Probability

1 Introduction to Discrete Probability

- Finite Probability
- Probability of Combination of Events
- Probabilistic Reasoning – Car & Goats

Terminology

Experiment

- A repeatable procedure that yields one of a given set (
- Rolling a die, for example

Sample space

- The set of possible outcomes
- For a die, that would be values 1 to 6

Event

- A subset of the sample experiment
- If you rolled a 4 on the die, the event is the 4

Probability

Experiment: We roll a single die, what are the possible outcomes?

{1,2,3,4,5,6}

The set of possible outcomes is called the **sample space.**

We roll a pair of dice, what is the sample space?

Depends on what we're going to ask.

Often convenient to choose a sample space of equally likely outcomes.

{(1,1),(1,2),(1,3),…,(2,1),…,(6,6)}

Probability definition: Equally Likely Outcomes

The probability of an event occurring (assuming equally likely outcomes) is:

$$p(E) = \frac{|E|}{|S|}$$

- Where E an event corresponds to a subset of outcomes.
  Note: $E \subseteq S$.
- Where S is a finite sample space of equally likely outcomes
- Note that $0 \leq |E| \leq |S|$
  - Thus, the probability will always between 0 and 1
  - An event that will never happen has probability 0
  - An event that will always happen has probability 1

Probability is always a value between 0 and 1

Something with a probability of 0 will never occur

Something with a probability of 1 will always occur

You cannot have a probability outside this range!

Note that when somebody says it has a "100% probability"

- That means it has a probability of 1

Dice probability

What is the probability of getting a 7 by rolling two dice?

- There are six combinations that can yield 7: (1,6), (2,5), (3,4), (4,3), (5,2), (6,1)
- Thus, $|E| = 6$, $|S| = 36$, $P(E) = 6/36 = 1/6$

Probability

Which is more likely:

Rolling an 8 when 2 dice are rolled?

Rolling an 8 when 3 dice are rolled?

No clue.

Probability

What is the probability of a total of [illegible] when 2 dice are rolled?

What is the size of the sample space? 36

How many rolls satisfy our property of interest? 5

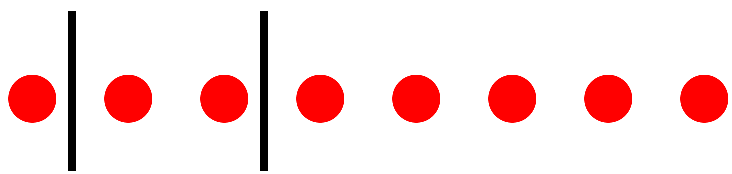
So the probability is 5/36 ≈ 0.139.

Probability

What is the probability of a total of 8 when 3 dice are rolled?

What is the size of the sample space? 216

How many rolls satisfy our condition of interest? C(7,2)



So the probability is 21/216 ≈ 0.097.

Poker probability: royal flush

What is the chance of getting a royal flush?

- That's the cards 10, J, Q, K, and A of the same suit

There are only 4 possible royal flushes.

Possibilities for 5 cards: C(52,5) = 2,598,960

Probability = 4/2,598,960 = 0.0000015

- **Or about 1 in 650,000**

Poker hand odds

The possible poker hands are (in increasing order):

– Nothing	1,302,540	0.5012
– One pair	1,098,240	0.4226
– Two pair	123,552	0.0475
– Three of a kind	54,912	0.0211
– Straight	10,200	0.00392
– Flush	5,108	0.00197
– Full house	3,744	0.00144
– Four of a kind	624	0.000240
– Straight flush	36	0.0000139
– Royal flush	4	0.00000154

Event Probabilities

Let E be an event in a sample space S. The probability of the complement of E is:

$$p(\overline{E}) = 1 - p(E)$$

Recall the probability for getting a royal flush is 0.0000015

- The probability of *not* getting a royal flush is 1-0.0000015 or 0.9999985

Recall the probability for getting a four of a kind is 0.00024

- The probability of *not* getting a four of a kind is 1- 0.00024 or 0.99976

Probability of the union of two events

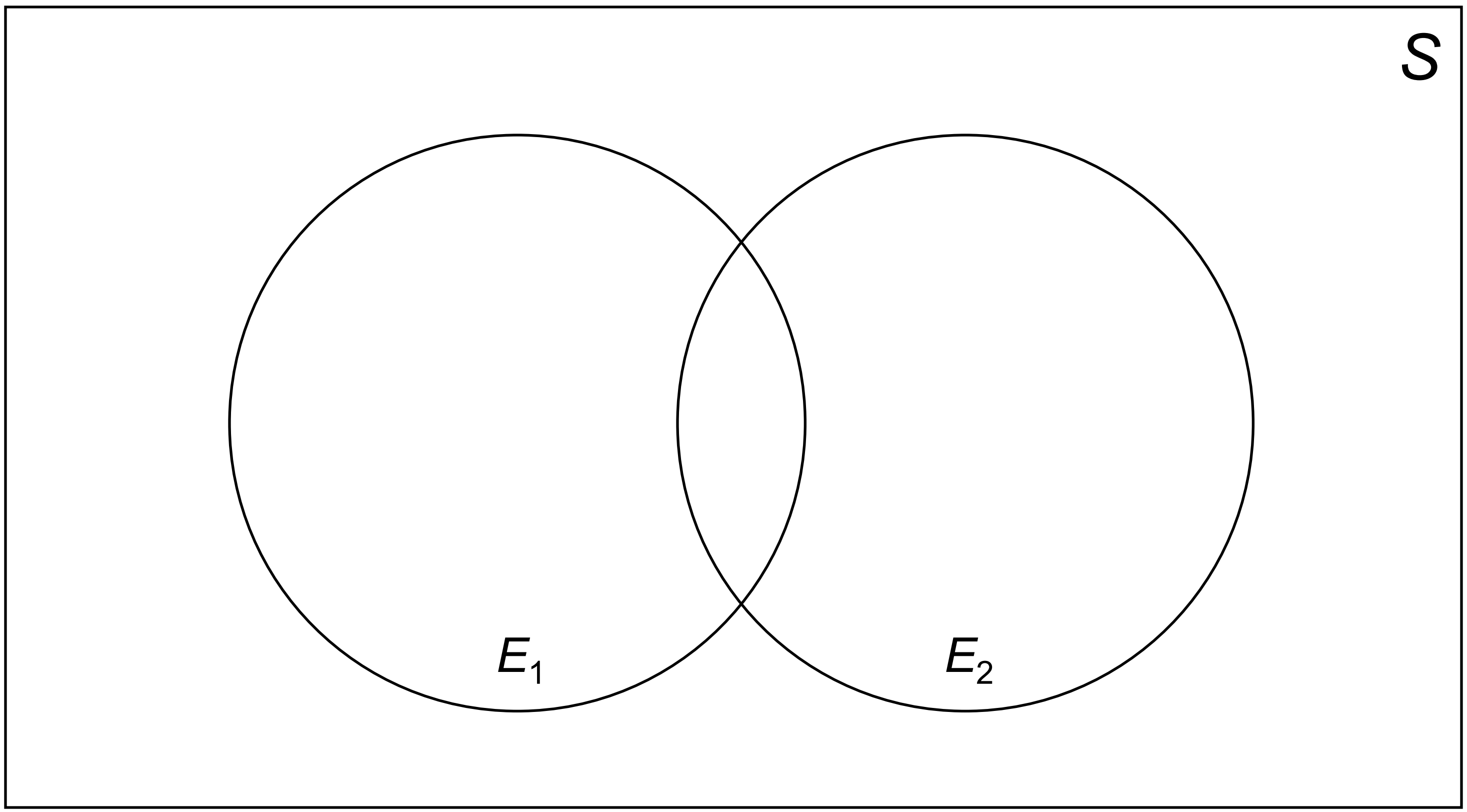
Let E_1 and E_2 be events in sample space S

Then $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$

Consider a Venn diagram dart-board

Probability of the union of two events

$p(E1 \cup E2)$



Probability of the union of two events

If you choose a number between 1 and 100, what is the probability that it is divisible by 2 or 5 or both?

Let n be the number chosen

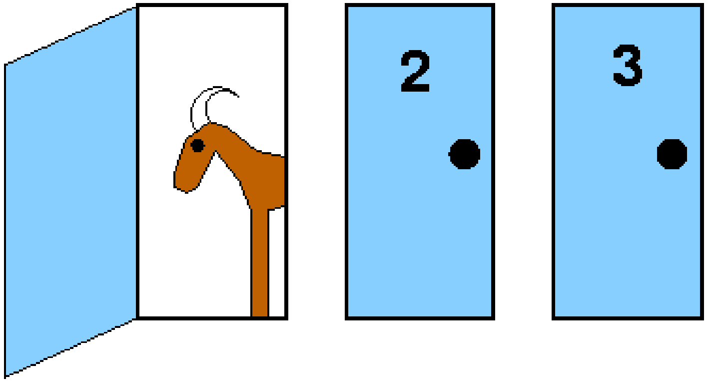
- $p(2 \text{ div } n) = 50/100$ (all the even numbers)
- $p(5 \text{ div } n) = 20/100$
- $p(2 \text{ div } n)$ and $p(5 \text{ div } n) = p(10 \text{ div } n) = 10/100$
- $p(2 \text{ div } n)$ or $p(5 \text{ div } n) = p(2 \text{ div } n) + p(5 \text{ div } n) - p(10 \text{ div } n)$

$$= 50/100 + 20/100 - 10/100$$

$$= 3/5$$

Probability
Monte Hall Puzzle

Choose a door to win a prize!



Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 3, and the host, who knows what's behind the doors, opens another door, say No. 1, which has a goat. He then says to you, "Do you want to pick door No. 2?"

Is it to your advantage to switch your choice? **If so, why? If not, why not?**

6.2 Probability Theory Topics

- Assigning Probabilities: Uniform Distribution
- Combination of Events - - - covered in 6.1
- Conditional Probability
- Independence
- Bernoulli Trials and the Binomial Distribution
- Random Variables – Added
- The Birthday Problem – Added
- Monte Carlo Algorithms – NOT ADDED
- The Probabilistic Method: NOT ADDED - Use in creating non-constructive existence proofs

Probability: General notion (non necessarily equally likely outcomes)

Define a probability measure on a set S to be a real-valued function, Pr, with domain 2^S so that:

For any subset A in 2^S, $0 \leq \Pr(A) \leq 1$.

$\Pr(\varnothing) = 0$, $\Pr(S) = 1$.

If subsets A and B are disjoint, then

$\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

Pr(A) is "the probability of event A."

A sample space, together with a probability measure, is called a <u>probability space</u>.

$S = \{1,2,3,4,5,6\}$

For $A \subseteq S$, $\Pr(A) = |A|/|S|$

(equally likely outcomes)

Ex. "Prob of an odd #"

$A = \{1,3,5\}$, $\Pr(A) = 3/6$

Aside: book first defines Pr per outcome.

Definition:

Suppose S is a set with *n* elements. The ***uniform distribution*** assigns the probability 1/n to each element of S.

The experiment of selecting an element from a sample space with a uniform a distribution is called selecting an element of S ***at random***.

When events are equally likely and there a finite number of possible outcomes, the second definition of probability coincides with the first definition of probability.

Alternative definition:

The probability of the event E is the sum of the probabilities of the outcomes in E. Thus

$$p(E) = \sum_{s \in E} p(s)$$

Note that when E is an infinite set, $\sum_{s \in E} p(s)$ is a convergent infinite series

Probability

As before:

If A is a subset of S, let ~A be the complement of A wrt S.

Then $\Pr(\sim A) = 1 - \Pr(A)$

Inclusion-Exclusion

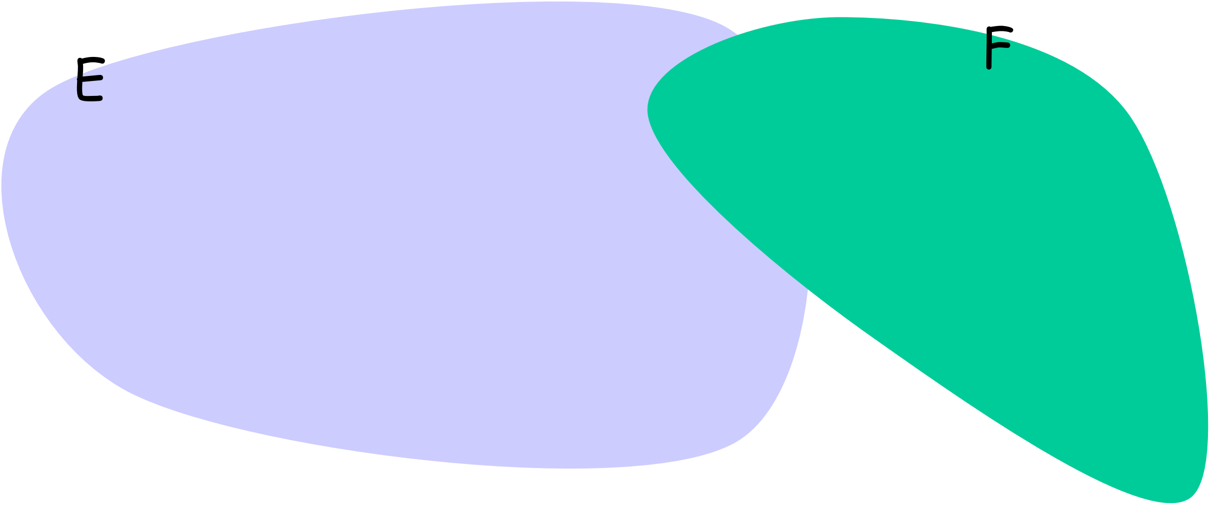
If A and B are subsets of S, then

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

Conditional Probability

Let E and F be events with Pr(F) > 0. The conditional probability of E given F, denoted by Pr(E|F) is defined to be:

$$\Pr(E|F) = \Pr(E \cap F) / \Pr(F).$$



Example: Conditional Probability

A bit string of length 4 is generated at random so that each of the 16 bit possible strings is equally likely. What is the probability that it contains at least two consecutive 0s, ***given that its first bit is a 0***?

So, to calculate:

$$\Pr(E|F) = \Pr(E \cap F) / \Pr(F).$$

where

F is the event that "first bit is 0", and

E the event that "string contains at least two consecutive 0s".

What is "the experiment"?

The random generation of a 4 bit string.

What is the "sample space"?

The set of all all possible outcomes, i.e., 16 possible strings. (equally likely)

A bit string of length 4 is generated at random so that each of the 16 bit strings is equally likely. What is the probability that it contains at least two consecutive 0s, ***given that its first bit is a 0***?

So, to calcuate:

$$\Pr(E|F) = \Pr(E \cap F) / \Pr(F).$$

where F is the event that first bit is 0 and E the event that string contains at least two consecutive 0's.

$\Pr(F) = ?$ 1/2

$\Pr(E \cap F)?$ 0000 0001 0010 0011 0100 (note: 1st bit fixed to 0)

$\Pr(E \cap F) = 5/16$ $\Pr(E|F) = 5/8$

Why does it go up?
Hmm. Does it?

1000 1001 ~~1010~~ ~~1011~~ 1100

So, $P(E) = 8/16 = 1/2$

A bit string of length 4 is generated at random so that each of the 16 bit strings is equally likely. What is the probability that the first bit is a 0, **given that it contains at least two consecutive 0s?**

So, to calculate:

$$\Pr(F|E) = \Pr(E \cap F) / \Pr(E)$$

$$= (\Pr(E|F) * \Pr(F)) / \Pr(E) \quad \text{Bayes' rule}$$

where F is the event that first bit is 0 and E the event that string contains at least two consecutive 0's.

We had:

$\Pr(E \cap F) = 5/16$

$\Pr(E|F) = 5/8$

$\Pr(F) = 1/2$

$\Pr(E) = 1/2$

So, $P(F|E) = (5/16) / (1/2) = 5/8$

$= ((5/8) * (1/2)) / (1/2)$

So, all fits together.

Sample space	F	E	E∩F)		
0000	0000	0000	0000	0000	0000
0001	0001	0001	0001	0001	0001
0010	0010	0010	0010	0010	0010
0011	0011	0011	0011	0011	0011
0100	0100	0100	0100	0100	0100
0101	0101			0101	
0110	0110		$Pr(E \cap F) = 5/16$	0110	
0111	0111			0111	
1000		1000			1000
1001	$P(F) = 1/2$	1001		$P(E\|F) = 5/8$	1001
1010					
1011					
1100		1100			1100
1101					
1110		$P(E) = 1/2$			$P(F\|E) = 5/8$
1111					

Independence

The events E and F are independent if and only if

$$\Pr(E \cap F) = \Pr(E) \times \Pr(F).$$

Note that in general: $\Pr(E \cap F) = \Pr(E) \times \Pr(F|E)$ (defn. cond. prob.)

So, independent iff $\Pr(F|E) = \Pr(F).$

(Also, $\Pr(F|E) = \Pr(E \cap F) / P(E) = (\Pr(E) x \Pr(F)) / P(E) = \Pr(F)$)

Example: P("Tails" | "It's raining outside") = P("Tails").

Independence

The events E and F are independent if and only if

$$\Pr(E \cap F) = \Pr(E) \times \Pr(F).$$

Let E be the event that a family of n children has children of both sexes.
Lef F be the event that a family of n children has at most one boy.

Are E and F independent if

$n = 2$? No Hmm. Why?

$S = \{(b,b), (b,g), (g,b), (g,g)\}$, $E = \{(b,g), (g,b)\}$, and $F = \{(b,g), (g, b), (g,g)\}$

So $\Pr(E \cap F) = ½$ and $\Pr(E) \times \Pr(F) = ½ \times ¾ = 3/8$

Independence

The events E and F are independent if and only if $\Pr(E \cap F) = \Pr(E) \times \Pr(F)$.

Let E be the event that a family of n children has children of both sexes.
Let F be the event that a family of n children has at most one boy.
Are E and F independent if

$n = 3$? Yes !!

Independence

The events E and F are independent if and only if $Pr(E \cap F) = Pr(E) \times Pr(F)$.

Let E be the event that a family of n children has children of both sexes.

Lef F be the event that a family of n children has at most one boy.

Are E and F independent if

So, dependence / independence really depends on detailed structure of the underlying probability space and events in question!! (often the only way is to "calculate" the probabilities to determine dependence / independence.

$n = 4$? No

$n = 5$? No

Bernoulli Trials

A ***Bernoulli trial*** is an experiment, like flipping a coin, where there are two possible outcomes. The probabilities of the two outcomes could be different.

Bernoulli Trials

A coin is tossed 8 times.

What is the probability of exactly 3 heads in the 8 tosses?

THHTTHTT is a tossing sequence...

How many ways of choosing 3 positions for the heads?

$C(8,3)$

What is the probability of a particular sequence?

$.5^8$

In general: The probability of exactly k successes in n independent Bernoulli trials with probability of success p, is

$$C(n,k)p^k(1-p)^{n-k}$$

Bernoulli Trials and Binomial Distribution

Bernoulli Formula: Consider an experiment which repeats a Bernoulli trial n times. Suppose each Bernoulli trial has possible outcomes *A, B* with respective probabilities *p* and 1-*p*. The probability that *A* occurs exactly *k* times in *n* trials is

$$C(n,k)\, p^{k} \cdot (1-p)^{n-k}$$

Binomial Distribution: denoted by b(k;n;p) – this function gives the probability of k successes in n independent Bernoulli trials with probability of success p and probability of failure q = 1- p

$$b(k;n;p) = C(n,k)\, p^{k} \cdot (1-p)^{n-k}$$

Bernoulli Trials

Consider flipping a fair coin n times.

A = coin comes up "heads"

B = coin comes up "tails"

$p = 1\text{-}p = ½$

Q: What is the probability of getting exactly 10 heads if you flip a coin 20 times?

Recall: P (A occurs k times out of n)

$$= C(n,k)\, p^{k} \cdot (1-p)^{n-k}$$

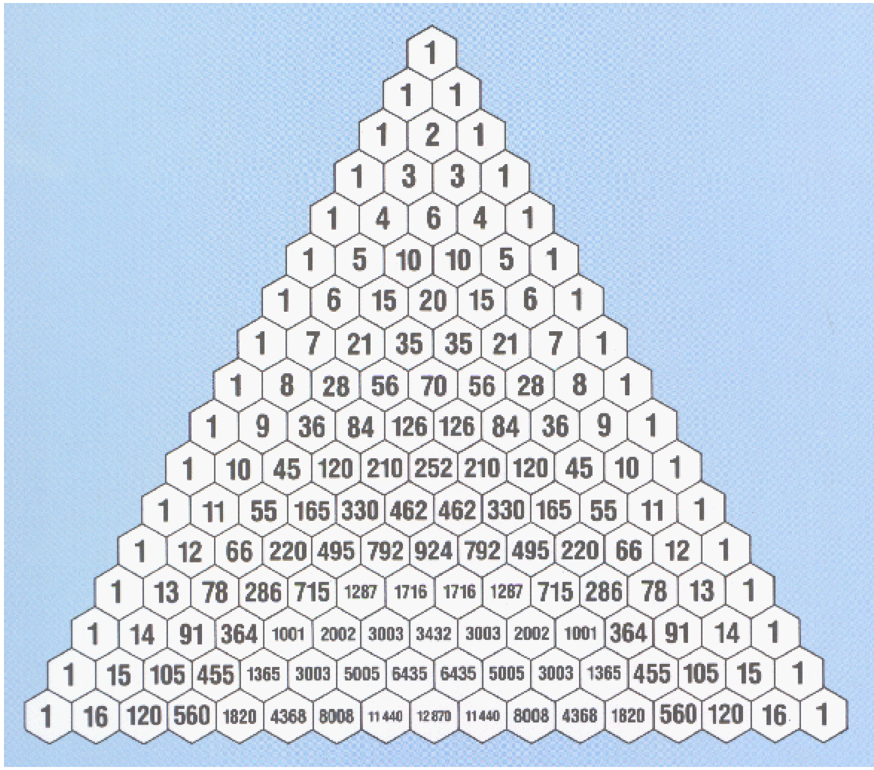
Bernoulli Trials: flipping fair coin

A: $(1/2)^{10} \cdot (1/2)^{10} \cdot C(20,10)$

$= \quad 184756 / 2^{20}$

$= \quad 184756 / 1048576$

$= \quad 0.1762\ldots$



Consider flipping a coin **n** times.

What is the most likely number of heads occurrence?

$n/2$

What probability?

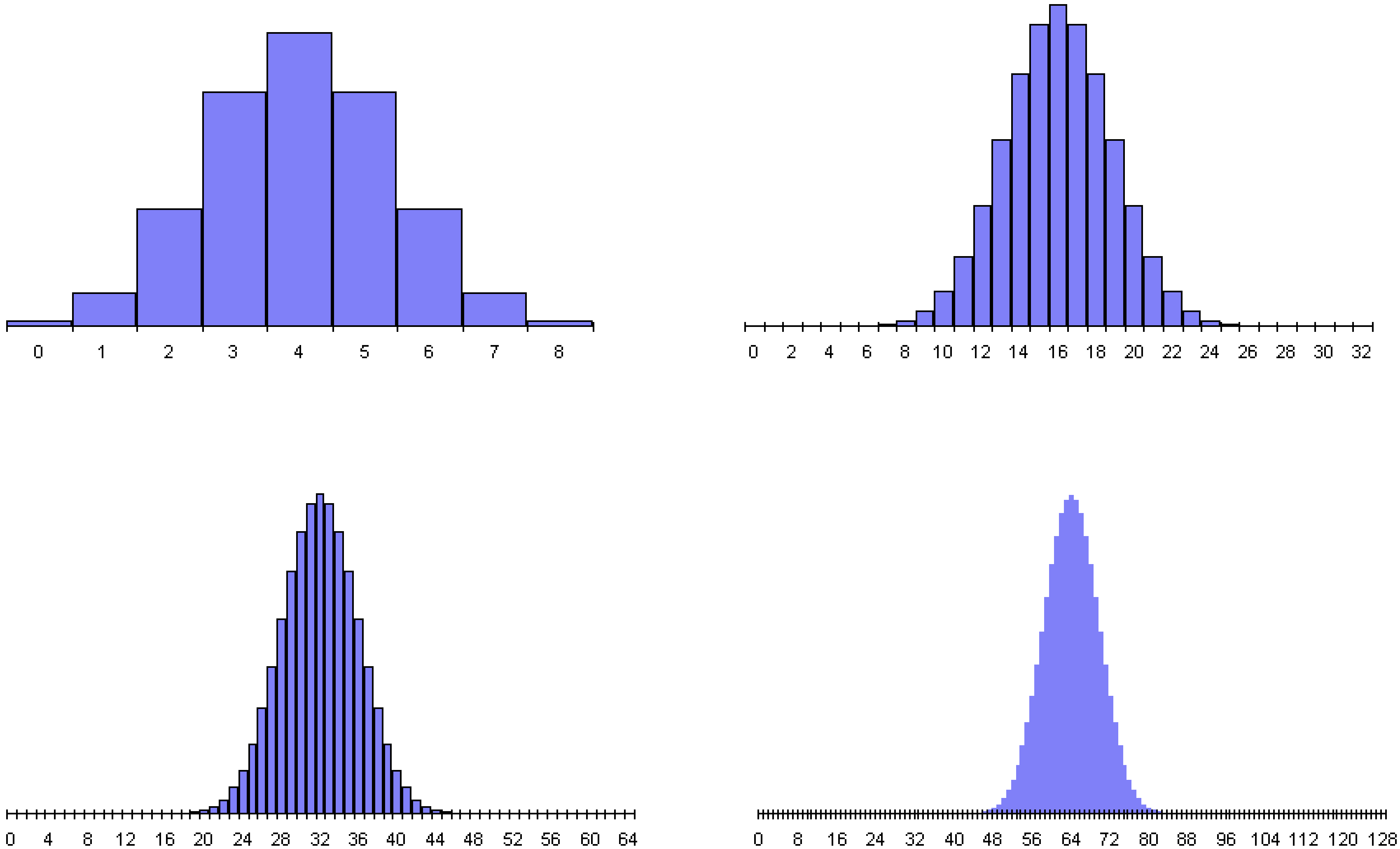
$C(n, n/2) \cdot (1/2)^n$

What is the least likely number?

0 or n

What probability?

$(1/2)^n$ (e.g. for n = 100 … it's "never")



What’s the “width”? …

O(sqrt(n))

Suppose a 0 bit is generated with probability 0.9 and a 1 bit is generated with probability 0.1., and that bits are generated independently. What is the probability that exactly eight 0 bits out of ten bits are generated?

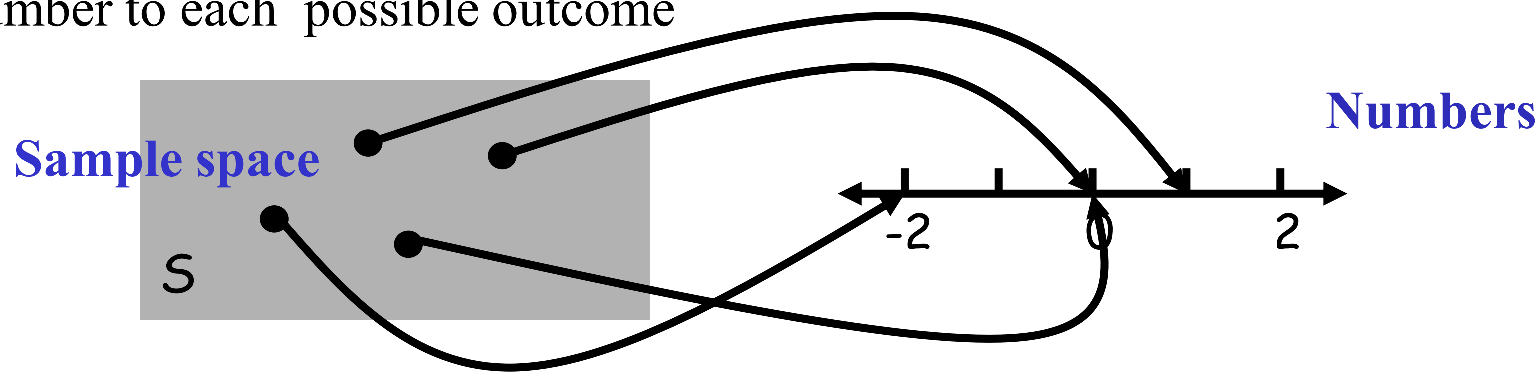
$b(8;10;0.9) = C(10,8)(0.9)^8(0.1)^2 = 0.1937102445$

Random Variables & Distributions
Also Birthday Problem

Added from Probability Part (b)

Random Variables

For a given sample space S, a random variable (r.v.) is any real valued function on S, i.e., a random variable is a function that assigns a real number to each possible outcome



Suppose our experiment is a roll of 2 dice. S is set of pairs.

Example random variables:

X = sum of two dice.	X((2,3)) = 5
Y = difference between two dice.	Y((2,3)) = 1
Z = max of two dice.	Z((2,3)) = 3

Random variable

Suppose a coin is flipped three times. Let X(t) be the random variable that equals the number of heads that appear when t is the outcome.

X(HHH) = 3

X(HHT) = X(HTH)=X(THH)=2

X(TTH)=X(THT)=X(HTT)=1

X(TTT)=0

Note: we generally drop the argument! We'll just say the "random variable X".

And write e.g. P(X = 2) for "the probability that the random variable X(t) takes on the value 2".

Or P(X=x) for "the probability that the random variable X(t) takes on the value x."

Distribution of Random Variable

Definition:

The distribution of a random variable X on a sample space S is the set of pairs (r, p(X=r)) for all $r \in X(S)$, where p(X=r) is the probability that X takes the value r.

A distribution is usually described specifying p(X=r) for each $r \in X(S)$.

A probability distribution on a r.v. X is just an allocation of the total probability mass, 1, over the possible values of X.

The Birthday Paradox

A: 23

Birthdays

How many people have to be in a room to assure that the probability that at least two of them have the same birthday is greater than 1/2?

a) 23
b) 183
c) 365
d) 730

Let p_n be the probability that no people share a birthday among n people in a room.

Then $1 - p_n$ is the probability that 2 or more share a birthday.

For L options answer is in the order of sqrt(L) ?

We want the smallest n so that $1 - p_n > 1/2$.

Informally, why??

Hmm. Why does such an n exist? Upper-bound?

Birthdays

Assumption:

Birthdays of the people are independent.

Each birthday is equally likely and that there are 366 days/year

Let p_n be the probability that no-one shares a birthday among n people in a room.

What is p_n? ("brute force" is fine)

Assume that people come in certain order; the probability that the second person has a birthday different than the first is 365/366; the probability that the third person has a different birthday form the two previous ones is 364/366.. For the jth person we have (366-(j-1))/366.

So, $$p_n = \frac{365}{366}\frac{364}{366}\frac{363}{366}\cdots\frac{367-n}{366}$$

$$1-p_n = 1-\frac{365}{366}\frac{364}{366}\frac{363}{366}\cdots\frac{367-n}{366}$$

After several tries, when n=22 $1 = p_n = 0.475$.

n=23 $1-p_n = 0.506$

Relevant to “hashing”. Why?

From Birthday Problem to Hashing Functions

Probability of a Collision in Hashing Functions

A hashing function h(k) is a mapping of the keys (or records, e.g., SSN, around
300x 10^6 in the US) to a much smaller storage location. A good hashing fucntio
yields few collisions. What is the probability that no two keys are mapped
to the same location by a hashing function?

Assume m is the number available storage locations, so the probability
of mapping a key to a location is $1/m$.
Assuming the keys are $k1, k2, kn,$ the probability of mapping the jth record to a
free location is after the first (j-1) records is (m-(j-1))/m.

$$p_n = \frac{m-1}{m}\frac{m-2}{m}\cdots\frac{m-n+1}{m}$$

$$1-p_n = 1-\frac{m-1}{m}\frac{m-2}{m}\cdots\frac{m-n+1}{m}$$

Given a certain m, find the smallest n
Such that the probability of a collision
is greater than a particular threshold p.

It can be shown that for p>1/2,

$n \approx 1.177\sqrt{m}$

m = 10,000, gives n = 117. Not that many!

END OF SLIDES

END OF DISCRETE PROBABILITY SLIDES FOR SECTIONS 6.1-6.2

Remaining topics in Probability Chapter Sections 6.3 – 6.4

- Some of these are covered in later slide sets by Selman
- Next slides indicate where some of the remaining topics are covered in Selman's slide sets Part (b) – Part (e)

Section 6.3: Bayes' Theorem

Topics covered in slides in Selman's Part (b) Slides

- Bayes' Theorem and the Generalized Bayes Theorem
- Bayesian Spam Filters

Section 6.4: Expected Value and Variance

Partially covered in slides for Part c and Part d

- Expected Values
- Linearity of Expectation
- Average-Case Computational Complexity
- The Geometric Distribution
- Independent Random Variables
- Variance
- Chebyshev's Inequality