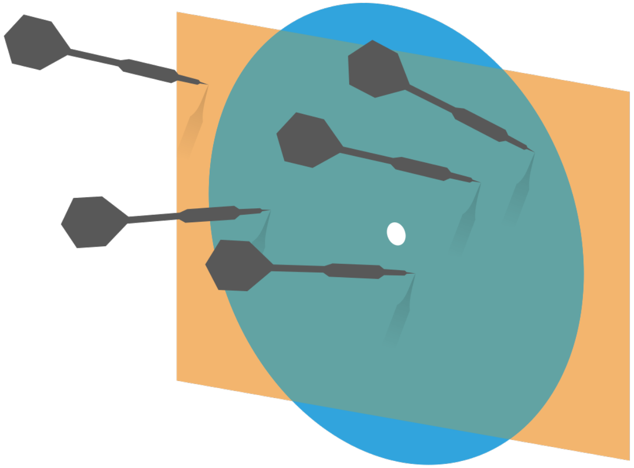
Probability and Statistics for Computer Science



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, 9.10.2020

Last time

Probability

More probability calculation

- Conditional probability

* Bayes rule

* Independence

Objectives

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

- Conditional Probability

* Product rule of joint prob.

* Bayes rule

* Independence

Counting: how many ways?

to put 7 hats (hats are indistinguishable) on 7 of 10 people randomly?

$$\binom{10}{7}$$

Warm up: which is larger?

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

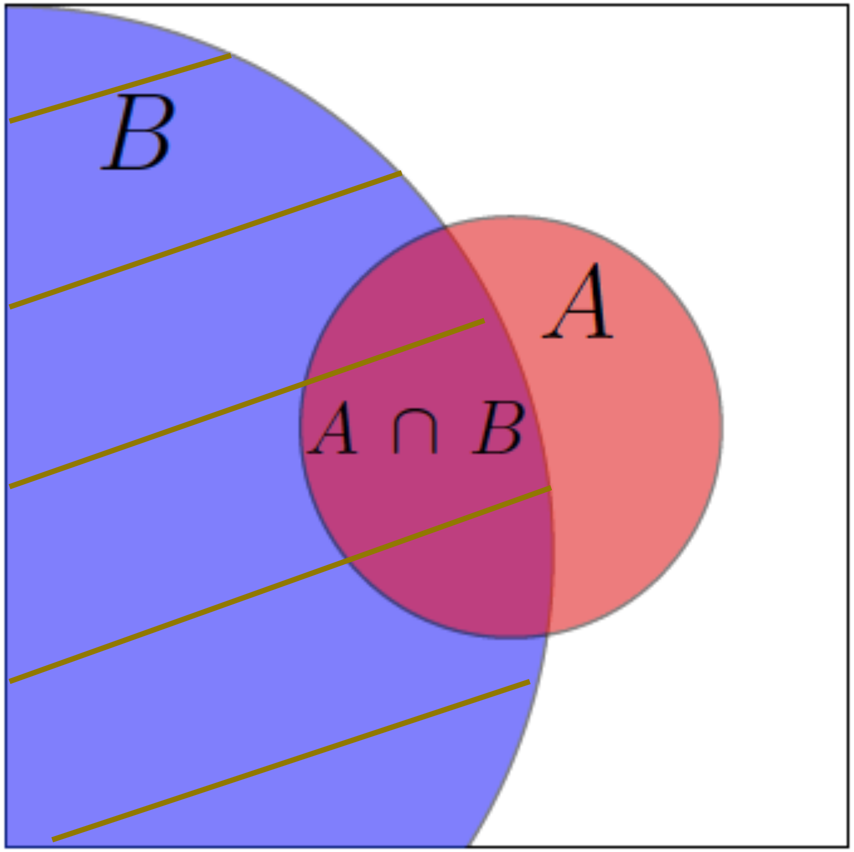
A) $P(A \cap B)$

B) $P(A|B)$

C) unsure

Conditional Probability

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



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

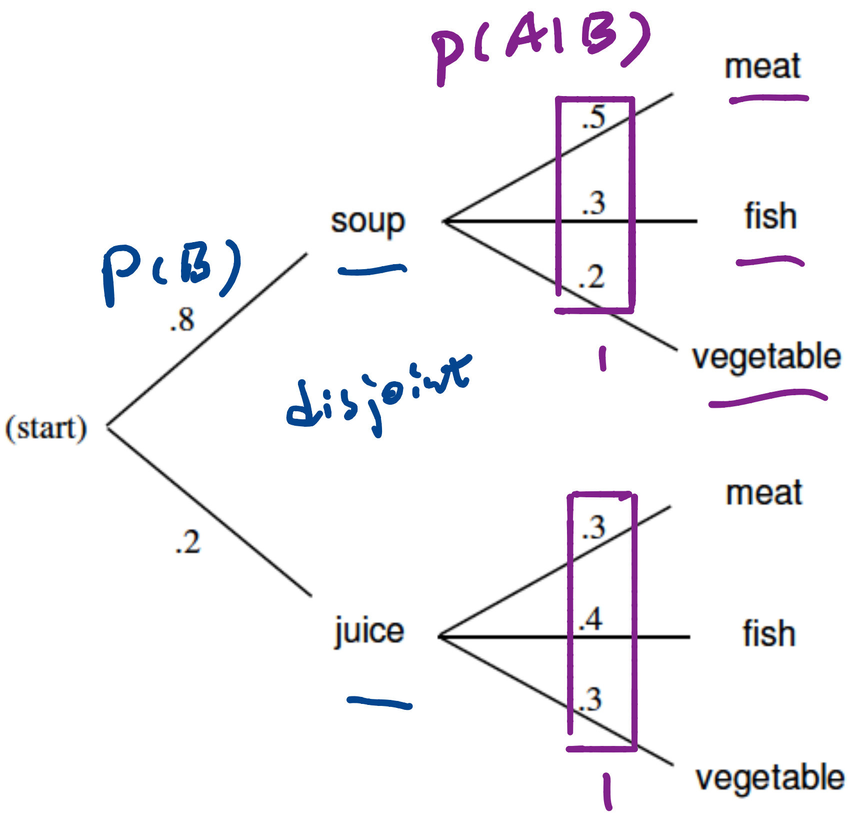
$$P(B) \neq 0$$

The line-crossed area is the new sample space for conditional P(A| B)

Joint Probability Calculation

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

$P(A \cap B)$
0.8×0.5



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

Bayes rule

- Given the definition of conditional probability and the symmetry of joint probability, we have:

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

And it leads to the famous Bayes rule:

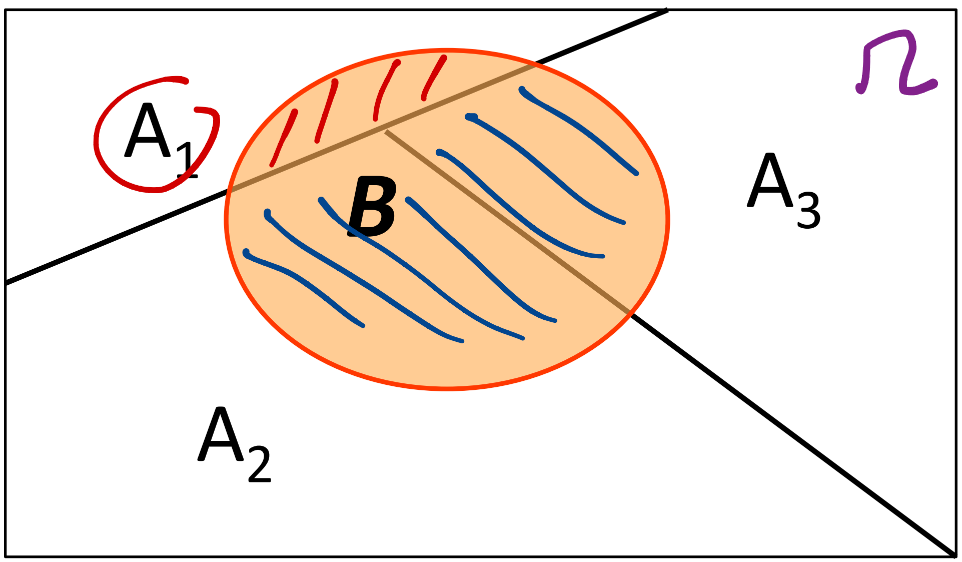
$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

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

Total probability

$$P(B) = P(B \cap A) + P(B \cap A^c)$$

$$= P(B|A)P(A) + P(B|A^c)P(A^c)$$



i.e.

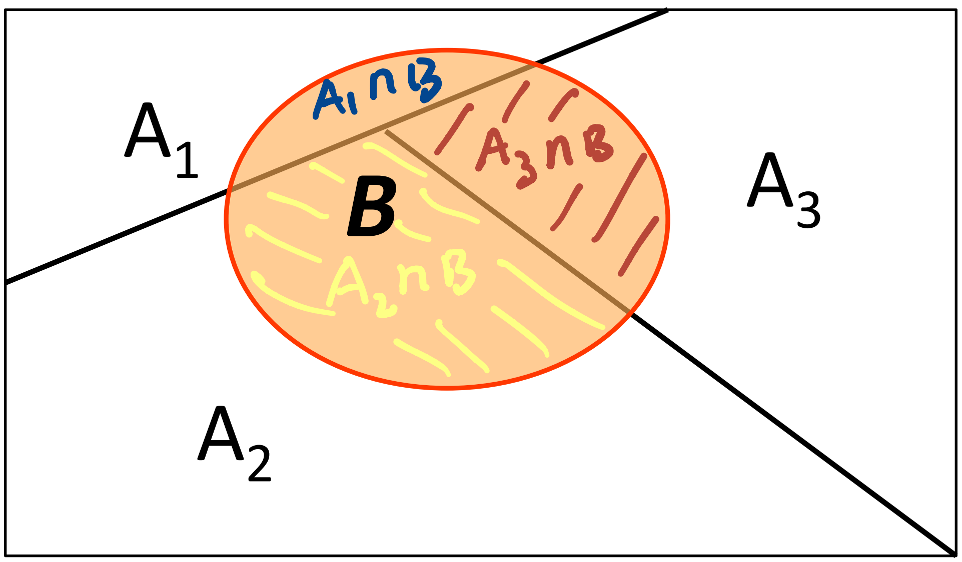
$A = A_1$

$A^c = A_2 \cup A_3$

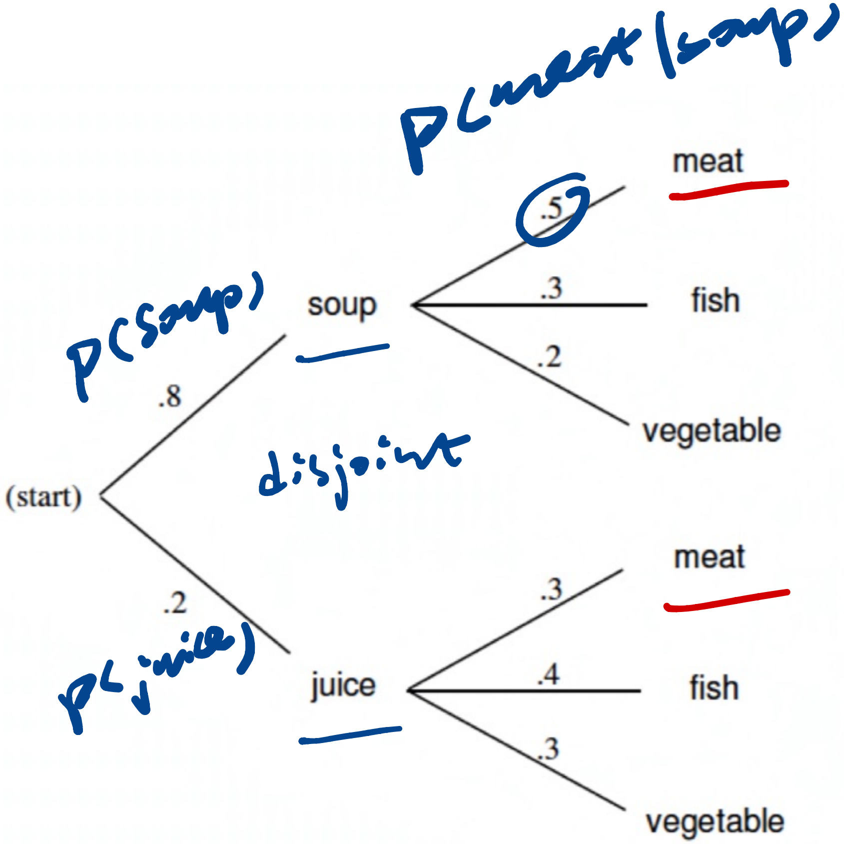
Total probability general form

$$P(B) = \sum_j P(B \cap A_j)$$

$$= \sum_j P(B \mid A_j) P(A_j)$$



Total probability:



$P(\text{meat}) = ?$ $P(\text{soup} \cap \text{meat}) + P(\text{juice} \cap \text{meat})$

$P(\text{meat} | \text{soup}) P(\text{soup})$

$+ P(\text{meat} | \text{juice}) P(\text{juice})$

$$P(\text{soup} | \text{meat}) = \frac{P(\text{meat} \cap \text{soup})}{P(\text{meat})}$$

Bayes rule using total prob.

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

$$\frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)\,P(A_j)}{\sum_j P(B|A_j)\,P(A_j)}$$

$A_j \cap A_i = \phi$ → disjoint

if $i \neq j$

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?

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

$$= \frac{0.95 \times 1/100000}{0.95 \times \frac{1}{100000} + 0.001 \times (1 - \frac{1}{100000})} = < 1\%$$

What about Covid test?

suppose freq. of Covid $= 1.21\%$

test accuracy $= 95\%$

false positive $= 0.001$

$P(T|D) = 0.95$

$P(D) = 1.21\%$

$P(D|T) = ?$

$P(T|D^c) = 0.001$

$P(D^c) = 1 - 1.21\%$

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

$$\doteq 92\%$$

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."

A H*

B HH TT

$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\{HH\}}{\{HH, TT\}} = \frac{1}{2} = P(A) = \frac{1}{2}$

- These two events are independent?

Independence

- Alternative definition

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

LHS by definition

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

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

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

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?

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

$p(E_1 \cap E_2) = \frac{3}{52}$

$p(E_1) = \frac{3 \times 4}{52} = \frac{3}{13}$

$p(E_2) = \frac{1}{4}$

$\frac{3}{13} \times \frac{1}{4} = \frac{3}{52}$

Independence vs Disjoint

- Q. Two disjoint events that have probability> 0 are certainly dependent to each other. ϕ

  A. True

  B. False

$P(A \cap B) = 0$

$P(A) > 0 \quad P(B) > 0$

$0 = P(A \cap B) \neq P(A)P(B)$

Independence of empty event

- Q. Any event is independent of empty event B.

  A. True

  B. False

$B = \phi$

$P(\phi) = 0$

$$P(\overset{\phi \cap A}{B \cap A}) = P(\overset{0}{\cancel{B}})P(A) = 0$$

$P(\phi) =$

Pairwise independence is not mutual independence in larger context

A_1	A_2
A_4	A_3

$A_1 \cup A_2 \cup A_3 \cup A_4 = \Omega$

$P(A_1) = P(A_2) = P(A_3) = P(A_4) = 1/4$

$A = A_1 \cup A_2; P(A) = \frac{1}{2}$

$B = A_1 \cup A_3; P(B) = \frac{1}{2}$

$C = A_1 \cup A_4; P(C) = \frac{1}{2}$

$P(A \cap B) = P(A)P(B)$ ✓

$P(B \cap C) = P(B)P(C)$ ✓

...

$P(A \cap B \cap C) = P(A_1) = \frac{1}{4}$

$P(A)P(B)P(C) = \left(\frac{1}{2}\right)^3$

$*P(ABC)$ *is the shorthand for* $P(A \cap B \cap C)$

if MI $\Rightarrow$ PI

Mutual independence

- Mutual independence of a collection of events $A_1, A_2, A_3 \ldots A_n$ is :

$$P(A_i \mid A_j \cap A_k \cap \cdots A_p) = P(A_i)$$

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

$$j, k, \ldots p \neq i$$

- It's very strong independence!

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

Probability using the property of Independence: Airline overbooking (1)

- An airline has a flight with 6 seats. They always sell 7 tickets for this flight. If ticket holders show up independently with probability **p**, what is the probability that the flight is overbooked ?

$\binom{7}{7} = 1$

p^7

$P(A_1 \cdots A_7)$

$= P(A_1) P(A_2) \cdots P(A_7)$

$= p \cdot p \cdots p$

Probability using the property of Independence: Airline overbooking (1)

- An airline has a flight with 6 seats. They always sell 7 tickets for this flight. If ticket holders show up independently with probability **p**, what is the probability that the flight is overbooked ?

P(7 passengers showed up)

Probability using the property of Independence: Airline overbooking (2)

- An airline has a flight with 6 seats. They always sell 8 tickets for this flight. If ticket holders show up independently with probability **p**, what is the probability that exactly 6 people showed up?

Event $= \bigcup A_j$

$(1-p)(1-p)$

$p \cdot p \cdot p \cdot p \cdot p \cdot p$

P(6 people showed up) =

$= \binom{8}{6} \cdot p^6 (1-p)^2$

Probability using the property of Independence: Airline overbooking (3)

- An airline has a flight with 6 seats. They always sell 8 tickets for this flight. If ticket holders show up independently with probability **p**, what is the probability that the flight is overbooked ?

P(overbooked) =

Probability using the property of Independence: Airline overbooking (4)

- An airline has a flight with **s** seats. They always sell **t** (**t**>**s**) tickets for this flight. If ticket holders show up independently with probability **p**, what is the probability that exactly **u** people showed up?

P(exactly **u** people showed up)

$$\binom{t}{u} \cdot p^{u} (1-p)^{t-u}$$

Probability using the property of Independence: Airline overbooking (5)

- An airline has a flight with **s** seats. They always sell **t** (**t**>**s**) tickets for this flight. If ticket holders show up independently with probability **p**, what is the probability that the flight is overbooked ?

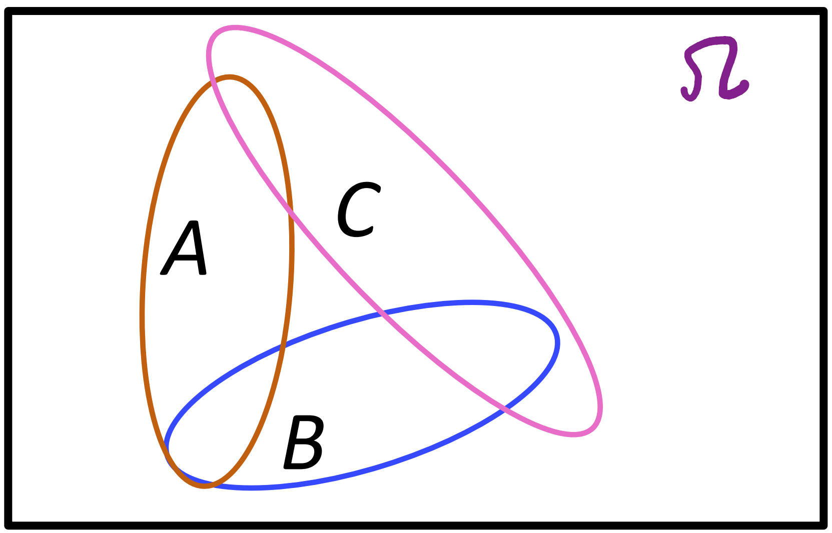
$t > s$

$s \quad s+1 \quad s+2 \quad \ldots \quad t$

P(overbooked) $\displaystyle\sum_{u=s+1}^{t} \binom{t}{u} p^{u} (1-p)^{t-u}$

Condition may affect Independence

- Assume event ***A*** and ***B*** are pairwise independent



Given ***C***, ***A*** and ***B*** are not independent any more because they become disjoint

Conditional Independence

- Event ***A*** and ***B*** are conditional independent given event ***C*** if the following is true.

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

See an example in Degroot et al. Example 2.2.10

Assignments

- HW3
- Finish Chapter 3 of the textbook
- Next time: Random variable

Additional References

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

Another counting problem

- There are several (>10) freshmen, sophomores, juniors and seniors in a dormitory. In how many ways can a team of 10 students be chosen to represent the dorm? There are no distinction to make between each individual student other than their year in school.

See you next time

See You!