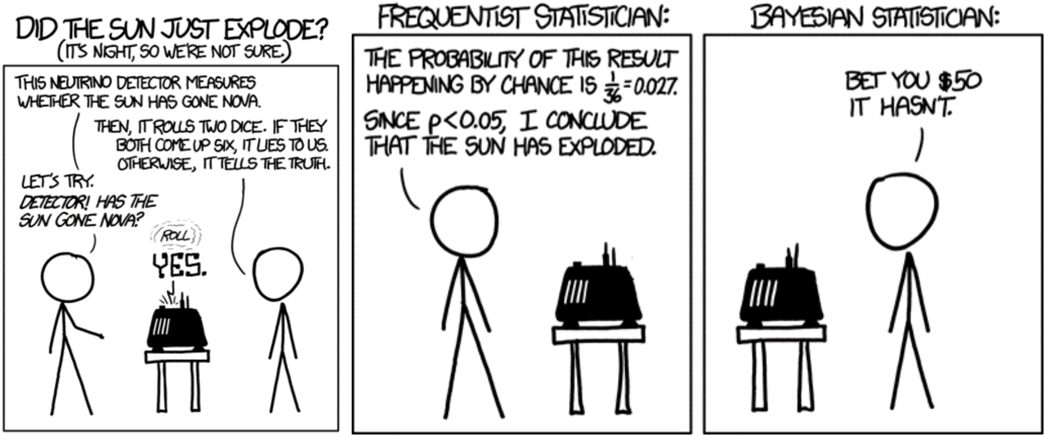
CSCI 3022 Intro to Data Science
Discrete pdfs



Announcements and To-Dos

Announcements:

1. HW 2 due Tuesday (not tonight, one extra day!)
   2 office hours: M 3-5 Tu 3-5.
2. Another nb day this Friday.

Last time we learned:

1. Bayes Theorem, finished up Probability theory

To do:

1. Finish that HW!

Boulder precip:

1) "Tr" → 0

2) "*" → ""

- string.replace
- Regex
- x[-1]

x = x[0:4].

library
string
re

3) omit rows

load the .txt, push backspace.

Probability Wrapup

OR Simulate the process of randomness, Count outcomes repeated experimentation.

- ~~If all outcomes are equally likely~~, we can just count outcomes:
  $P(A) = \frac{|A|}{|\Omega|} = \frac{\text{\# of ways A can happen}}{\text{\# of total outcomes in sample space}}$
- *Conditional Probability*: $P(A|B) = \frac{P(A \cap B)}{P(B)}$, P(both) / P(conditional).
- *Multiplication Rule*: $P(A \cap B) = P(A|B)P(B)$
- The following are equivalent: Two events A and B are said to be *independent*;
  $P(A|B) = P(A)$; $P(B|A) = P(B)$; $\boxed{P(A \cap B) = P(A)P(B).}$
- *Law of Total Probability*: Given disjoint $E_1, E_2, \ldots E_k$ such that $E_1 \cup E_2 \cup \cdots \cup E_k = \Omega$, for any A: A and E_1 A and E_2 A and E_k
  $P(\boxed{A}) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k)$
- *Bayes*: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$

  B happens as 1) $B \cap A$ 2) $B \cap A^c$

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

Random Variables

INPUT:

Definition: *Random Variable*
A *random variable* is a (measurable) function that maps elements or events in the sample space Ω to the real numbers $a_1, a_2, \ldots$ (or, more generally, to a measurable space...whatever that is!)

OUTPUT

Example: Consider rolling two dice. The *Sample Space* is the full list of outcomes $\{\omega_1, \omega_2\}$.
But what if we only care about summing the two dice? We could skip the sample space and just count the *random variable:*

$X :=$ the sum of the two dice.

Probability Distributions

Definition: *Probability Density Function*
A *Probability density function* (pdf) is a function f that describes the probability distribution of a random variable X.

If X is discrete, the pdf provides answers to questions like $P(X=2)=f(2)$. It is also called a probability mass function (pmf).

Capital X: variable
little x: one outcome

tiny

If X is continuous, then $P(X=x)$ =0 for all x. Why?! In this case, the distribution function is called a probability density function. In the continuous case, the pdf provides answers to questions like:

Probability Distributions

Definition: *Probability Density Function*
A *Probability density function* (pdf) is a function f that describes the probability distribution of a random variable X.

If X is discrete, the pdf provides answers to questions like $\underline{f(x) = P(X = x)}$. It is also called a probability mass function (pmf).

If X is continuous, then $\underline{P(X = x)=0}$ for all x. Why?! In this case, the distribution function is called a probability density function. In the continuous case, the pdf provides answers to questions like:
"What is the probability that X takes on a value between a and b?"

Properties of pdfs

For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:

for every outcome x:

1) Positive / non-negative

$$f(x) \geq 0$$

2. (For discrete distributions:)

2) Sum/add up to 1.

$$\sum_{\text{all } x} f(x) = 1$$

f is called a *probability mass function* because it describes how all of the possible outcomes in Ω have some probability or "mass" associated with them.

Properties of pdfs

For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:

1. $$f(x) = P(X = x) \geq 0 \qquad \forall x \text{ (with events in } \Omega)$$

2. (For discrete distributions:)

$$\sum_{x \in \Omega} f(x) = \sum_{x \in \Omega} P(X = x) = 1$$

f is called a *probability mass function* because it describes how all of the possible outcomes in Ω have some probability or "mass" associated with them.

Discrete pdfs

Example:
A lab has 6 computers. Let X denote the number of these computers that are in use during lunch hour, so

$$\Omega = \{0, 1, 2, \ldots, 6\}.$$

Suppose that the probability distribution of X is as given in the following table:

x	0	1	$\boxed{2}$	3	4	5	6
$P(X = x)$	.05,	.1	$\boxed{.15}$	.25	.2	.15	.1

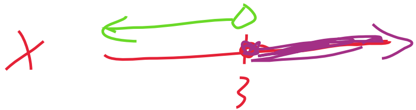
Sanity Checks!

1) $f(x) \geq 0$ for $x = 0, 1, \ldots 6$

2) $\sum f(x)$ $= .05 + 0.1 + 0.15 \ldots$

P(# of CPUs in use $= 2$) ?

Discrete pdfs



Example, cont'd:

x	0	1	2	3	4	5	6
$P(X=x)$	.05,	.1	.15	.25	.2	.15	.1

From here, we can find almost anything we might want to know about X.

1. Probability that at most 2 computers are in use

   P(0 or 1 or 2 CPUs in use)
   exclusive or!

   $P(0 \text{ in use}) + P(1 \text{ in use}) + P(2 \text{ in use}) = f(0)+f(1)+f(2) = .05+.1+.15$
   $= .3$

2. Probability that at least half of the computers are in use

   $P(X \geq 3) = P(x=3 \text{ or } 4 \text{ or } 5 \text{ or } 6) = P(x \neq 0 \text{ or } 1 \text{ or } 2) = P(\text{not } X<3)$

   $P(X \geq 3) = 1 - P(\underbrace{X<3}_{0 \text{ or } 1 \text{ or } 2}) = 1 - .3$
   $= .7$

3. Probability that there are 3 or 4 computers free

   $P(X=3 \text{ or } 4)$
   $= f(3)+f(4)$

Discrete pdfs

Example, cont'd:

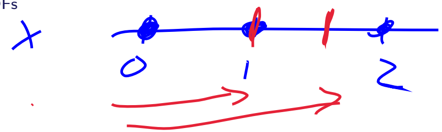
x	0	1	2	3	4	5	6
$P(X=x)$	.05,	.1	.15	.25	.2	.15	.1

From here, we can find almost anything we might want to know about X.

1. Probability that at most 2 computers are in use
   $P(X=0)+P(X=1)+P(X=2)=.3$

2. Probability that at least half of the computers are in use
   $P(X \geq 3)=1-P(X<3)=1-(P(X=0)+P(X=1)+P(X=2))=1-.3=.7$

3. Probability that there are 3 or 4 computers free
   $P(X \geq 3)=1-P(X=3\, or\, X=4)=1-(P(X=3)+P(X=4))=1-(.25+.2)=.55$



A Discrete pdf Example

Example: Suppose we are given the following pmf:

$$P(X=x) = f(x) = \begin{cases} .5 & x = 0 \\ .167 & x = 1 \\ .333 & x = 2 \\ 0 & \boxed{else} \end{cases}$$

Prob Outcomes

between

$P(X \leq 1.5) = P(X \leq 1)$

1. Calculate: $F(0), F(1), F(2)$.
2. What is $\underbrace{\boxed{F(1.5)}}_{=F(1)}$? $F(20.5)$?
3. Is $P(X < 1) = P(X \leq 1)$?

$F(0) = P(X \leq 0) = \underbrace{P(X=0)}_{.5} + \underbrace{P(X<0)}_{0} = .5$

$F(1) = P(X$ "up to 1"$) = P(X \leq 1) = .5 + .167$

$F(2) = P(X \leq 2) = P(X = 0 \text{ or } 1 \text{ or } 2) = 1.$

pdf Example; Soln

Example: Suppose we are given the following pmf:

$$P(X = x) = f(x) = \begin{cases} .5 & x = 0 \\ .167 & x = 1 \\ .333 & x = 2 \\ 0 & else \end{cases}$$

1. Calculate: $F(0), F(1), F(2)$.
2. What is $F(1.5)$? $F(20.5)$?
3. Is $P(X < 1) = P(X \leq 1)$?

pdf Example; Soln

Example: Suppose we are given the following pmf:

$$P(X = x) = f(x) = \begin{cases} .5 & x = 0 \\ .167 & x = 1 \\ .333 & x = 2 \\ 0 & else \end{cases}$$

everything else is not "supported"

1. Calculate: $F(0), F(1), F(2)$.
   $F(0) = P(X \leq 0) = .5$; $F(0) = P(X \leq 1) = .667$; $F(0) = P(X \leq 2) = 1$
2. What is $F(1.5)$? $F(20.5)$?
   $F(1.5) = P(X \leq 1.5) = P(X \leq 1) = .667$; $F(0) = P(X \leq 2) = 1$
   20.5
3. Is $P(X < 1) = P(X \leq 1)$?
   Most certainly not!

$P(X \leq 1) = P(X < 1) + P(X = 1)$

1/6!

Cumulative Distribution Functions

Definition: *Cumulative Density Function*
For a discrete r.v. X with pdf $f(x) = P(X = x)$, the *cumulative density function*, denoted $F(x)$, is defined for every real number x to be the probability that the observed value of X will be at most x.
Mathematically:

random variable

$$F(x) = P(X \leq x)$$

up to x.

Example: If I roll a single fair die, what is the cdf?

6 outcomes; $P(X=x) = f(x) = 1/6 \quad \forall x$.

1. $F(0)$: $P(X \leq 0)$ $= P(\text{die is 0 or less}) = 0$
2. $F(1)$: $P(X \leq 1) = P(\text{die is 1 or less}) = P(\text{die} = 1) = 1/6$.
3. $F(2)$: $P(X \leq 2) = P(\text{die is 2 or less}) = P(1 \text{ or } 2) = 1/6 + 1/6 = 1/3$
4. $F(6)$: $P(X \leq 6) = P(\text{"all outcomes up to 6 on fair die"}) = 1$.

Cumulative Distribution Functions

Definition: *Cumulative Density Function*
For a discrete r.v. X with pdf $f(x) = P(X = x)$, the *cumulative density function*, denoted $F(x)$, is defined for every real number x to be the probability that the observed value of X will be at most x.
Mathematically:

$$F(x) = P(X \leq x)$$

Example: If I roll a single fair die, what is the cdf?

1. $F(0) = 0$
2. $F(1) = 1/6$
3. $F(2) = 2/6$
4. $F(6) = 1$: with probability 1, our roll will be $\leq$ 6.

pdf to cdf

The relationship between pdf and cdf is very important!

$$F(a) = P(X \leq a) = \sum_{x \leq a} P(X = x)$$

every "x" up to "a" add up!

Example: What is the probability that if I roll two dice, they add up to at least 9. Write in terms of $F(x)$, then compute.

Die 1

Die 2

Sum

pdf to cdf

The relationship between pdf and cdf is very important!

$$F(a) = P(X \leq a) = \sum_{x \leq a} P(X = x)$$

Example: What is the probability that if I roll two dice, they add up to at least 9. Write in terms of $F(x)$, then compute.
$X :=$ the sum of the two dice, we want

Cumulative density function

$$P(X \geq 9) = 1 - \underline{P(X < 9)} = 1 - \underline{P(X \leq 8)} = 1 - F(8).$$

The easier probability is probably the
$P(X \geq 9) = \underbrace{P(X = 9)}_{4/36} + \underbrace{P(X = 10)}_{3/36} + \underbrace{P(X = 11)}_{2/36} + \underbrace{P(X = 12)}_{1/36} = 10/36.$

2d6; Ω and X

Suppose we roll two fair, 6-sided dice. Let $X :=$ the value representing the maximum of the two dice.

Process: Die 1 Die 2 => 36 outcomes? => MAX 6 outcomes.

1. What are the possible values of X?
2. Which elements of the sample space map to which values of X?
3. What is the pmf of X?

$\Omega = \{\text{Die 1}, \text{Die 2}\}$

$w = \{3, 2\} \Rightarrow X = 3$

2d6; Ω and X

Suppose we roll two fair, 6-sided dice. Let $X :=$ the value representing the maximum of the two dice.

1. What are the possible values of X?
2. Which elements of the sample space map to which values of X?
3. What is the pmf of X?

All equally likely! : count the ways.

1. $X \in \{1, 2, 3, 4, 5, 6\}$

2.

	Die 2					
Die 1	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

Red Die

Blue die

3. The pmf is: $P(X = x)$; or

$$f(x) = \begin{cases} 1/36 & X = 1 \\ 3/36 & X = 2 \\ 5/36 & X = 3 \\ 7/36 & X = 4 \\ 9/36 & X = 5 \\ 11/36 & X = 6 \end{cases}$$

2d6; The Max

Now we have

$$f(x) = \begin{cases} 1/36 & X = 1 \\ 3/36 & X = 2 \\ 5/36 & X = 3 \\ 7/36 & X = 4 \\ 9/36 & X = 5 \\ 11/36 & X = 6 \end{cases}$$

What are:

1. $P(X \text{ is even})$?
2. $P(X \text{ is 3 or less})$?
3. What is the cdf for X?

F(x) sums up rows of f(x)

2d6; The Max

Now we have

$$f(x) = \begin{cases} 1/36 & X = 1 \\ 3/36 & X = 2 \\ 5/36 & X = 3 \\ 7/36 & X = 4 \\ 9/36 & X = 5 \\ 11/36 & X = 6 \end{cases}$$

What are:

1. $P(X$ is even)?

$$\frac{3+5+7}{36}$$

2. $P(X$ is 3 or less)?

$$\frac{1+3+5}{36}$$

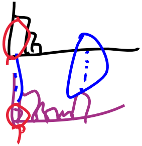
3. What is the cdf for X?

$$F(x) = \begin{cases} 0 & X < 1 \\ 1/36 & 1 \leq X < 2 \\ 4/36 & 2 \leq X < 3 \\ 9/36 & 3 \leq X < 4 \\ 16/36 & 4 \leq X < 5 \\ 25/36 & 5 \leq X < 6 \\ 36/36 & X > 6 \end{cases}$$

$$P(\text{surv} \mid \text{Pclass}=1) = \frac{P(\text{both})}{P(\text{Pclass}=1)} = \frac{\text{count}(\text{both})}{\text{count}(\text{class}:1)}$$

CDFs

hist:



hist.

$P(\text{young} \mid \text{surv.})$

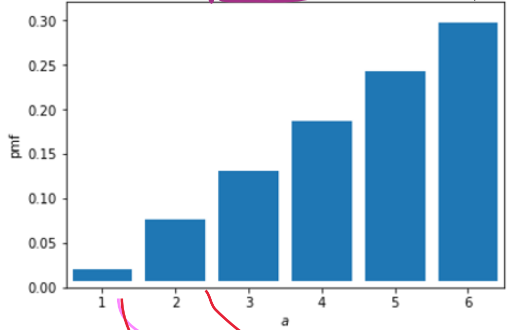
versus

$P(\text{young} \mid \text{died})$

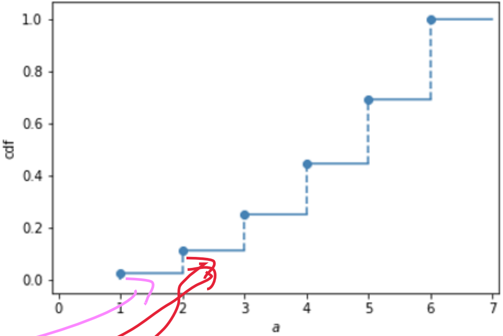
A picture denoting the pdf and cdf of our X:

$$P(A \mid B) = \boxed{\frac{P(\text{Both})}{P(B)}} = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

PDF:



CDF:



Making a pdf

Recall: we did an opening **example**: Suppose we flip a coin with a p chance per flip of landing on heads. Define $X =$ the number of tails flips before we see a heads. What is $P(X = 0)$? $P(X = 1)$? $P(X = i)$? Verify that $P(X) = 1$ over all of Ω.

- State space:
- Associated r.v. possible values or *support:*
- pdf $P(X = x) =$ probability of x tails before a heads:

Making a pdf

Recall: we did an opening **example**: Suppose we flip a coin with a p chance per flip of landing on heads. Define X = the number of tails flips before we see a heads. What is $P(X = 0)$? $P(X = 1)$? $P(X = i)$? Verify that $P(X) = 1$ over all of Ω.

- State space: $\{H, TH, TTH, TTTH, \dots\}$
- Associated r.v. possible values or *support:* $\{0, 1, 2, 3, \dots\}$
- pdf $P(X = x)$ = probability of x tails before a heads:

$$P(X = x) = P(\{T \dots TH\}) = P(\{T\})^x P(\{H\}) = (1 - p)^x \cdot p$$

So we report $f(x) = (1 - p)^x \cdot p$

Discrete Random Variables

Discrete random variables can be categorized into different types or classes.
Each type/class models many different real-world situations. They can loosely be broken down into a few groups:

1. The *Discrete Uniform* for modeling n equally likely (*fair*) outcomes
2. Distributions built on counting trials-until-event (how rolls until I get a 6, etc.) when the trials are identical and repeated

   Examples: Binomial, Geometric, etc.
3. Counting occurrences of an event over fixed areas of time/space.

   Example: Poisson

The Bernoulli

A random variable whose only possible values are 0 or 1.

This is a discrete random variable – why?

This distribution is specified by a single parameter:
The probability of a heads/"success" p! This gives the pdf:

We denote the Bernoulli random variable X by

The Bernoulli

A random variable whose only possible values are 0 or 1.

This is a discrete random variable – why?

Countable outcomes

This distribution is specified by a single parameter:

The probability of a heads/"success" p! This gives the pdf:

$$P(X = x) = f(x) = \begin{cases} p & x = 1 \\ (1 - p) & x = 0 \\ 0 & else \end{cases}$$

We denote the Bernoulli random variable X by $\underline{X \sim Bern(p)}$

The Bernoulli

A random variable whose only possible values are 0 or 1.

This is a discrete random variable – why?

This distribution is specified by a single parameter:
The probability of a heads/"success" p! This gives the pdf:

$$P(X = x) = f(x) = \begin{cases} p & x = 1 \\ (1-p) & x = 0 \\ 0 & else \end{cases}$$

It turns out, it's nice to write the pdf as a single line whenever possible. The nicest way to do so for the Bernoulli:

$$f(x) = p^x(1-p)^{1-x}$$

which works as long as we remember x can only be 0 or 1.
We denote the Bernoulli random variable X by $\underline{X \sim Bern(p)}$

Overview

The Bernoulli random variable is the building block for numerous important probability distributions that reflect **repeated** measurements.

Statistics and data science on repeated measurements requires us understand principles of **counting**!

Overview

The Bernoulli random variable is the building block for numerous important probability distributions that reflect **repeated** measurements.

Statistics and data science on repeated measurements requires us understand principles of **counting**!

1. Some counting is easy: how many integers are there in $[0, 9]$?

Overview

The Bernoulli random variable is the building block for numerous important probability distributions that reflect **repeated** measurements.

Statistics and data science on repeated measurements requires us understand principles of **counting**!

2. Zach, Felix, Rachel, and Ioana line up at a coffee stand. How many different orders could they stand in?

Overview

The Bernoulli random variable is the building block for numerous important probability distributions that reflect **repeated** measurements.

Statistics and data science on repeated measurements requires us understand principles of **counting**!

2. Zach, Felix, Rachel, and Ioana line up at a coffee stand. How many different orders could they stand in?
   This is a *permutation:* it counts distinct orderings

Overview

The Bernoulli random variable is the building block for numerous important probability distributions that reflect **repeated** measurements.

Statistics and data science on repeated measurements requires us understand principles of **counting**!

3. There are 10 problems on an exam, and you need 7 correct to pass. How many different ways are there to pass?

Overview

The Bernoulli random variable is the building block for numerous important probability distributions that reflect **repeated** measurements.

Statistics and data science on repeated measurements requires us understand principles of **counting**!

3. There are 10 problems on an exam, and you need 7 correct to pass. How many different ways are there to pass?
   This is a *combination*: it counts ways a set can be split into subsets

Permutations

How many ways can you order a set of one object; e.g. $\{A\}$?

How many ways can you order a set of two objects; e.g. $\{A, B\}$?

How many ways can you order a set of three objects; e.g. $\{ABC\}$?

What's the pattern? How many ways could you order n objects?

Permutations

How many ways can you order a set of one object; e.g. $\{A\}$?
A: 1 way. $\{A\}$.

How many ways can you order a set of two objects; e.g. $\{A, B\}$?
A: 2 ways. $\{AB, BA\}$.

How many ways can you order a set of three objects; e.g. $\{ABC\}$?
A: 6 ways. $\{ABC, ACB, BAC, BCA, CBA, CAB\}$.

What's the pattern? How many ways could you order n objects?
A: $n!$

Permutations; General

What if you have n objects, but only want to permute r of them?

How many 3-character strings can we make if each character is a distinct letter from the English alphabet?

What is the general form for an r-permutation of n objects?

Permutations; General

What if you have n objects, but only want to permute r of them?

How many 3-character strings can we make if each character is a distinct letter from the English alphabet?
A: There are 24 that start with $\{AB\}$. There are 25 letters (including B) that could have followed an A. There are 26 options to start with. That multiplies to $26 \cdot 25 \cdot 24$.

What is the general form for an r-permutation of n objects?
A: $P(n, r) = \frac{n!}{(n-r)!}$

This should feel a lot like **sampling without replacement**.. because it is, only without probabilities.

Combinations

Counting *combinations* means counting the number of ways an object can be sliced into subsets. The big difference: **order doesn't matter**.

> How many 3-character *combinations* can we make if each character is a distinct letter from the English alphabet?

Combinations

Counting *combinations* means counting the number of ways an object can be sliced into subsets. The big difference: **order doesn't matter**.

How many 3-character *combinations* can we make if each character is a distinct letter from the English alphabet?
Start with the number of permutations: $P(n, r) = 26 \cdot 25 \cdot 24$, then ask how many times we "overcounted," because now we don't want subsets with the same elements.
Ex: How many times did we include a subset with $\{A, B, C\}$?

Combinations

Counting *combinations* means counting the number of ways an object can be sliced into subsets. The big difference: **order doesn't matter**.

How many 3-character *combinations* can we make if each character is a distinct letter from the English alphabet?
Start with the number of permutations: $P(n, r) = 26 \cdot 25 \cdot 24$, then ask how many times we "overcounted," because now we don't want subsets with the same elements.
Ex: How many times did we include a subset with $\{A, B, C\}$?

Our permutation set had $\{ABC\}, \{ACB\}, \{BAC\}, \{BCA\}$,$\{CBA\}$, and $\{CAB\}$ as distinct... or all 6 orderings of those 3 elements! So:

$$C(n, r) = \frac{n!}{(n-r)!(r!)}$$

Combinations; Example

Combinations often use a variety of notations, including

$$C(n, r) = \binom{n}{k} = \frac{n!}{(n-r)!r!} := \text{“n choose k”}$$

Example: If there are 10 problems on an exam, and you need 7 correct to pass, how many different ways are there to pass?

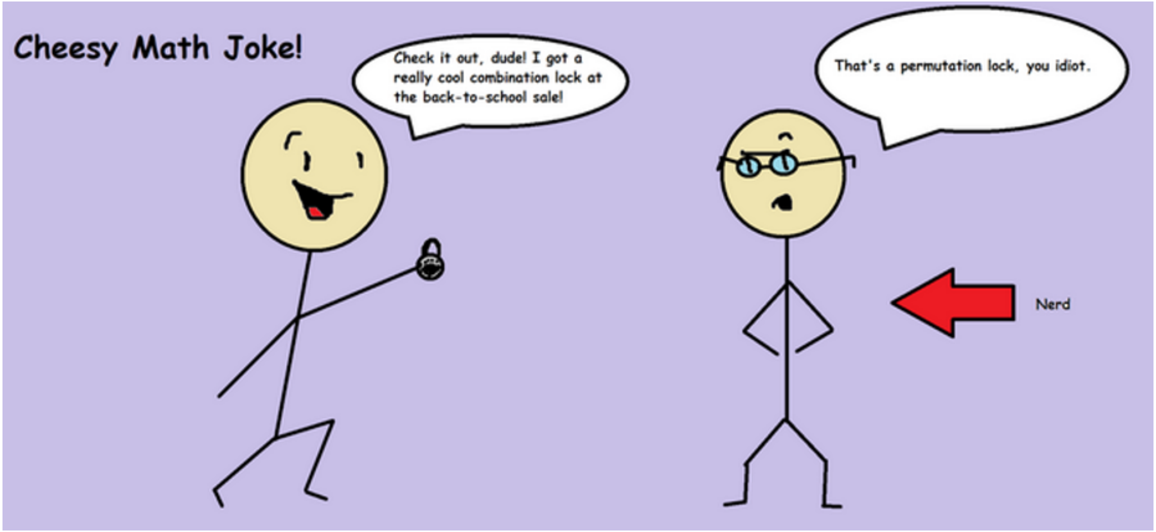
Combinations; Example

Combinations often use a variety of notations, including

$$C(n,r) = \binom{n}{k} = \frac{n!}{(n-r)!r!} := \text{“n choose k”}$$

Example: If there are 10 problems on an exam, and you need 7 correct to pass, how many different ways are there to pass?
Answer: $C(10,7) + C(10,8) + C(10,9) + C(10,10)$

Perms and Combs; Summary



Binomials

Exponents are useful, and pretty common: a lot of both data science and computational problems often involve solutions that look like polynomials. For example, let's consider:

1. Expand $(x + y)^1$
2. Expand $(x + y)^2$
3. Expand $(x + y)^3$
4. Expand $(x + y)^4$

Binomials

Exponents are useful, and pretty common: a lot of both data science and computational problems often involve solutions that look like polynomials. For example, let's consider:

1. Expand $(x+y)^1$
   Solution: $(x+y)^1 = x+y$
2. Expand $(x+y)^2$
   Solution: $(x+y)^2 = x^2+2xy+y^2$
3. Expand $(x+y)^3$
   Solution: $(x+y)^1 = (x+y)(x^2+2xy+y^2) = x^3+3x^2y+3xy^2+1$
4. Expand $(x+y)^4$
   Solution: $(x+y)^1 = (x+y)(x^3+3x^2y+3xy^2+1) = x^4+4x^3y+6x^2y^2+4xy^3+1$

Binomials

Exponents are useful, and pretty common: a lot of both data science and computational problems often involve solutions that look like polynomials. For example, let's consider:

1. Expand $(x+y)^1$
   Solution: $(x+y)^1 = x + y$
2. Expand $(x+y)^2$
   Solution: $(x+y)^2 = x^2 + 2xy + y^2$
3. Expand $(x+y)^3$
   Solution: $(x+y)^1 = (x+y)(x^2+2xy+y^2) = x^3 + 3x^2y + 3xy^2 + 1$
4. Expand $(x+y)^4$
   Solution: $(x+y)^1 = (x+y)(x^3+3x^2y+3xy^2+1) = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + 1$

What are some patterns? It's definitely symmetric - the coefficient are palindromic - and it seems to always start with 1 and then n (the power)

Binomials

One way to think about a binomial (two term) expansion is using "choose." Think about foiling:

$$(x_0 + x_1)(a + b) = \underbrace{ax_0}_{\text{first}} + \underbrace{bx_0}_{\text{outer}} + \underbrace{ax_1}_{\text{inner}} + \underbrace{bx_1}_{\text{last}}$$

There are 4 terms, but these are the same 4 terms as we would get from a multiplication rule: "choose" one of the first 2 terms and "choose" one of the second 2 terms for $2 \cdot 2$ total.

For our problem, we have to worry about repeating terms, though! If we think about:

$$(x + y)^4 = (x + y)(x + y)(x + y)(x + y)$$

it's making 4 choices: "choose x or y," then "choose x or y," then "choose x or y," then "choose x or y." The coefficient of the x^2y^2 term is the number of ways we could "choose x or y" 4 times and end up with 2 x's and 2 y's.

Binomials, Cont'd

So we're expanding

$$\begin{aligned}(x+y)^4 &= (x+y)\,(x+y)\,(x+y)\,(x+y)\\ &= (x+y)(x^3+3x^2y+3xy^2+1)\\ &= x^4+4x^3y+6x^2y^2+4xy^3+1\end{aligned}$$

and the coefficient of the x^2y^2 term is the number of ways we could "choose x or y" 4 times and end up with 2 x's and 2 y's.

Let's check. We're looking for all of the ways you could get e.g. $xxyy$, $yyxx$, $xyyx$, etc. This is the same as asking for the number of ways to choose 2 of the 4 "slots" to be x or choosing 2 of the 4 slots to be y, or $C(4,2) = \frac{4!}{2!}$.

Binomial Theorem

Theorem: Let x and y be variables and n be a non-negative integer. Then

$$(x+y)^n = \sum_{k=0}^{n} C(n,k)x^{n-k}y^k = C(n,0)x^n y^0 + C(n,1)x^{n-1}y^1 + \cdots + C(n,n)x^0 y^n$$

In other words, $C(n,k)$ is the coefficient of $x^k y^{n-k}$ *and* $x^{n-k}y^k$. We usually write the C numbers in choose notation:

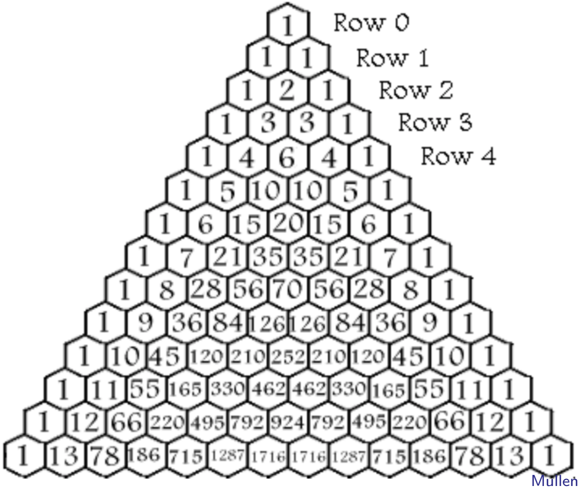
$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k}y^k = \binom{n}{0}x^n y^0 + \binom{n}{1}x^{n-1}y^1 + \cdots + \binom{n}{n}x^0 y^n$$

Pascal's Triangle

For small expansions, an easy trick to find the binomial coefficients is Pascal's triangle. Each entry of the triangle is the sum of the two entries above it:

Pascal's Triangle

For small expansions, an easy trick to find the binomial coefficients is Pascal's triangle. Each entry of the triangle is the sum of the two entries above it:



The Binomial

Example: A fair Bernoulli coin is tossed eight times. A "successful toss" is defined to be the coin landing on heads.
Let $X = \#$ of successes or heads in 8 tosses.

1. How many *ways* in Ω can $X = 3$?

2. What is $P(X = 3)$ for each *one* of those ways?

3. What is $P(X = 3)$?

The Binomial

Example: A fair Bernoulli coin is tossed eight times. A "successful toss" is defined to be the coin landing on heads.
Let $X = \#$ of successes or heads in 8 tosses.

1. How many *ways* in Ω can $X = 3$?

   $C(8,3)$ OR $C(8,5)$

2. What is $P(X = 3)$ for each *one* of those ways?

   One such way is $\{HHHTTTTT\}$ which has probability $P(\{H\})^3 \cdot P(\{T\})^5$.

3. What is $P(X = 3)$? The product of these two things!

The Binomial

Lets generalize those ideas to derive the Binomial pdf for n trials of an underlying Bern(p).

Let $X :=$ the number of successes of n trials of a Bern(p). Then:

NOTATION: We write ____________ to indicate that X is a Binomial rv with success probability p and n trials.

The Binomial

Lets generalize those ideas to derive the Binomial pdf for n trials of an underlying Bern(p).
Let $X :=$ the number of successes of n trials of a Bern(p). Then:

$$P(X = i) = (\text{\# of ways that } X = i) \cdot P(\text{of one such outcome})$$

NOTATION: We write $\underline{X \sim bin(n, p)}$ to indicate that X is a Binomial rv with success probability p and n trials.

The Binomial

Lets generalize those ideas to derive the Binomial pdf for n trials of an underlying Bern(p).
Let $X :=$ the number of successes of n trials of a Bern(p). Then:

$$P(X = i) = (\# \text{ of ways that } X = i) \cdot P(\text{of one such outcome})$$

$$P(X = i) = \binom{n}{i} \cdot P(n \text{ successes}) \cdot P(n - i \text{ failures}).$$

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{(n-i)}$$

$$f(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{(n-x)}; \quad x \in \{0, 1, 2, \ldots, n\}$$

NOTATION: We write $\underline{X \sim bin(n, p)}$ to indicate that X is a Binomial rv with success probability p and n trials.

The Binomial

The Binomial r.v. counts the total number of successes out of n trials, where X is the number of successes.

Important Assumptions:

1. Each trial must be *independent* of the previous experiment.
2. The probability of success must be *identical* for each trial.

The binomial is often defined and derived as the sum of n *independent, identically distributed* Bernoulli random variables.

In practice, any time we try to study a proportion on an underlying population, we gather a smaller sample where the observed proportion can often be thought of as a binomial random variable.

Daily Recap

Today we learned

1. pdfs and cdfs!

Moving forward:

- nb day Friday!
- Tuesday: HW 2

Next time in lecture:

- More common pdfs names!