CSE 312

Foundations of Computing II

Lecture 5: Conditional Probability and Bayes Theorem

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

Anna R. Karlin

Slide Credit: Based on Stefano Tessaro's slides for 312 19au
incorporating ideas from Alex Tsun, Rachel Lin, Hunter Schafer & myself ☺

Thank you for your feedback!!! ❤

- Several people mentioned that I was going too fast.
  - Slow me down! Ask questions!!!
  - Watch Summer 2020 videos before class (at half speed)
  - Do the reading before class.

Use edstem thread!

- Some people said they wanted more practice
  - Problems in textbook
  - Do the section problems!
  - MIT "Mathematics for Computer Science" 6.042J (sections on counting & probability)
  - Get the book "A First Course in Probability" by Sheldon Ross
- More office hours?

Review Probability

Definition. A **sample space** Ω is the set of all possible outcomes of an experiment.

Examples:

- Single coin flip: $\Omega = \{H, T\}$
- Two coin flips: $\Omega = \{HH, HT, TH, TT\}$
- Roll of a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Definition. An **event** $E \subseteq \Omega$ is a subset of possible outcomes.

Examples:

- Getting at least one head in two coin flips: $E = \{HH, HT, TH\}$
- Rolling an even number on a die : $E = \{2, 4, 6\}$

$(\Omega, Pr(\cdot))$

$Pr(\omega) \in [0,1] \quad \forall \omega \in \Omega$

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

uniform prob space

$$Pr(\omega) = \frac{1}{|\Omega|}$$

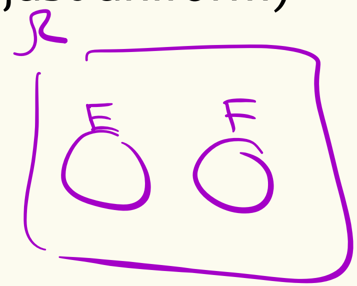
$$Pr(E) = \sum_{\omega \in E} Pr(\omega)$$

$$Pr(E) = \frac{|E|}{|\Omega|}$$

Review Axioms of Probability

Let Ω denote the sample space and $E, F \subseteq \Omega$ be events. Note this is more general to **any** probability space (not just uniform)

Axiom 1 (Non-negativity): $P(E) \geq 0$
Axiom 2 (Normalization): $P(\Omega) = 1$
Axiom 3 (Countable Additivity): If E and F are mutually exclusive,
then $P(E \cup F) = P(E) + P(F)$



Corollary 1 (Complementation): $P(E^c) = 1 - P(E)$
Corollary 2 (Monotonicity): If $E \subseteq F$, $P(E) \leq P(F)$
Corollary 3 (Inclusion-Exclusion): $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Agenda

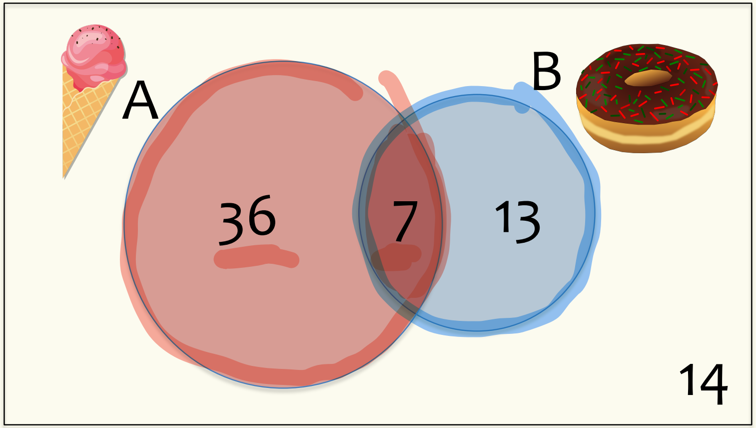
- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- Bayes Theorem + Law of Total Probability
- More Examples

$36+7+13+14=70$

$|\Omega|=70$

$Pr(\omega)=\frac{1}{70}$

Conditional Probability (Idea)



$Pr(A)=\frac{36+7}{70}$

What's the probability that $\underbrace{\text{someone likes ice cream}}_{A}$ **given** they $\underbrace{\text{like donuts?}}_{B}$

$$Pr(A \underset{\substack{\uparrow \\ \text{given that}}}{|} B) = \frac{7}{7+13} = \frac{|A\cap B|}{|B|} = \frac{|A\cap B|/|\Omega|}{|B|/|\Omega|} = \frac{Pr(A\cap B)}{Pr(B)}$$

$Pr(E|F)$

Conditional Probability

$Pr(A) = Pr(A|\Omega) = \frac{Pr(A \cap \Omega)}{Pr(\Omega)}$

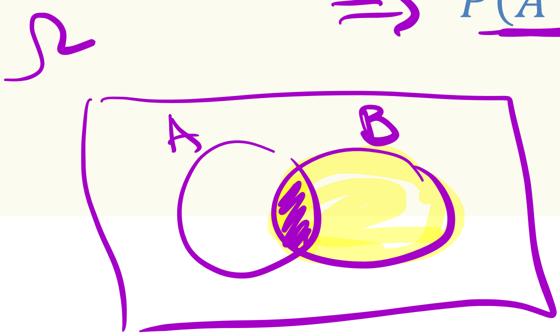
Definition. The **conditional probability** of event A **given** an event B happened (assuming $P(B) \neq 0$) is

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

Prob of A given B (Prob of A conditioned on B)

An equivalent and useful formula is

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



Reversing Conditional Probability

Question: Does $P(A|B) = P(B|A)$?

No! The following is purely for intuition and makes no sense in terms of probability

- Let A be the event you are wet
- Let B be the event you are swimming

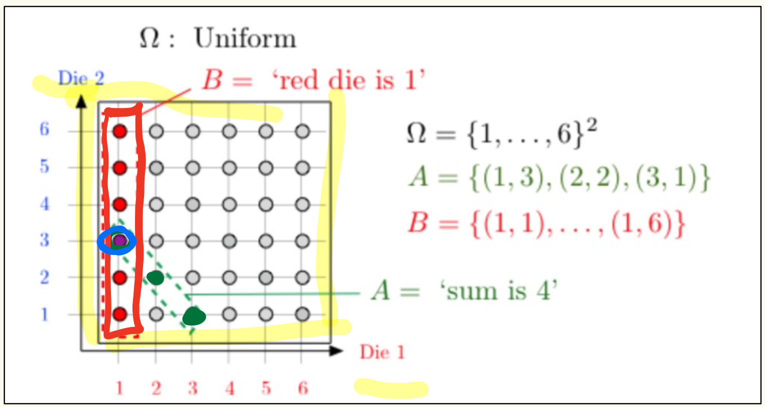
$$P(A|B) = 1$$
$$P(B|A) \neq 1$$

Example with Conditional Probability

https://pollev.com/ annakarlin185

Toss a red die and a blue die (both 6 sided and all outcomes equally likely). What is $P(B)$? What is $P(B|A)$?

	$P(B)$	$P(B\|A)$
a)	1/6	1/6
b)	1/6	1/3
c)	1/6	3/36
d)	1/9	1/3



$\Pr(\omega) = \frac{1}{36}$

A: sum is 4

B: red die shows 1

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

$\Pr(B) = \frac{|B|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$

$\Pr(B|A) = \frac{1/36}{3/36} = \frac{1}{3}$

$$\Pr(E|F) = \frac{\Pr(E\cap F)}{\Pr(F)} = \frac{\frac{|E\cap F|}{|\Omega|}}{\frac{|F|}{|\Omega|}} = \frac{|E\cap F|}{|F|}$$

Gambler's fallacy

Ω = set of all seqs of length 51 of H's & T's,

$|\Omega| = 2^{51}$ uniform prob space.

Assume we toss **51** fair coins.
Assume we have seen **50** coins, and they are all "tails".
What are the odds the **51**st coin is "heads"?

$\mathcal{A}$ = first 50 coins are "tails" $= \{\overbrace{TT\ldots TT}^{50}, \overbrace{TT\ldots TH}^{50}\}$

B = first 50 coins are "tails", 51st coin is "heads" $\{\overbrace{TT\ldots TH}^{50}\}$

$$\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\Pr(B\cap A)}{\Pr(A)} = \boxed{\frac{|B\cap A|}{|A|}} = \frac{1}{2}$$

$$\frac{\Pr(B\cap A)}{\Pr(A)} = \boxed{\frac{|B\cap A| / \cancel{|\Omega|}}{|A| / \cancel{|\Omega|}}}$$

Gambler's fallacy

Assume we toss **51** fair coins.
Assume we have seen **50** coins, and they are all "tails".
What are the odds the **51**$^{\text{st}}$ coin is "heads"?

$\mathcal{A}$ = first 50 coins are "tails"

B = first 50 coins are "tails", 51$^{\text{st}}$ coin is "heads"

$$\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A} \cap \mathcal{B})}{\mathbb{P}(\mathcal{A})} = \frac{1/2^{51}}{2/2^{51}} = \frac{1}{2}$$

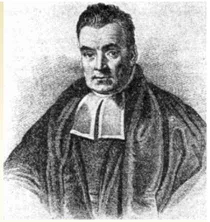
51$^{\text{st}}$ coin is independent of outcomes of first 50 tosses!

Gambler's fallacy = Feels like it's time for " heads"!?

Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- Bayes Theorem + Law of Total Probability
- More Examples

Bayes Theorem



A formula to let us "reverse" the conditional.

Theorem. (Bayes Rule) For events A and B, where $P(A), P(B) > 0$,

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

$P(A)$ is called the **prior** (our belief without knowing anything)

$P(A|B)$ is called the **posterior** (our belief after learning B)

Bayes Theorem Proof

defn of cond prob.

$$P(\overset{E}{A}|\overset{F}{B}) = \frac{P(\overset{E}{A} \cap \overset{F}{B})}{\underset{F}{P(B)}}$$

$$P(ice|do) = \frac{P(ic \cap D)}{P(D)}$$

$$P(do|ice) = \frac{P(D \cap ic)}{P(ic)}$$

By definition of conditional probability

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

$$\|$$

Swapping A, B gives

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

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

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

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

Dividing both sides by $P(B)$ gives

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

Our First Machine Learning Task: Spam Filtering

S: email is spam

F: subject line contains "Free"

Subject: "FREE $$$ CLICK HERE"

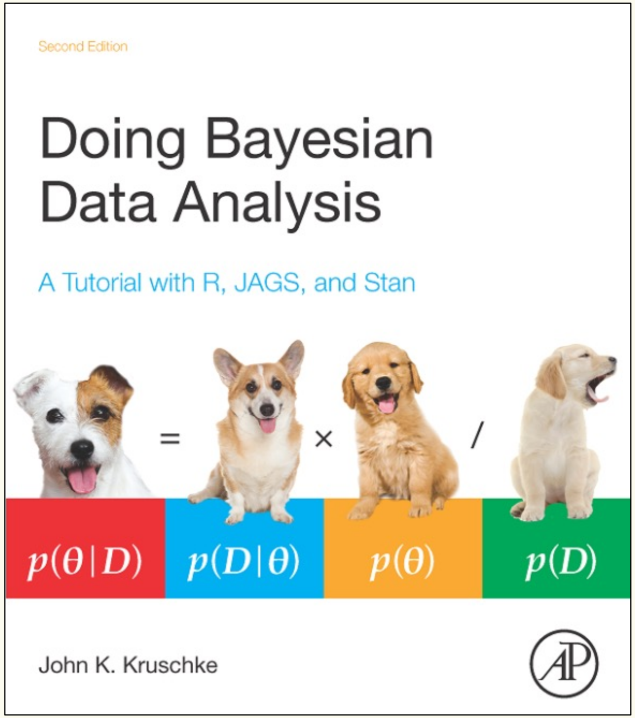
What is the probability this email is spam, given the subject contains "FREE"?
Some useful stats:

- 10% of ham (i.e., not spam) emails contain the word "FREE" in the subject. $\Pr(F \mid \bar{S}) = 0.1$
- 70% of spam emails contain the word "FREE" in the subject. $\Pr(F \mid S) = 0.7$
- 80% of emails you receive are spam. $\Pr(S) = 0.8$

$$\Pr(S \mid F) = \frac{\Pr(F \mid S)\,P(S)}{\Pr(F)} = \frac{0.7 \cdot 0.8}{\Pr(F)}$$

?

Brain Break


Second Edition
Doing Bayesian Data Analysis
A Tutorial with R, JAGS, and Stan
=
×
/
$p(\theta|D)$
$p(D|\theta)$
$p(\theta)$
$p(D)$
John K. Kruschke

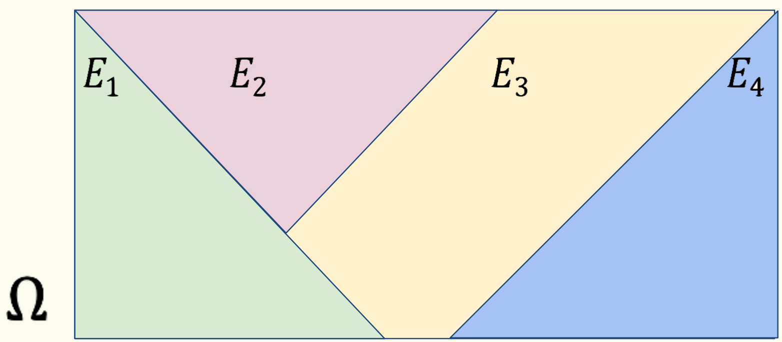
Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- Bayes Theorem + Law of Total Probability
- More Examples

Partitions (Idea)

These events **partition** the sample space

1. They "cover" the whole space
2. They don't overlap



Partition

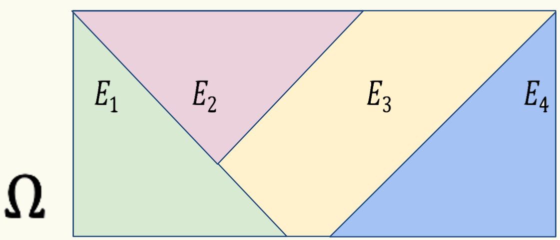
Definition. Non-empty events $E_1, E_2, \ldots, E_n$ **partition** the sample space Ω if

(Exhaustive)

$$E_1 \cup E_2 \cup \cdots \cup E_n = \bigcup_{i=1}^{n} E_i = \Omega$$

(Pairwise Mutually Exclusive)

$$\forall_i \forall_{i \neq j} E_i \cap E_j = \emptyset$$

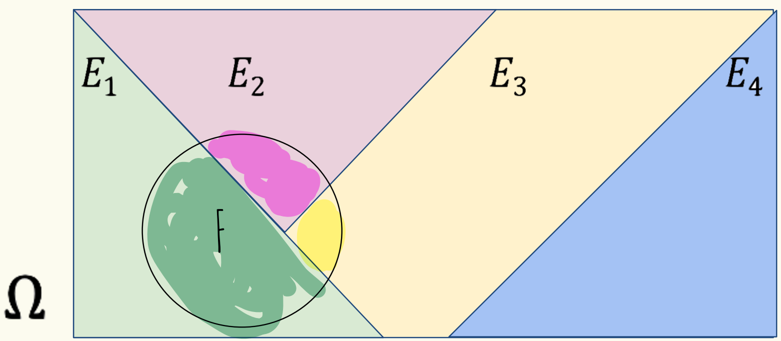


Law of Total Probability (Idea)

If we know $E_1, E_2, \dots, E_n$ partition Ω, what can we say about $P(F)$

$$\Pr(F) = \Pr(F \cap E_1) + P(F \cap E_2) + P(F \cap E_3) + \cancel{P(F \cap E_4)}$$

0



Law of Total Probability (LTP)

Definition. If events $E_1, E_2, \ldots, E_n$ partition the sample space Ω, then for any event F

$$P(F) = P(F \cap E_1) + \ \ldots + P(F \cap E_n) = \boxed{\sum_{i=1}^{n} P(F \cap E_i)}$$

Using the definition of conditional probability $\boxed{P(F \cap E) = P(F|E)P(E)}$
We can get the alternate form of this that show

$$P(F) = \boxed{\sum_{i=1}^{n} P(F|E_i)P(E_i)}$$

Another Contrived Example

Alice has two pockets:

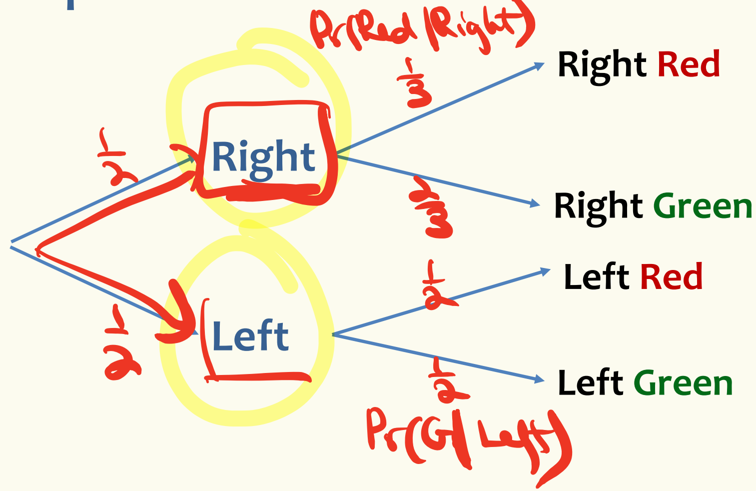
- **Left pocket:** Two red balls, two green balls
- **Right pocket:** One red ball, two green balls.

Alice picks a random ball from a random pocket.
[Both pockets equally likely, each ball equally likely.]

What is $\mathbb{P}(\mathbf{R})$?

red ball.

Sequential Process – Non-Uniform Case



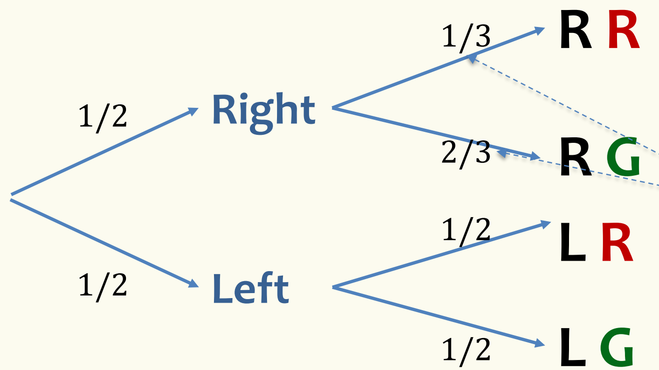
- **Left pocket:** Two red, two green
- **Right pocket:** One red, two green.

- Alice picks a random ball from a random pocket

$$\Pr(\text{Red}) = \underset{\frac{1}{3}}{\Pr(\text{Red} \mid \text{Right})}\,\underset{\frac{1}{2}}{\Pr(\text{Right})} + \underset{\frac{1}{2}}{\Pr(\text{Red} \mid \text{Left})}\,\underset{\frac{1}{2}}{\Pr(\text{Left})}$$

$$\Pr(F) = \sum_{i=1}^{n} \Pr(F \mid E_i)\,P(E_i) \quad \text{LTP.}$$

Sequential Process – Non-Uniform Case



- **Left pocket:** Two red, two green
- **Right pocket:** One red, two green.

1/3 = $\mathcal{P}(R \mid \mathrm{R})$ and 2/3 = $\mathcal{P}(G \mid \mathrm{R})$

$$\mathbb{P}(\mathbf{R}) = \mathbb{P}(\mathbf{R} \cap \textbf{Left}) + \mathbb{P}(\mathbf{R} \cap \textbf{Right})$$

(Law of total probability)

$$= \mathbb{P}(\textbf{Left}) \times \mathbb{P}(\mathbf{R} | \textbf{Left}) + \mathbb{P}(\textbf{Right}) \times \mathbb{P}(\mathbf{R} | \textbf{Right})$$

$$= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{3} = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$$

Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- Bayes Theorem + Law of Total Probability
- More Examples

Our First Machine Learning Task: Spam Filtering

S : email is spam

F : subject line contains "Free"

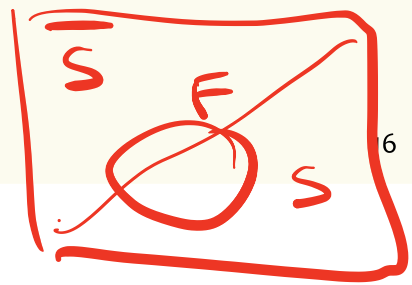
Subject: "FREE $$$ CLICK HERE"

What is the probability this email is spam, given the subject contains "FREE"?
Some useful stats:

- 10% of ham (i.e., not spam) emails contain the word "FREE" in the subject. $\Pr(F \mid \bar{S}) = 0.1$
- 70% of spam emails contain the word "FREE" in the subject. $\Pr(F \mid S) = 0.7$
- 80% of emails you receive are spam. $\Pr(S) = 0.8$

$$\Pr(S \mid F) = \frac{\Pr(F \mid S)\,P(S)}{\Pr(F)} = \frac{0.7 \cdot 0.8}{\Pr(F)}$$

?



$$\Pr(F) = \underset{0.7}{\Pr(F \mid S)}\,\underset{0.8}{P(S)} + \underset{0.1}{P(F \mid \bar{S})}\,\underset{0.2}{\Pr(\bar{S})}$$

Bayes Theorem with Law of Total Probability

Bayes Theorem with LTP: Let $E_1, E_2, \dots, E_n$ be a partition of the sample space, and F and event. Then,

$$P(E_1|F) = \frac{P(F|E_1)P(E_1)}{P(F)} = \frac{P(F|E_1)P(E_1)}{\sum_{i=1}^{n} P(F|E_i)P(E_i)}$$

Simple Partition: In particular, if E is an event with non-zero probability, then

$$P(E|F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^C)P(E^C)}$$