Unit 9
Probability

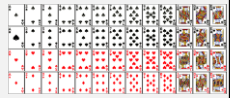
Computer Science 1002

Introduction to Logic for Computer Scientists

1

Puzzle: playing poker

- There are 52 cards in a standard deck; 4 suits of 13 ranks each.
- In poker, some 5-card combinations ("hands") are special:
  - For example, a "three of a kind" consists of three cards with the same rank, together with two arbitrary cards.
- What are the chances to get
  - a three of a kind hand?
  - A two pairs hand (5 cards with 2 same-rank pairs)?
  - Other hands?...



2

Finite probability

- **Experiments**: producing an **outcome** out of possible choices
  - Tossing a coin: outcome can be "heads"
  - Getting a lottery ticket: outcome can be "win"
- **Sample space S**: set of all possible outcomes.
  - {heads, tails} for a coin toss
  - {1,2,3,4,5,6} x {1,2,3,4,5,6} for rolling two dice
- **Event A** $\subseteq S$: subset of outcomes
  - Both dice came up even.

3

Finite probability

- **Probability** of an event if all outcomes are **equally likely**:
  - Pr(A)= |A|/|S| (fraction of the outcomes that are in the event A).
  - Probability of both dice coming up even:
    - A={ (2,2),(2,4),(4,2),(2,6),(6,2),(4,4), (4,6), (6,4) , (6,6)}.
    - |A| =9, |S|=36
    - P(A)=9/36=1/4
- Probability is always a number between 0 (event cannot happen, A=∅) and 1 (event always happens, A=S).
- If $\Pr(A) = p$, then $\Pr(\bar{A}) = 1 - p$ (that is, probability that A does not happen is 1 – probability of A.)

Let's use combinatorics we just studied to calculate probabilities!

4

Puzzle: playing poker

- There are 52 cards in a standard deck; 4 suits of 13 ranks each.
- In poker, some 5-card combinations ("hands") are special:
  - For example, a "three of a kind" consists of three cards with the same rank, together with two cards of other different ranks.
- How *many ways are there to choose* (ignoring the order)
  - A royal flush?
  - a three of a kind hand?
  - a two pairs hand?
  - other hands?...

5

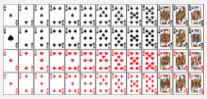
Puzzle: playing poker

- There are 52 cards in a standard deck; 4 suits of 13 ranks each.
- How many ways are there to choose a royal flush
  - all same suit, ranks A, K, Q, J, 10.
  - Pick a suit: C(4,1) = 4
- How many ways to choose three of a kind?
  - pick the rank: 13=C(13,1)
  - Pick 3 out of 4 kinds of this rank: 4=C(4,3)
  - Pick two other ranks: C(12,2)= 66
  - Pick a suit of each of the other ranks: C(4,1)*C(4,1)=16
  - Total: 13*4*66*16=54912

6

Puzzle: playing poker

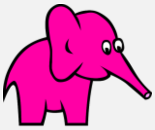
- What is the probability of getting a three of a kind hand?
  - Experiment: drawing a hand (5 cards).
  - Sample space S: set of all possible hands.
    - Size of S is C(52, 5) = $\binom{52}{5} = 2{,}598{,}962$
  - Event A: getting a three of a kind hand
    - Size of the event A: 54,912
  - Probability of A (all hands are equally likely):
    - Pr(A) = $\frac{|A|}{|S|} = 0.0211..$



7

Probabilities and pink elephants

- What is the probability that walking down George street you'd see a pink elephant?
  - Your friend says: "It is ½! You will either see the pink elephant, or not!"
    - Do you agree?

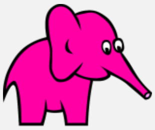


8

9

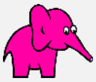
Probabilities and pink elephants

- What is the probability that walking down George street you'd see a pink elephant?
  - Your friend says: "It is ½! You will either see the pink elephant, or not!"
    - Do you agree?



10

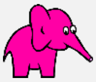
Probabilities and distributions



- What if outcomes are not equally likely?
  - Biased coins, pink elephants, etc.
- A function $Pr: S \to \mathbb{R}$ is a **probability distribution** on (a finite set) S if Pr satisfies the following:
  1) For any outcome $s \in S$, $0 \le \Pr(s) \le 1$
  2) $\Sigma_{\{s \in S\}} \Pr(s) = 1$
- **Uniform distribution**: for all $s \in S, \Pr(s) = 1/|S|$
  - all outcomes are equally likely
  - Fair coin: $\Pr(heads) = \Pr(tails) = \frac{1}{2}$

11

Biased coin



- Lots of distributions other than uniform!
- Biased coin: say heads twice as likely as tails.
  - $\Pr(heads) + \Pr(tails) = 1.$
  - $\Pr(heads) = 2 * \Pr(tails)$
  - So $\Pr(heads) = \frac{2}{3}, \Pr(tails) = \frac{1}{3}$

12

Probabilities of events

- Probability of an event A is a sum of probabilities of the outcomes in A
  $$\Pr(A) = \Sigma_{\{a \in A\}} \Pr(a)$$
  - Probability of A not occurring: $\Pr(\bar{A}) = 1 - \Pr(A)$
- Probability of the union of two events (either A or B happens) is
  $$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$
    - By principle of inclusion-exclusion
  - If A and B are disjoint, $\Pr(A \cap B) = 0$, so $\Pr(A \cup B) = \Pr(A) + \Pr(B)$
- In general, if events $A_1 \ldots A_n$ are **pairwise disjoint**
    - that is, $\forall i, j$ if $i \neq j$ then $A_i \cap A_j = \emptyset$
  - Then $\Pr(\cup_{i=1}^{n} A_i) = \Pr(A_1 \cup A_2 \cup \cdots \cup A_n) = \Sigma_{i=1}^{n} \Pr(A_i)$

13

Probabilities of events

- Suppose a die is biased so that 3 appears twice as often as any other number (others equally likely).
  - Probability of 3: 2/7. Probabilities of others: 1/7
- What is the probability that an odd number appears?
  - Event: A={1,3,5}
  - Pr(A) = 1/7+2/7+1/7=4/7.
- What is a probability that an odd number or a number divisible by 3 appears?
  - A={1,3,5}, B = {3,6}, $A \cap B = \{3\}$
  - Pr(A) = 4/7. Pr(B) = 3/7. $\Pr(A \cap B) = 2/7$
  - $\Pr(A \cup B) = \Pr(\{1,3,5,6\}) = \frac{4}{7} + \frac{3}{7} - \frac{2}{7} = \frac{1}{7} + \frac{2}{7} + \frac{1}{7} + \frac{1}{7} = \frac{5}{7}$

14

HAPPY BIRTHDAY

Birthday paradox

- How many people have to be in the room so that probability that two of them have the same birthday is at least ½?

15

16

HAPPY BIRTHDAY

Birthday paradox

- How many people have to be in the room so that probability that two of them have the same birthday is at least ½?

17

Birthday paradox

- How many people have to be in the room so that probability that two of them have the same birthday is at least ½?
  - Considering all birthdays independent: no twins!
    - If there are twins in the room, then the probability is 1.
    - So no twins is the harder case.
  - And considering all days equally likely
    - Otherwise again probability would be higher.
  - Even counting leap years: 366 days.
    - So each person has probability $1/366$ to have a birthday on any given day.

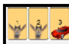
18

Birthday paradox

- Let's look at the first two people, first three, etc.
  - And calculate the probabilities of them all having different birthdays.
- Product rule: number of combinations of distinct birthdays of the first i people is P(366,i) = 366*365*…*(366-i+1)
  - Probability that the first i people all have different birthday is $\frac{P(366,i)}{366^i} = \frac{365}{366}\frac{364}{366}\dots\frac{(366-i+1)}{366}$
  - So with probability $1 - \frac{P(366,i)}{366^i}$ at least two out of first i people have birthday on the same day.
  - That works out to about i =23 people to reach ½.

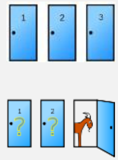
Only 23! That's why it is called a paradox.

19



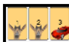
Puzzle: Monty Hall problem

- Let's make a deal!
  - A player picks a door.
  - Behind one door is a car.
  - Behind two others are goats.
- A player chooses a door.
  - A host opens another door
  - Shows a goat behind it.
  - And asks the player if she wants to change her choice.
- Should she switch?



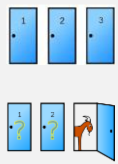
20

21



Puzzle: Monty Hall problem

- Let's make a deal!
  - A player has to pick a door.
  - Behind one door is a car.
  - Behind two others are goats.
- A player chooses a door.
  - A host opens another door
  - Shows a goat behind it.
  - And asks the player if she wants to change her choice.
- Should she switch?

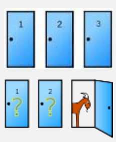


22

Puzzle: Monty Hall problem

- A player has to pick a door. Behind one door is a car. Behind other doors are goats.
- After the player chooses a door, the host opens another door, shows a goat behind it, and asks the player if she wants to change her choice.

- Should she switch?
  - Originally, probability of picking the car is 1/3
  - If she first picked a door with a car: (1/3 probability)
    - Then she would switch to a goat.
  - If she first picked a door with a goat (2/3 probability)
    - Then she would switch to a car.
  - Yes, she should switch to increase her probability of getting a car!



23

Conditional probabilities

- **Conditional probability** of an event A given event B, denoted $\Pr(A|B)$, is the probability of A if we know that B occurred.
  - Probability of rain given that it is cloudy is higher than just probability of rain.
    - Probability of rain is lower given that it is sunny.
  - In the Monty Hall puzzle
    - Probability of car behind door 2 is 1/3
    - Probability of a car behind door 2 conditional the host showing a goat behind door 3 is 2/3
- It is probability of both A and B, given that B happened for sure:

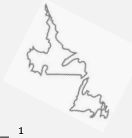
$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

  - Assume that $\Pr(B) > 0$: after all, B did happen.

24


Conditional probabilities

- Suppose that there are 9,000 black bears and 1000 polar bears in our province, NL. Moreover, all polar bears and 2000 of black bears are in Labrador.
  - Let A be an event that a bear is a polar bear.
  - Let B be an event that a bear is in Labrador (rather than Newfoundland).



- Probability that a random bear in NL is a polar bear:
  - $\Pr(A) = \frac{1000}{1000+9000} = \frac{1}{10}$
- Probability that a random bear in NL is in Labrador:
  - $\Pr(\mathrm{B}) = \frac{1000+2000}{1000+9000} = \frac{3}{10}$
- Probability of being both a polar bear and in Labrador is $\Pr(A \cap B) = \frac{1}{10}$
  - Same as probability of being a polar bear, since all polar bears are in Labrador
- Probability that a random bear is a polar bear conditional on it being in Labrador:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/10}{3/10} = \frac{1}{3} = \frac{1000}{1000 + 2000}$$

  - With this data $\Pr(A|\bar{B}) = 0$. A random bear in Newfoundland is a black bear.

25


Independent events

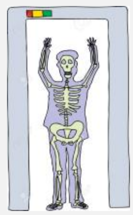
- If knowing B gives us no information about A and vice versa, then A and B are **independent** events:
  - Then Pr(A) = Pr(A|B) = $\frac{\Pr(A \cap B)}{\Pr(B)}$.
  - So A and B are independent iff $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$.
    - In general, events $A_1 \dots A_n$ can be **pairwise independent** (that is, any two A_i, A_j) are independent, or (stronger condition) **mutually independent**: $\forall\, \mathrm{T} \subseteq \{A_1, \dots, A_n\}\ \Pr_{A_i \in T}(A_i) = \prod_{A_i \in T} \Pr(A_i)$

      That is, for every subset of these events, probability of them occurring together is the product of individual probabilities.
  - Different coin tosses/dice rolls are usually considered independent.

26

27

Contrapositive vs. Converse



- "If a person is carrying a weapon, then the airport metal detector will ring".
  - Same as "If the airport metal detector does not ring, then the person is not carrying a weapon".
  - Not the same as: "If the airport metal detector rings, then the person is carrying a weapon."
- "If the person is sick, then the test is positive".
- "If he is a murderer, his fingerprints are on the knife".

28

Sensitivity and specificity

- Let A: "metal detector rings", B: "person carries a weapon",

Sensitivity:

percentage of correct positives

- Pr(A|B) = $\Pr(A \cap B)\ /\Pr(B)$
- Probability that if a person has a weapon, then detector rings
- Probability that if the person is sick, then the test is positive
- Test that always says "yes" has 100% sensitivity

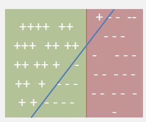
Specificity:

percentage of correct negatives

- $\Pr(\bar{A}|\bar{B})$ = $\Pr(\bar{A} \cap \bar{B})\ /\Pr(\bar{B})$
- Probability that if the detector rings, then the person has a weapon
- Probability that if the person is not sick, then the test is negative
- Test that always says "no" has 100% specificity.

29

Sensitivity and specificity



Sensitivity: percentage of correct (true) positives
- Pr(A|B) = $\Pr(A \cap B)\ /\Pr(B)$

Specificity: percentage of correct (true) negatives
- $\Pr(\bar{A}|\bar{B})$ = $\Pr(\bar{A} \cap \bar{B})\ /\Pr(\bar{B})$

- Let green area be what test detected as positive (red as negative)
  - Event A: in green area. Event B: "+"
- The blue line represents correct separation between positives (+) and negatives (-)
  - Positive can mean "sick", negative "healthy"
- The green part under blue line is false positives.
- The red part above blue line is false negatives.

- 24 + (positives),
- 30 - (negatives)
- 23 true positives
  - One false negative
- 22 true negatives
  - 8 false positives
- Sensitivity: 100*23/24=96%
- Specificity: 100*22/30=73%

30

Medical test problem

- Consider a medical test that checks for a disease. This test
  - Has false positive rate of 3% (healthy labeled as sick)
    - **Specificity** 97%
  - Has false negative rate of 1% (sick labeled as healthy).
    - **Sensitivity** 99%
- What is the probability that a person has the disease given that the test came positive?
  - Let A: person tested positive, B: person is sick. $\Pr(B|A)$?
  - Pr(A|B) = 0.99, $\Pr(\bar{A}|B) = 0.01\ldots$
- Not enough information!

31

Bayes theorem

- **Bayes theorem** allows us to get $\Pr(B|A)$ from $\Pr(A|B)$, if we know probabilities of A and B:

$$\Pr(B|A) = \frac{\Pr(A|B)\cdot\Pr(B)}{\Pr(A)} = \frac{\Pr(A|B)\cdot\Pr(B)}{\Pr(A|B)\Pr(B)+\Pr(A|\bar{B})\Pr(\bar{B})}$$

- Proof:
  - $\Pr(A|B) = \frac{\Pr(A\cap B)}{\Pr(B)}$. $\Pr(B|A) = \frac{\Pr(A\cap B)}{\Pr(A)}$.
  - $\Pr(A \cap B) = \Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$
  - So $\Pr(B|A) = \Pr(A|B)\Pr(B)/\Pr(A)$
- The formula $\Pr(A) = \Pr(A|B)\Pr(B) + \Pr(A|\bar{B})\Pr(\bar{B})$ comes from writing probability of A (e.g., a positive test) as sum of probabilities of A for B (sick people) and for $\bar{B}$ (healthy people).

32

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

- Consider a medical test that
  - Has false positive rate of 3% (healthy labeled as sick).
  - Has false negative rate of 1% (sick labeled as healthy).
  - Tests for a disease that occurs in 5 in 1000 people.
- What is the probability that a person is sick given that the test came positive?
  - Let A: person tested positive, B: person is sick. $\Pr(B|A)$?
  - Pr(A|B) = 0.99, $\Pr(\bar{A}|B) = 0.01$. $\Pr(\bar{A}|\bar{B})$ = 0.97 , $\Pr(A|\bar{B}) = 0.03$
  - $\Pr(B) = 0.005$.
  - $\Pr(A) = \Pr(A|B)\Pr(B) + \Pr(A|\bar{B})\Pr(\bar{B}) = 0.0348$
  - By Bayes theorem, Pr(B|A) = $\frac{\Pr(A|B)\Pr(B)}{\Pr(A)}$ = 0.1422
- So the probability that a person who tested positive has the disease is just 0.1422, that is, 14.22%.
  - By a similar argument, probability that a person who tested negative does not have a disease is whopping 0.99995 = 99.995%.

33

Hat check problem

- In Victorian England, n men came to an event, and checked in their hats at the door.
  - On the way out, in a hurry, they each picked up a random hat.
  - On average, how many men picked their own hat?

34

35

Expectations

- Often we are interested in what outcome we would see "on average".
  - How fast does this program run "on average"?
- Let possible outcomes of an experiment be numbers $a_1, \ldots, a_n$
  - E.g., time a program takes to sort n elements
- Its **expected value (mean)** is $\Sigma_{k=1}^{n} a_k \Pr(a_k)$
  - Often phrased in terms of a "random variable" X, where X is a *function* from outcomes to numbers.
  - Write E(X) to mean the expected value (mean, expectation) of X.

36

Random variables

- "Random variable" X is a *function* from outcomes to numbers.
  - In Computer Science applications, usually X counts something.
    - Number of heads out of n coin tosses.
    - Number of steps a program takes on an input
  - An *indicator* random variable X takes value 0 or 1 depending on whether an event occurred or not.
- Expectation of a random variable X is

$$E(X) = \sum_{i \in Outcomes} X(i) * \Pr(i)$$

37

Expectation example

Suppose we roll two fair dice. What is the expected sum of their values?

- X can take values from 2 to 12.
  - Pr(X=2) = Pr(X=12) = 1/36
  - Pr(X =3) = Pr(X =11) =2/36=1/18,
  - Pr(X =4) = Pr(X =10) =3/36=1/12,
  - Pr(X =5) = Pr(X =9) =4/36=1/9,
  - Pr(X =6) = Pr(X =8) =5/36,
  - Pr(X =7) =6/36=1/6

$$E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{1}{18} + 4 \cdot \frac{1}{12} + 5 \cdot \frac{1}{9} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{1}{6} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{1}{9} + 10 \cdot \frac{1}{12} + 11 \cdot \frac{1}{18} + 12 \cdot \frac{1}{36} = 7$$

38

Expected win in a lottery

- Rules of Lotto 6/49:
  - A player chooses 6 numbers, 1 to 49.
  - During a draw, 6 randomly generated numbers are revealed.
  - If all 6 numbers chosen by the player match 6 numbers in the draw, the player gets the jackpot of $5,000,000 or more.
    - There are also smaller prizes; let's ignore them for simplicity.
  - A ticket costs $3.
  - According to Atlantic Lotto Corporation, chances of winning the jackpot are roughly 1/14,000,000.

39

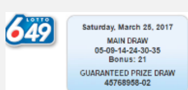
Expected win in a lottery

- What is the expected amount a player would win if the jackpot is 5,000,000?
  - Pr(win) = 1/14,000,000. Pr(loss)=1-Pr(win) = 13,999,999/14,000,000.
  - Let the random variable X encode the amount a player wins.
    - For all but one player, that amount is -3. So Pr(X=-3)=Pr(loss)
    - For the lucky one, the amount is the jackpot minus ticket price. Pr(X=4,999,997)=Pr(win)
  - Expected amount to win is E(X) = Pr(loss)*(-3) +Pr(win)*(5,000,000-3) = *-2.64*
    - If counting smaller prizes , just add their amount*odds to the sum, and adjust Pr(loss)
    - E(X)=Pr(loss)*(-3)+Pr(jackpot)*(4,999,997)+Pr(5/6+bonus)*374,997+Pr(5/6)*312,497…

40

Expected win in a lottery

- How large should be the jackpot so that the players expect at least to break even?
  - Let's call the jackpot amount J.
  - Expected amount to win is E(X) = Pr(loss)*(-3) +Pr(win)*(J-3).
    - To break even, want E(X)=0.
  - J= 3+(E(X)-Pr(loss)*(-3))/Pr(win) = *42,000,000*



41

42

Bernoulli trials and repeated experiments

- Suppose an experiment has two outcomes, 1 and 0 (success/failure), with Pr(1) = p.
  - Such experiment is called a **Bernoulli trial**.
- What happens if the experiment is repeated multiple times (independently)?

  - A sample space after carrying out n Bernoulli trials is a set of all possible n-tuples of elements in {0,1} (or {success, fail}).
  - Number of n-tuples with k 1s is $\binom{n}{k}$
  - Probability of getting 1 in any given trial is p, of getting 0 is (1-p).
  - Probability of getting exactly k 1s (successes) out of n trials is $\binom{n}{k}p^k(1-p)^{n-k}$
  - Probability of getting the first success on exactly the k^{th} trial is $p(1-p)^{k-1}$
- How many trials do we need, on average, to get a success?

43

Expected number until…

- Suppose that Alan insists on buying lottery tickets until he wins. If probability of winning is p, how many tickets in expectation ("on average") would he have to buy?
  - Let X be a random variable for how many tickets he has to buy.
  - The probability of winning on exactly i^{th} ticket is $p(1-p)^{i-1}$
  - $E(X) = \Sigma_{i\in\mathbb{N}}\, i * \Pr(X = i) = \frac{1}{p}$
    - So for Lotto 6/49 he'd have to buy 14,000,000 tickets (and spend $42,000,000 -- that's jackpot that would let him break even!)

44

Expected number until…

- Suppose we have Bernoulli trials with success probability p. What is the expected number of trials to see success?
  - Let X be a random variable for the number of steps till success.
  - $E(X) = \Sigma_{i\in\mathbb{N}}\, i * \Pr(X = i) = \frac{1}{p}$
- Same reasoning applies to other processes, where there is a fixed probability of something happening at each experiment or time step.
  - A system has a 1% probability of hanging in any given hour. How long, on average, will it stay up?
    - 100 hours: a little over 4 days.

45

46

Linearity of expectation

- Expectation is a very well-behaved operation:
  - $E(X_1 + X_2 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$
  - $E(aX + b) = a\,E(X) + b$
    - Where $X_1 \ldots X_n$ are random variables on some sample space S, and $a, b \in \mathbb{R}$
- Proof:
  - $E(X_1 + X_2) = \Sigma_{s\in S}\, p(s)(X_1(s) + X_2(s))$
    $= \Sigma_{s\in S}\, p(s)X_1(s) + \Sigma_{s\in S}\, p(s)X_2(s) = E(X_1) + E(X_2)$
  - Similar for $E(aX + b) = a\,E(X) + b$
    - Using the fact that $\Sigma_{s\in S}\, p(s) = 1$

47

Hat check problem

- In Victorian England, n men came to an event, and checked in their hats at the door.
  - On the way out, in a hurry, they each picked up a random hat.
  - On average, how many men picked their own hat?

48

Hat check problem

In Victorian England, n men came to an event, and checked in their hats at the door.
- On the way out, in a hurry, they each picked up a random hat.
- On average, how many men picked their own hat?

- For each man, introduce a random variable X_i, where $X_i = 1$ iff he picked his own hat
  - Such random variables are called **indicator variables**.
  - The quantity we want is $E(X_1 + \cdots + X_n)$
  - Now, for each X_i, $E(X_i) = 1 \cdot \Pr(X_i = 1) = \frac{1}{n}$
  - By linearity of expectation, $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = \frac{1}{n} + \cdots + \frac{1}{n} = n \cdot \frac{1}{n} = 1$
- So on average, just one man will go home with his own hat!

49

50

Hat check problem 2

- In Victorian England, n men came to an event, and checked in their hats at the door.
  - On the way out, in a hurry, they each picked up a random hat.
  - *What is the probability that at least k of them got their own hat?*

51

Deviation from the expectation

" three-quarters (73 percent) of U.S. drivers consider themselves better-than-average drivers" [AAA].

- Let R be a (nonnegative) random variable. What is the probability that $R \geq x$?
- **Markov's inequality**: $\Pr(R \geq x) \leq \frac{E(R)}{x}$
  - Since R is nonnegative, $E(R) \geq \ xPr(R \geq x) + 0\Pr(R < x)$
  - So $E(R) \geq xPr(R \geq x)$.

52

Using Markov's inequality

- **Markov's inequality**: $\Pr(R \geq x) \leq \frac{E(R)}{x}$
- Let's use Markov's inequality in the Hat Check Problem
  - Let $R = X_1 + \cdots + X_n$
  - each X_i is an indicator variable for i^th man getting his hat.
  - By Markov's inequality, $\Pr(R \geq k) \leq \frac{E(R)}{k} = \frac{1}{k}$.
- But that's a pretty weak bound!
  - Probability that everybody gets their hat is $\Pr(R \geq n) \leq \frac{1}{n}$ by Markov's inequality
  - However, we can calculate that it is only $\frac{1}{n!}$

53

Variance

Let's define the expectation of the difference from the mean.

Let X be a random variable on a sample space S. Define **variance** of X, denoted $V(X)$ to be

$$V(X) = E_{s \in S}\left(\left(X - E(X)\right)^2\right) = \sum_{s \in S}\left(X(s) - E(X)\right)^2 \cdot Pr(s)$$

A **standard deviation** of X, denoted $\sigma(X)$ where σ is a Greek letter sigma, is $\sigma(X) = \sqrt{V(X)}$

Useful equality: $V(X) = E(X^2) - \left(E(X)\right)^2$
- That is, the variance of X can be expressed as the difference between the expectation of X^2 and square of $E(X)$.
- To prove this, just open parentheses in our original definition of $V(X)$, and use the fact that $\sum_{s \in S} Pr(s) = 1$.

54

Variance

Variance of a random variable X on a sample space S is:

$$V(X) = E_{s\in S}\left(\left(X - E(X)\right)^2\right) = \sum_{s\in S}\left(X(s) - E(X)\right)^2 \cdot Pr(s) = E(X^2) - \left(E(X)\right)^2$$

A **standard deviation** of X is $\sigma(X) = \sqrt{V(X)}$

Example: consider again throwing a fair die

- and let X be a random variable for the number that comes up
  - So X takes values 1,2,3,4,5,6 with equal probability 1/6.
- $E(X) = \sum_{r\in\{1,2,3,4,5,6\}} r \cdot \frac{1}{6} = \frac{7}{2}$.
- $E(X^2) = \sum_{r\in\{1,2,3,4,5,6\}} r^2 \cdot \frac{1}{6} = \frac{1}{6}(1^2 + 2^2 + \cdots + 6^2) = \frac{91}{6}$.
- $V(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$. And $\sigma(X) = \sqrt{\frac{35}{12}} \approx 1.7$

55

Variance in the hat check problem

Variance of a random variable X on a sample space S is: $V(X) = E(X^2) - \left(E(X)\right)^2$

- Example: variance in the hat check problem.
  - Let $R = X_1 + \cdots + X_n$ for indicator variables X_i: man i gets his hat.
  - $E(R) = 1$.
  - $E(R^2) = E(\ (\sum_{i=1}^n X_i\ )^2\ ) = E\left(\sum_{i=1}^n X_i^2 + \sum_{i=1}^n \sum_{j=1,\ j\neq i}^n X_i X_j\right) = E(\sum_{i=1}^n X_i^2) + E(\sum_{i=1}^n \sum_{j=1,\ j\neq i}^n X_i X_j)$
  - $E(X_i^2) = 1^2 \cdot \frac{1}{n} = \frac{1}{n}$, $so\ E(\sum_{i=1}^n X_i^2) = \sum_{i=1}^n E(X_i^2) = 1$
  - $E(X_i X_j) = \frac{1}{n(n-1)}$, so $E\left(\sum_{i=1}^n \sum_{j=1,\ j\neq i}^n X_i X_j\right) = (n^2 - n) \cdot \frac{1}{n(n-1)} = 1$
  - So $E(R^2) = E(\ (\sum_{i=1}^n X_i\ )^2\ ) = 1 + 1 = 2$.

  Finally, $V(R) = E(R^2) - \left(E(R)\right)^2 = 2 - (1)^2 = 1$.

56

Independent random variables

- Similar to independent events:

  - Random variables X, Y are **independent** iff
    $\forall r_1, r_2 \in \mathbb{R}\ \ \Pr(X = r_1 \wedge Y = r_2) = \Pr(X = r_1) \cdot \Pr(Y = r_2)$

- Example: Suppose a pair of fair dice is thrown, with
  X_1: number on the first die, X_2: number on the second die, Y: total amount
  - Then $\Pr(X_1 = i \wedge X_2 = j) = \Pr(X_1 = i) \cdot \Pr(X_2 = j) = \frac{1}{36}$ for all $i, j \in \{1,2,3,4,5,6\}$ which are the only values X_1 and X_2 take. So X_1 and X_2 are independent.
  - But X_1 and Y are not independent: $\Pr(X_1 = 1 \wedge Y = 12) = 0$, whereas
  $\Pr(X_1 = 1) = \frac{1}{6}$, and $\Pr(Y = 12) = \frac{1}{36}$, so $\Pr(X_1 = 1) \cdot \Pr(Y = 12) = \frac{1}{216} \neq 0$.

57

Variance of Bernoulli trials

- When random variables X, Y are independent, then
  - $E(XY) = E(X) \cdot E(Y)$
  - $V(X + Y) = V(X) + V(Y)$

- With this, we can calculate the variance of the number of successes in n Bernoulli trials where in each trial the probability of success is p.
  - For a single trial: let $X = 1$ iff success, $X = 0$ iff failure.
    - Then $E(X) = p$, $E(X^2) = p$, so $V(X) = E(X^2) - \left(E(X)\right)^2 = p - p^2 = p(1-p)$
  - Let $X_1 \ldots X_n$ be indicator variables so that $X_i = 1$ iff i^{th} trial is a success.
  - They are independent, so $V(X_1 + \cdots + X_n) = \sum_{i=1}^n V(X_i) = np(1-p)$

58

Chebyshev's inequality

- Now that we have the notion of variance, we can get a better bound on the probability of being away from the mean:
- **Chebyshev's inequality**: Let $x > 0$. Then

$$\Pr(|R - E(R)| \geq x) \leq \frac{V(R)}{x^2}$$

Proof:

- $|R - E(R)| \geq x$ is equivalent to $\left(R - E(R)\right)^2 \geq x^2$.
- By Markov's inequality, $\Pr\left(\left(R - E(R)\right)^2 \geq x^2\right) \leq \frac{E\left(\left(R - E(R)\right)^2\right)}{x^2} = \frac{V(R)}{x^2}$

59

Better estimate for Hat Check

- **Chebyshev's inequality**: Let $x > 0$. Then

$$\Pr(|R - E(R)| \geq x) \leq \frac{V(R)}{x^2}$$

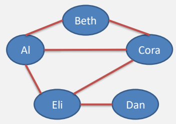
- Example: hat check problem again.
  - Let $R = X_1 + \cdots + X_n$ for indicator variables X_i: man i gets his hat.
  - We have shown that $E(R) = 1$, and that $V(R) = 1$ as well.
  - Let $k \geq 2$. Since $R \geq 0$, $\Pr(R \geq k) = \Pr(|R - 1| \geq k - 1) \leq 1/(k-1)^2$
    - This gives a tighter bound than Markov's inequality.

60

Team selection puzzle

- Imagine that your friend is a project manager, and her team consist of great programmers – if only she could get them to stop fighting among each other!
  - She decides to split them in two smaller teams
    - To minimize fighting within each team.
  - She knows who fights with whom (the "CONFLICT relation"), but how can she do the splitting?
    - And is it possible at all to eliminate at least half the conflicts? If not, why bother…
- Do you think it is possible to split any group into two teams
  - to eliminate all conflicts?
  - How about eliminating half the conflicts?
  - How would you do the splitting?

- Suppose this is the graph of the CONFLICT relation for a group.
  - Here, lines are double-direction arrows, since CONFLICT is symmetric.



  - What do you think is the best split?

61