Counting and Probability

Appendix C

Revised from Prof. Galen Sasaki's slides

1

Introduction

- Probability is a way to model uncertainty
- It can be used to analyze situations
- It can be used to solve problems

2

Example Situations

Coin flipping

Will the coin come up heads?
How many flips before a head?

NBA Free Throw Shooting

How many will Kobe make?

Poker Card Playing

Will the hand be a full house?

3

Algorithms: applied to uncertain situations

Hiring the "best" engineer among 100

- You are to hire a new engineer among 100 candidates
- You interview them one at a time
- At the end of each interview you
  - You evaluate the candidate – giving a score
  - Determine if you will hire or interview the next candidate
  - You cannot go back to a previous candidate

Leader election

- You are in a team that must elect a leader, but everybody is shy
- You each have a coin
- You all flip your coins
- If there is exactly one person with a heads, that's your leader
- Otherwise, you keep flipping coins

4

Algorithms: applied to uncertain situations

Shuffling cards – mixing them up

- Objective: "randomly" mixing them up thoroughly
- What does it mean by randomly mixing them up thoroughly?
- How many shuffles?
  - What do we mean exactly by a shuffle?

- References
- Gilbert-Shannon-Reeds Model
- https://en.wikipedia.org/wiki/Gilbert%E2%80%93Shannon%E2%80%93Reeds_model
- Riffle shuffle permutation
- https://en.wikipedia.org/wiki/Riffle_shuffle_permutation
- Seven shuffles to randomize the deck
- https://en.wikipedia.org/wiki/Persi_Diaconis

5

Digressing: Shuffling

- Magic card trick, shuffling and binary numbers
- https://www.youtube.com/watch?v=Y2lXsxmBx7E
- Persi Diaconis discussion about shuffling cards
- https://www.youtube.com/watch?v=AxJubaijQbI

6

What do we need in our model of an uncertain system?

- Model is of some random experiment taking place
  - Coin flip
  - Entire basketball game
  - A poker hand
- "Sample Space" S = Set of all possible outcomes
  - Coin flip sample space S = {Heads, Tails}

  - Roll of a die sample space S = {1, 2, 3, 4, 5, 6}

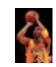
  - Kobe shooting a one for one S = {(no), (yes,no), (yes,yes)}

7

What do we need in our model of an uncertain system?

- An "event" is a subset of the sample space S
  - We're interested in certain events

- Example events
  - Roll of a dice: {outcome is 7} = {(1,6), (2,5),..., (6,1)}
  - Kobe shooting a one for one: { Kobe scores } = { (yes,no), (yes,yes) }
- Special events
  - *Certain event (universe)* = Sample space S
  - *Null event* = Empty set
  - *Elementary events* {x}
    - We often write just write them as x



8

What do we need in our model of an uncertain system?

- For our experiment, for each outcome $x \in S$, we assign a *likelihood* of x
- This likelihood is the *probability* that the outcome x will occur
  - It is a number from [0, 1]
- Examples
  - Flip a fair coin: The likelihood of heads is 0.5
  - Flip an unfair coin: The likelihood of heads is 0.6

  - Throwing dice: The likelihood of (1,1) is 1/36

- Having probabilities for every outcome in S gives us a *probability distribution (True for countable set S)*

9

What do we need in our model of an uncertain system?

- We have
  - Sample space S, all the possible outcomes
  - Probability distribution
- We can compute probability of an *event*, which is a subset of S
- *P(E)* is the probability of event E

$$P(E) = \sum_{x \in E} P(x)$$

$$P(S) = \sum_{x \in S} P(x) = 1$$

- Examples
  - P({Kobe makes at least one free throw}) or P(Kobe makes at least one free throw)
  - P({Poker hand is a full house }) or P(Poker hand is a full house)
  - Roll of dice: outcome (2,5), P({ (2,5) }) or P((2,5))

10

What do we need in our model of an uncertain system?

- How do we get the probabilities?
- We can make it up
  - Create a model that could be applied to different situations
  - Example: flipping multiple coins
    - Applied to coin flipping
    - Applied to physics
    - Applied to people voting
  - Well known probabilities -- Example: uniform probability (see next)
- We can take statistics – (then apply to the future)
  - Take statistics of batting tendencies of baseball players
  - Take statistics of a google search for a word and the likelihood of a purchase of an item

11

Uniform probability distribution

- Applies to a finite sample space S
- Each outcome x is equally likely $\quad P(x) = \frac{1}{N(S)} = \frac{1}{|S|}$
- $N(E)$ = number of elements in E (= $|E|$)
- For an event E,

$$P(E) = \sum_{x \in E} P(x) = \sum_{x \in E} \frac{1}{N(S)} = \frac{1}{N(S)} \sum_{x \in E} 1 = \frac{1}{N(S)} N(E) = \frac{N(E)}{N(S)}$$

- Counting is important to computing probabilities

12

Outline

- Counting (Read Appendix C.1)
- Probability (Read Appendix C.2)
- Discrete Random Variables (Read Appendix C.3)
- Geometric and Binomial Distributions (Read Appendix C.4)

13

Counting

- Strings
- Permutations
- Binomial coefficients
- Grabbing balls out of a bucket –
  - variations, with and without replacement (put a ball back in the bucket or not)
- Pidgeon hole principle

14

Strings

- A **string** over a finite set *S* is a sequence of element of *S*
- Example: Binary strings of length 3, *S* = {0, 1}
  - 000, 001, 010, 011, 100, 101, 110, 111
- String of length *k* is a **k-string**
  - The number of *k*-strings of *S* is $|S|^k$
- A **substring** s' of a string s is an ordered subsequence of s
- A substring of length *k* is a **k-substring**

15

Permutations

- A *permutation* of a finite set S is an ordered sequence of all the elements of *S* with each element appearing exactly once
  - Example: *S* = {a,b,c}.
    - Permutations = abc, acb, bac, bca, cab, cba
- There are n! permutations of n elements
  - Recall n! = n x (n-1) x...x 2 x 1
  - with 0! = 1! = 1
  - Let's check: S has 3! permutations. 3! = 3 x 2 x 1 = 6

16

Permutations

- A ***k-permutation*** of a finite set *S* is an ordered sequence of *k* elements of *S* with each element appearing exactly once
  - Example: *S* = {a,b,c,d}
  - 2-permutations = {ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc}
- The number of *k*-permutations is *P*(*n*,*k*)
  - $P(n,k) = n \times (n-1) \times \ldots \times (n-k+1) = n!/(n-k)!$
  - Let's check:
    - The number of 2-permutations of S is 4!/2! = 4 x 3 = 12

17

Example: Permutations colored balls

- How many ways are there to rearrange *m* red balls, *n* green balls, and *p* yellow balls
  - If all the balls had distinct numbers then the number of ways is $(m+n+p)!$
  - If the red balls are indistinguishable then the number of ways is $(m+n+p)!/m!$
  - If the red balls are indistinguishable from each other and green balls are indistinguishable from each other: $(m+n+p)/(m! \times n!)$
  - If the balls are indistinguishable except for their color: $(m+n+p)!/(m! \times n! \times p!)$

18

Counting

- A *k-combination* of a finite set S is a k-subset of S
- Number of k-combinations of S ("n choose k")

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

  - Example: S = {a,b,c,d}.
    - 2-permutations= ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc
  - Number of 2-permutations = 4!/2! = 12
  - Note that {a,b} is the same as {b,a}
  - Number of 2-combinations = 12/2! = 4!/(2! 2!)

Useful Property

$$\binom{n}{k} = \binom{n}{n-k}$$

19

Binomial Coefficients

- Binomial expansion $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$
  - n terms = (x+y)(x+y)(x+y)...(x+y)
  - Each product term selects an x or y from each term
  - A product term $x^k y^{n-k}$ correspond to a k-subset of terms where x is selected
  - There are $\binom{n}{k}$ of those terms
- Special case

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

20

Binomial Bounds

- $\binom{n}{k} = \left(\frac{n}{k}\right)\left(\frac{n-1}{k-1}\right)\left(\frac{n-2}{k-2}\right)\ldots\left(\frac{n-k+1}{1}\right) \geq \left(\frac{n}{k}\right)^k$
- $\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{k!} \leq \frac{n^k}{k!} \leq \left(\frac{en}{k}\right)^k,$
  - where the last inequality comes from k! ≥ $(k/e)^k$, which comes from Stirling's approximation
- $\binom{n}{k} \leq \frac{n^n}{k^k(n-k)^{n-k}}$ by induction
- Let $k = \alpha k$, where α is a positive fraction
- Then $\binom{n}{k} \leq 2^{H(\alpha)}$, where $H(\alpha) = -\alpha \lg \alpha - (1-\alpha)\lg(1-\alpha)$, the **(binary) entropy function**

21

Theorem: Permutations with sets of indistinguishable objects

- Suppose a collection of n objects of which
  - n_1 are of type 1 and are indistinguishable from each other
  - n_2 are of type 2 and are indistinguishable from each other
  - ...
  - n_k are of type k and are indistinguishable from each other
  - and suppose $n_1 + n_2 + \ldots + n_k = n$
- Then, the number of distinguishable permutations of the n objects is

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\ldots\binom{n-n_1-n_2-\ldots-n_{k-1}}{n_k}$$
$$= \frac{n!}{n_1!n_2!\ldots n_k!}$$

22

r-Combinations with Repetition Allowed

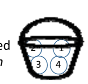
- **Definition:** An ***r-combination with repetition allowed*** (or ***multiset***) of size *r*
  - Chosen from a set *X* of *n* elements
  - Is an unordered selection of elements taken from X with repetition allowed
- Example
  - *X* = {1, 2, 3, 4}
  - 3-combination with repetition allowed
  - Number of combinations: Let's try 4^3 / *m*! for some *m*. It won't work
  - Representation: [Category 1 | Category 2 | Category 3 | Category 4]
    - {1, 2, 2, 4} = [X | X X | | X] --- Number of Xs is the number of occurences
    - Xs and vertical bars ("|") are symbols
    - Number of symbols = $n + r - 1$
  - Number of combinations: $\binom{n+r-1}{r}$

23

r-Combinations with Repetition Allowed

- **Definition**
  - An ***r-combination with repetition allowed*** (or ***multiset***) of size *n*
    - Chosen from a set *X* of *n* elements
  - Is an unordered selection of elements taken from *X* with repetition allowed
- Example

n balls numbered 1, 2, ..., *n*

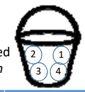


How many ways are to pick *k* balls?

  - Combination: (# balls 1, # balls 2, ..., # balls *n*), and sum of balls = *k*
    - Example: (1, 2, 0, 3) -- number of ball selected k = 6, number of types of balls n = 4
    - Representation: [X | X X | | X] --- "X" = occurrence of a ball, "|" = separates ball types
    - "X" and "|" are symbols, and there are n-1 "|" and k "X"
    - Each representation (of a combination) has $n + k - 1$ symbols or which *n-1* of them are "|"
  - Number of combinations: $\binom{n+k-1}{r}$

24

Review of Formulas



n balls numbered 1, 2, …, n

How many ways are to pick k balls?

	Order matters	Order does not matter
Repetition is allowed (pick a ball and but it back)	n^k A combination: (1st ball, 2nd ball, …, kth ball)	$\binom{n+k-1}{k}$ Subset of k balls, each ball can take a value from {1, 2,…, n} A combination: (# balls 1, # balls 2, …., # balls n) sum of balls = k
Repetition is not allowed	$n!/(n-k-1)!$ A combination: (1st ball, 2nd ball, …, kth ball)	$\binom{n}{k}$ A combination: Subset of k balls

25

The Pigeonhole Principle

- A function from a finite set X to a smaller finite set Y cannot be done one-to-one:
  - There must be at least two elements in the domain that have the same image in the co-domain
- Suppose there are m pigeons and n holes, and $m > n$
  - There must be at least one hole with more than one pigeon
  - Proof by contradiction
    - Suppose all each hole has at most one pigeon
    - Since there are n holes, there is at most n pigeons
    - This contradicts m > n
- Suppose there are m pigeons and n holes
  - For any positive integer k, if $k < n/m$
  - Then there is some hole with at least k+1 pigeons

26

Probability

- Basics
- Axioms
- Probability distribution
- Independence
- Bayes theorem

27

Events

- *Certain event* = Sample space S
- Null event = Empty set
- Mutually exclusive = they're disjoint
  - All elementary events are mutually exclusive of each other
- Elementary events {x}

  We often write just write them as x

28

Axioms

- A *probability distribution* Pr{ } on a sample space S is a mapping from events of S to the unit interval [0, 1]
  - Pr{ A } $\rightarrow$ [0,1]
    - like a function except that its inputs A are subsets
  - Pr{event} models the likelihood of the event occurring
    - Example: Pr{event} = 0.15 means that the event has a 15% chance of occurring
    - Pr{event} is the *probability* of the event
  - Pr{ } is often written as P( ) or P[ ]
  - Consequences
    - For any event A, $P[A] = \sum_{x \in A} P[x]$
    - If A and B are disjoint,

$$P[A \cup B] = \sum_{x \in A \cup B} P[x] = \sum_{x \in A} P[x] + \sum_{x \in B} P[x]$$

$$P[A \cup B] = P[A] + P[B]$$

29

Axioms continued

- A probability distribution Pr{ } on a sample space S must satisfy the following probability axioms:
  - For any event A, P(A) $\geq$ 0
  - P(S) = 1
  - For any two mutually exclusive events A and B, P(A $\cup$ B) = P(A) + P(B)
  - More generally, for any finite or countably infinite sequence of events A_1, A_2, … that are pairwise mutually exclusive, $P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$

30

Example: Flipping a fair coin

- Flip a fair coin n times
- Elementary event = an n-string of Heads and Tails
- For an elementary event x, P(x) = $1/2^n$
- P({exactly k Heads, and (n-k) Tails}) = $2^{-n}\binom{n}{k}$
  k-subset of n flips

31

Simple Results

- Let the *complement* of *A* be $\overline{A} = S - A \quad or \quad S \setminus A$
- Result: $P(\overline{A}) = 1 - P(A)$
- Result: $P(\phi) = 0$
- Result: If $A \subseteq B$ then $P(A) \le P(B)$
- Result: $P(A \cup B) = P(A) + P(B) - P(A \cap B) \le P(A) + P(B)$

(union bound)

32

Discrete Probability Distribution

- A probability distribution is *discrete* if it is defined over a finite or countably infinite sample space S
- Result: $P(A) = \sum_{x \in A} P(x)$
- *Uniform probability distribution*
  - For all elementary events x, P(x) = 1/|S|
  - In other words, each elementary event is equally likely

33

Continuous Uniform Probability Distribution

- The sample space S is over a closed interval [a, b], or [a,b), or (a,b), or (a,b]
  - In many cases, a = 0 and b = 1, i.e., the unit interval
  - Note: S is an uncountable set, and with this distribution P(x) = 0 for any x in the interval
  - Probabilities still satisfy
    - For all subsets A of S, P{A} $\geq$ 0
    - P{S} = 1
    - For mutually disjoint events, $P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$
- P{[c,d]} = $\frac{d-c}{b-a}$ = ratio of length of [c,d] over length of [a,b]
  - It's the fraction of S
  - Note that P{ [c,d] } = P{ (c,d] } = P{ (c,d) } = P{ [c,d) } because P{{c}} = P{{ d}} = 0

34

Conditional Probability, Baye's Formula and Independent Events

- Suppose we have two events *C* and *T*
- How are they statistically related?
- Example:
  - *C* = Current Weather, *T* = Tomorrow's Weather
  - Weather outcomes = { Wet, Dry }
  - Sample space S = { (CW, TW), (CW, TD), (CD, TW), (CD, TD) }

Probabilities		Tomorrow (*T*)	
		Wet	Dry
Current (*C*)	Wet	0.2	0.1
	Dry	0.1	0.6

Sanity Check:
- Probabilities are nonnegative
- Sum of all the probabilities is 1

35

Conditional Probability

- Example continued
- P(*T* = Dry) = 0.1 + 0.6 = 0.7

Probabilities		Tomorrow (*T*)	
		Wet	Dry
Current (*C*)	Wet	0.2	0.1
	Dry	0.1	0.6

- What if you know the current weather: will that help?
- Suppose *C* = Dry
  - We can eliminate cases that won't occur
  - Probability that { *T* = Dry } = 0.6?

Probabilities		Tomorrow (*T*)	
		Wet	Dry
Current (*C*)	Wet	~~0.2~~	~~0.1~~
	Dry	0.1	0.6

    - Wrong since remaining probabilities don't sum to 1
    - Sum of remaining probabilities equals P(*C* = Dry)
    - Normalize remaining probabilities by dividing by P(*C* = Dry) so new probs sum to 1
  - Probability that {*T* = Dry} given {*C* = Dry} = 0.6/P(*C*=Dry) = 0.6/0.7 = 0.857

36

Conditional Probability

- Conditional probability of an event *A* given another event *B*

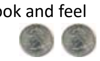
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$ ← Normalization given we know *B* occurs

whenever $P(B) \neq 0$
Also note $P(A \cap B) = P(A \mid B)\, P(B)$

- “P(*A* | *B*)” as the “**probability of A given B**”
- In the example
  - P(Tomorrow = Dry | Current = Dry)
  - B = { Current = Dry }
  - A = { Tomorrow = Dry }

37

Another Example – Baye’s Theorem



- Suppose we have two coins that are identical in look and feel
  - One coin is fair
  - The other coin is biased – always comes up heads
- How can we check which is biased?
  - Suppose are super busy and have time to make only two coin flips
- We can pick up one at random and flip it twice
  - If we get a tails in at least one flip then we have the fair coin
  - What if we get two heads?
  - It makes sense to “guess” that it’s the biased coin but can we quantify this
  - Given {Two heads} what is the probability that it’s the biased coin?

38

Another Example – Baye’s Theorem

- What do we know?
  - We picked up a coin at random
    - *B* = {picked up biased coin}
    - $P(B) = 1/2$
    - B^c = {picked up fair coin }
    - $P(B^c) = 1/2$
  - We got { Two heads }
    - *A* = { Two heads }
    - $P(A \mid B) = 1$
    - $P(A \mid B^c) = 1/4$
    - Assume uniform distribution of all four outcome of two coin flips
  - **But** what are we trying to calculate is $P(B \mid A)$
    - Probability of a biased coin given two heads
    - We have to switch *A* and *B* in the conditional probabilities

39

Baye’s Theorem

- $P(A \mid B) = P(A)\, P(B \mid A) / P(B)$
  - **Proof**
  - $P(A \cap B) = P(A \mid B)\, P(B)$ *[definition of conditional probabilities]*
  - $P(A \cap B) = P(B \mid A)\, P(A)$ *[definition of conditional probabilities]*
  - $P(A \mid B)\, P(B) = P(B \mid A)\, P(A)$
  - Divide both sides by $P(B)$ to get the theorem

40

Example continued -- Apply Baye’s Theorem

- What do we know?
  - Events
    - *B* = {picked up biased coin}
    - B^c = {picked up fair coin }
    - *A* = { Two heads }
  - ***P(B) = 1/2***
  - $P(B^c) = 1/2$
  - ***P(A | B) = 1***
  - $P(A \mid B^c) = 1/4$
  - Baye’s theorem: $P(B \mid A) = P(B)\, P(A \mid B) / P(A)$
    - We need to compute P(A)
    - $P(A) = P(A \mid B)\, P(B) + P(A \mid B^c)\, P(B^c)$
      $= 1 \cdot (1/2) + (1/4) \cdot (1/2) = (4/8) + (1/8) = (5/8)$
    - $P(B \mid A) = (1/2)\,(1) / (5/8) =$ **(4/8) / (5/8) = 4/5 = 0.8**

41

Another useful version of Baye’s Theorem

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(\bar{A})P(B \mid \bar{A})}$$ ← *P(B)*

42

Independence

- Two events A and B are (statistically) **independent** if $P(A \cap B) = P(A)P(B)$
  - This is equivalent to $P(A \mid B) = P(A)$ if $P(B) \neq 0$
  - This means the statistics of event A doesn't change given event B
- A collection of events are **pair-wise independent** if all pairs of events A and B are independent
- A collection of events are **mutually independent** if all subsets of events have the property that $P(\bigcap_{k=1}^{m} A_m) = \prod_{k=1}^{m} P(A_m)$
- Mutual independence implies pair-wise independence
- Pair-wise independence does not imply mutual independence

43

Example

- Example: 3 coin flips
  - P(HHH) = 1/8
    - P(H) = 1/2
  - P(Coin flip 3 = H | Coin Flip 1 = H, Coin Flip 2 = H) = ?
    - = P(coin flips = HHH)/P(first two flips = HH)
    - = (1/8)/(1/4) = 1/2
  - The third coin flip is statistically independent of the first two

44

Discrete Random Variable

- A random variable X for a probability space
  - Sample space S
  - Probability $P(\ )$
- X: $S \rightarrow R$, the real numbers
  - Maps an outcome $x \in S$ to a real number
  - X will take a random value depending on the outcome
- Example: S = {sunny, raining, foggy, snowing}
  - X(sunny) = 2
  - X(raining) = -1
  - X(foggy) = 0
  - X(snowing) = -2

45

Discrete Random Variable

- For all real numbers x, $\{ X = x \}$ is an event
  - $\{ X = x \} = \{ s \in S\text{: } X(s) = x \}$
  - Other important events $\{ X \leq x \}$, $\{ X > x \}$, etc
- $P(\{ X = x \})$ makes sense
  - We often write this short-hand as $P(X = x)$
  - Example: X = Outcome (total) of a dice throw
    - $P(X = 7)$ = probability of {(1,6), (2,5), ..., (6,1)} = 6/36
- Sanity check
  - For real numbers x, $P(X = x) \geq 0$
  - $\sum_x P(X = x) = 1$, where the sum is over all real numbers x
- *Density function* $f_X(x) = P(X = x)$ --- easier notation
  - For real numbers x, $f_X(x) \geq 0$
  - $\sum_x f_X(x) = 1$, where the sum is over all real numbers x

46

Multiple Random Variables

- You can have more than one random variable over a sample space
  - Example: S = {cloudy, windy, sunny, rainy}
    - X(cloudy) = 1, X(windy) = 0, X(sunny) = 0, X(rainy) = 1
    - Y(cloudy) = 0, Y(windy) = -1, Y(sunny) = 2, X(rainy) = -2
- **Joint density function** is over multiple random variables
  - $f_{XY}(x,y) = P(X = x \text{ and } Y = y) = P(X = x, Y = y)$
    - comma means "and"
- ***Marginal density function*** for fixed $X = x$
  - $f_X(x) = P(X = x) = \sum_y P(X = x, Y = y) = \sum_y f_{XY}(x, y)$

47

Joint Distribution Example

$f_{XY}(\ )$	$X = 0$	$X = 1$	$X = 2$	$f_Y(\ )$
Y = 0	0.052	0.124	0.111	0.287
Y = 1	0.034	0.073	0.118	0.225
Y = 2	0.207	0.168	0.113	0.488
$f_X(\ )$	0.293	0.365	0.342	

Sum of rows

Sum of columns

48

Functions of Random Variables

- All functions of random variables are random variables
- Example: Suppose X and Y are random variables
  - $X + Y$
  - $\max(X, Y)$
  - XY + constant
  - $g(X,Y)$

49

Expectations

- ***Expected value*** (or ***mean***) of a random variable X
  - $\sum_x xP(X = x)$
  - $\sum_x xf_X(x)$
  - Represents the "average" value
  - It's a constant

50

Linearity of Expectations

- If X and Y are random variables and a and b are constants
- E[a X + b Y] = a E[X] + b E[Y]
- Proof on the next slide

51

Verify

$$
\begin{aligned}
E[aX+bY] &= \sum_x \sum_y (ax+by)P(X=x,Y=y) \\
&= \sum_x \sum_y [axP(X=x,Y=y) + byP(X=x,Y=y)] \\
&= \sum_x \sum_y axP(X=x,Y=y) + \sum_x \sum_y byP(X=x,Y=y)] \\
&= \sum_x \sum_y axP(X=x,Y=y) + \sum_y \sum_x byP(X=x,Y=y)] \\
&= a\sum_x x\underbrace{\sum_y P(X=x,Y=y)} + b\sum_y y\underbrace{\sum_x P(X=x,Y=y)]} \\
&= a\sum_x xP(X=x) + b\sum_y yP(Y=y) \\
&= aE[X] + bE[Y]
\end{aligned}
$$

52

Conditional Probability and Independence

- *Conditional probability*

$$f_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f_{XY}(x,y)}{f_Y(y)}$$

- Two random variables (X,Y) are *(statistically) independent* if for all (x,y), $P(X = x, Y = y) = P(X = x)P(Y = y)$
  - Consequences $f_{XY}(x,y) = f_X(x)f_Y(y)$
- The definition of independence is extended to multiple variables – mutual independence

$$f_{X|Y}(x \mid y) = f_X(x)$$

53

Product of Random Variables

- Suppose X and Y are independent
  - $E[XY] = E[X]\,E[Y]$
- Suppose $X(1),\ldots, X(n)$ are mutually independent
  - $E[X(1)\,X(2)\ldots X(n)] = E[X(1)]\,E[X(2)]\ldots E[X(n)]$

54

E[XY] = E[X]E[Y]

$$E[XY] = \sum_x \sum_y xy f_{XY}(x,y)$$

Independence

$$= \sum_x \sum_y xy f_X(x) f_Y(y)$$

Factor out terms not dependent on y

$$= \sum_x x f_X(x) \sum_y y f_Y(y)$$

Definition of E[Y]

$$= \sum_x x f_X(x) E[Y]$$

Factor out the constant E[Y]

$$= E[Y] \sum_x x f_X(x)$$

Definition of E[X]

$$= E[Y]E[X]$$

55

Indicator Functions

- Indicator function I{ A } = 1 if A is true, and 0 if A is false
- Let's show E[X] = $\sum_{i=1}^{\infty} P\{X \geq i\}$
- E[X] = $\sum_{i=0}^{\infty} iP\{X = i\}$

$$= \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} I\{j \leq i\} P\{X = i\}$$

$$= \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} I\{j \leq i\} P\{X = i\}$$

$$= \sum_{j=1}^{\infty} P\{X \geq j\}$$

56

Variance

Variance $E[(X - E[X])^2] = E[X^2] - (E[X])^2$

Is a kind of measure of uncertainty
Measures how far away X can be from its mean

$Var[aX + b] = a^2 Var[X]$ Constants *a* and *b*

For pair-wise independent random variables $Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i]$

57

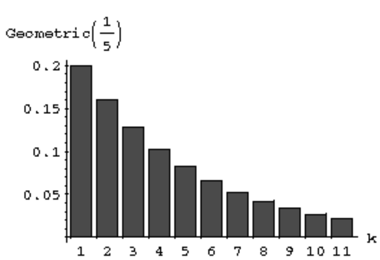
Geometric and Bernoulli Distributions

- Coin flip can be thought of as a *Bernoulli trial*
  - Probability *p* it's heads (success)
  - Probability *q* = 1-*p* it's tails (failure)
- Let *G* be the random variable that equals the number of independent trials before a success
  - *P*(*G* = 1) = *p*
  - *P*(*G* = 2) = *q* x *p*
  - *G* is a *geometric* random variable with a geometric distribution

$$P(G = k) = p \cdot q^{k-1}$$

58

Geometric Probability Distribution



59

Expectations of a Geometric Random Variable

$$E[G] = \sum_{k=1}^{\infty} kP[G = k]$$

Definition of distn

$$= \sum_{k=1}^{\infty} kpq^{k-1}$$

Factor out p

$$= p\sum_{k=1}^{\infty} kq^{k-1}$$

Use this identity from taking derivative

$$= p\sum_{k=1}^{\infty} \frac{\partial}{\partial q} q^k$$

Move derivative outside the sum. Okay since derivative is a kind of sum

$$= p\frac{\partial}{\partial q}\sum_{k=1}^{\infty} q^k$$

By recursion

$$= p\frac{\partial}{\partial q}\frac{q}{1-q}$$

Take derivative

$$= p\frac{1}{(1-q)^2}$$

$$= p\frac{1}{(p)^2} = \frac{1}{p}$$

60

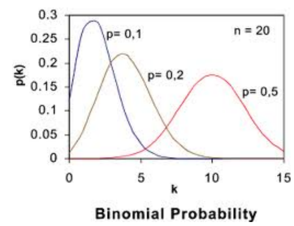
Binomial Distribution

- n Bernoulli trials (independent)
- *X* = number of successes
- *X* has the Binomial distribution $P(X = k) = \binom{n}{k} p^k \cdot q^{n-k}$

- *E*[*X*] = *np*
  - *X* = 1(trial 1 is true) + ... + 1(trial n is true)
  - 1(*F*) is an *indicator* function
    - Equals 1 if *F* is true
    - Equals 0 otherwise
    - Useful because E[1(F)] = 1 x P(F) + 0 = P(F)

61

Binomial Distribution



62

Example

- ALOHA network
- Slotted ALOHA protocol
- Tree algorithm

63

ALOHA System (ALOHA network)

- First public demonstration of a wireless network in 1971, University of Hawaii
  - Norm Abramson
- Systems influenced by ALOHA
  - Ethernet
  - SMS messaging
  - GPRS – data access for GSM systems
  - Various satellite networks

64

Motivation for ALOHA

- Scenario: A main computer connected to terminals by wireless communication
  - Given a wireless bandwidth B
  - # terminals = n
  - Time division multiplexing (TDM) or frequency division multiplexing (FDM)
    - An existing solution in 1971
    - Splits the bandwidth B into n subchannels with bandwidths B/n (or approximately)
    - A GSM cell phone connection works this way
    - Wastes bandwidth if terminals are not in constant use, e.g., data traffic is "bursty"

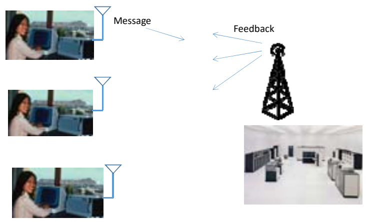
65

ALOHA designed for data

- Protocol (called "random access ALOHA")
  - Terminal
    - When it has something to send, it sends it at full bandwidth B
    - It waits for the main computer for feedback about whether the transmission was a success or not ("collision")
    - If not then it waits a random amount of time then retransmits
    - It keeps trying after random waits until the message gets through to the main computer
  - Main Computer
    - Sends feedback about whether its received transmissions were successful or not

66

ALOHA



67

Slotted ALOHA

- System assumptions
  - Time is divided into equal length time slots
  - A packet can be transmitted in a slot
  - Packet sizes are fixed length equal to the transmission in a slot
  - A collision occurs in a slot if two or more terminals transmit
  - A packet is successfully transmitted in a slot if exactly one packet is transmitted

68

Slotted ALOHA

- Protocol
  - Packets are transmitted in FIFO order
  - If the packet at the head of the FIFO is new then it is transmitted
  - If there is a collision then the terminal retransmits the packet after a random delay
    - Terminal keeps trying until the packet gets through
- Variation
  - Packets are transmitted in FIFO order
  - A packet at the head of the FIFO (whether new or not) is transmitted with probability p

69

Slotted ALOHA

- Variation
  - Packets are transmitted in FIFO order
  - A packet at the head of the FIFO is transmitted with probability p
- What should p be?
  - Let's assume that the network is heavily congested, i.e., all n terminals have lots of packets in their FIFO queues
  - P(success in a slot) = $\binom{n}{1}p(1-p)^{n-1}$
  - Maximize with respect to p
    - Take derivative wrt *p* and set to zero

70

Slotted ALOHA

- What should p be?
  - Let's assume that the network is heavily congested, i.e., all terminals have lots of packets in their FIFO queues
  - P(success in a slot) = $\binom{n}{1}p(1-p)^{n-1}$
- p = 1/n maximizes P(success)
- P(success) = $\binom{n}{1}p(1-p)^{n-1}$

$$= \left(1-\frac{1}{n}\right)^{n-1}$$

$$= e^{-1} = 0.367$$

= "Throughput"

71

Tree Algorithm

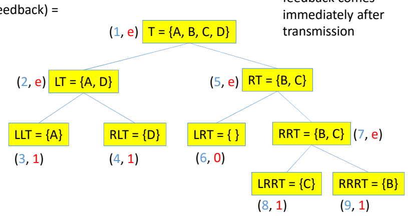
- ALOHA is a way to choose one among a collection of terminals in distributed way
- Three algorithm is another approach
- Suppose there is a set T of n terminals that want to transmit
  - No other terminals want to transmit
- They transmit in the next slot
  - If n > 1 then there is a collision
  - Set T splits in two: left subset LT and right subset RT
    - By coin flipping
  - Left subset LT transmits before right subset RT

72

Example

(Time slot, Outcome)
Outcome (feedback) =
0 (idle),
1 (success),
e (collision)

Assume the feedback comes immediately after transmission

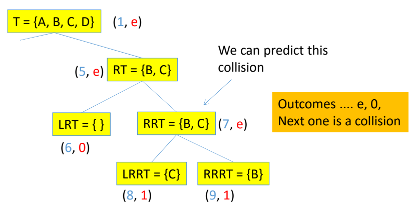


Stack Implementation

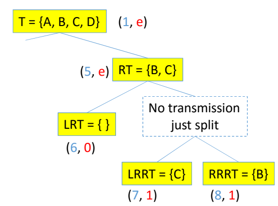
Transmit the top of the stack
Collision results in splitting with right side going into the stack, and the left side on the top of the stack

Slot	Outcome	Top of Stack	Rest of Stack
1	e	{A,B,C,D}	
2	e	{A,D}	{B,C}
3	1	{A}	{D}, {B,C}
4	1	{D}	{B,C}
5	e	{B,C}	
6	0	{ }	{B,C}
7	e	{B,C}	
8	1	{C}	{B}
9	1	{B}	

Improvement



Improvement



Terminal's Algorithm

- Keeps track of where it is in the stack using a counter (0 = top of the stack)
- counter = 0 means transmit
- Modify counter based on Outcome
  - Outcome = e
    - counter = 0 means flip a coin
      - Tails (right side) means counter = counter + 1
      - Heads (left side) keep counter = 0
    - counter > 0 means counter = counter + 1
  - Outcome = 0 or 1
    - counter = counter – 1

Tree Algorithm

- Better throughput than slotted ALOHA (around 0.43 throughput)
- Interesting observations
  - Slotted ALOHA and Tree Algorithm are *distributed* algorithms
    - ALOHA is more robust to errors
  - Nodes in the tree algorithm are coordinated because
    - All nodes use the same algorithm
    - All have the same input (the outcomes of the transmissions)