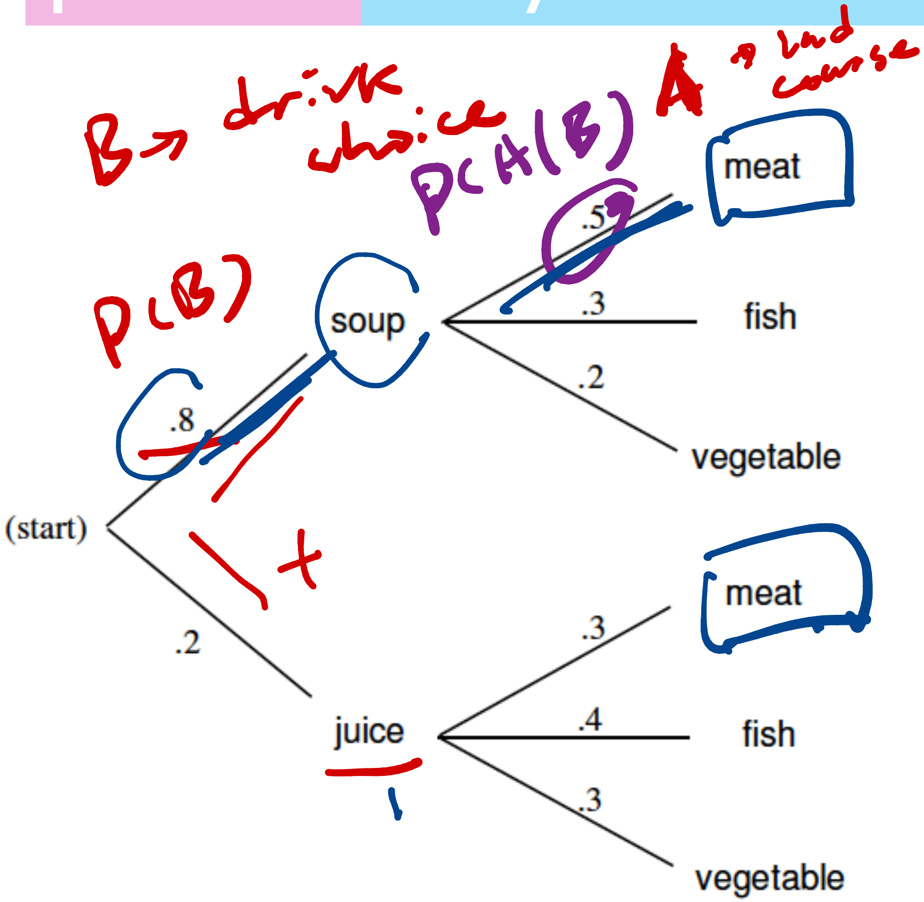
- Joint event

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

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

B → data

Multiplication using conditional probability



$$\underline{P(\text{soup} \cap \text{meat})}$$

$$= 0.8 \times 0.5$$

$$= \underbrace{P(\text{soup})}_{\text{prior}} \cdot \underbrace{P(\text{meat} \mid \text{soup})}_{\text{likelyhood}}$$

$$P(\text{meat}) = ?$$

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.

$B \cap A = A \cap B$

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

$$\Rightarrow$$

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

$P(A) \neq 0$

$P(B) \neq 0$

The famous **Bayes** rule

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

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

$p(A \mid (B_1, B_2 \ldots))$ D

$D_1 \to D_2$

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**?

B : a bad car from the dealer

A : it came from Fac B

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{1002}}{\frac{12}{1002}} = \frac{2}{12} = \frac{1}{6}$$

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?

$$P(B|L) = \frac{P(L|B)P(B)}{P(L)} = \frac{1 \times \frac{2}{1002}}{\frac{12}{1002}} = \frac{2}{12} = \frac{1}{6}$$

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.

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.

$$\boxed{52 \times 12}$$

Permutations (order matters)

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

Permutations (order matters)

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

10×9×8 = P(10,3) = 720

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

Combinations (order not important)

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

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

Combinations (order not important)

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

$$\boxed{\mathbf{C(N,2) = N\times(N-1)/2}}$$

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

Partition

- How many ways are there to rearrange ILLINOIS?

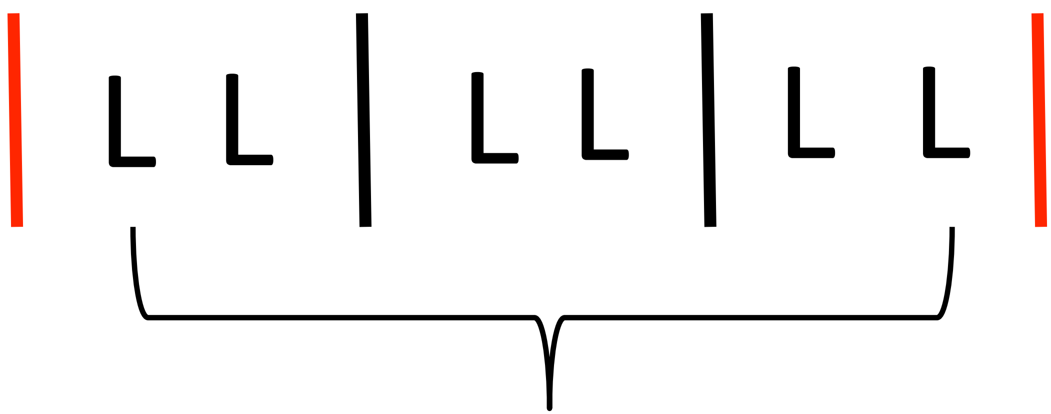
$$\frac{8!}{\underset{\substack{\uparrow \\ I}}{3!}\,\underset{\substack{\uparrow \\ L}}{2!}\,1!\,1!\,1!}$$

- General form

$$\frac{n!}{n_1!\,n_2!\ldots n_r!}$$

Allocation

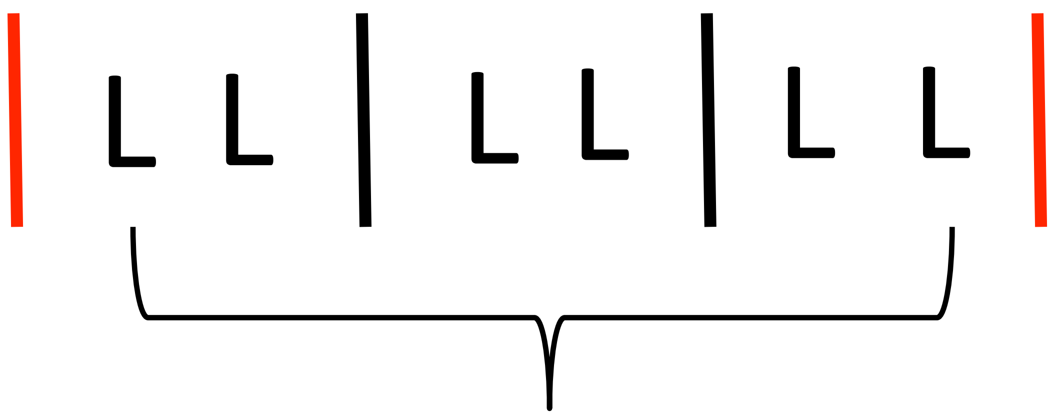
- Putting 6 identical letters into 3 mailboxs (empty allowed)



Choose 2 from the 8 positions

Allocation

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



Choose 2 from the 8 positions: $\boxed{\mathbf{C(8,2) = 28}}$

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)

$$C(3019, 19) = \frac{3019!}{19!3000!}$$

See you next time

See You!