Basic Principle of Counting and Permutations

Q: Suppose we flip a coin n times, what is the probability that we get heads exactly k times?

$n = 4$, $k = \underline{2}$

Total number of outcomes

HHHH	HTHH
HHHT	HTHT
HHTH	HTTH
HHTT	HTTT
THHH	TTHH
THHT	TTHT
THTH	TTTH
THTT	TTTT

16 outcomes

6 outcomes with exactly 2 heads

$$\text{probability} = \frac{6}{16}$$

Combinatorial analysis

= "the art of counting"

"Basic principle of counting"

- Suppose we perform 2 "experiments"
- The first experiment has m possible outcomes
- For each outcome of the first experiment, the second experiment has n possible outcomes

THEN, together there are $m \cdot n$ possible outcomes for the <u>pair</u> of experiments.

$$m\left\{\underbrace{\begin{array}{cccc} (1,1) & (1,2) & \cdots & (1,n) \\ (2,1) & (2,2) & & \\ (3,1) & (3,2) & \ddots & \\ \vdots & & & \\ (m,1) & & & (m,n) \end{array}}_{n}\right.$$

Total

$m \cdot n$

Extension:

r experiments

the ith experiment has n_i outcomes

Then all told there are

$$n_1 \cdot n_2 \cdot n_3 \cdot \cdots \cdot n_r$$

Total outcomes.

Ex Committee of 4 students, one from each class (Fresh, soph, junior, senior)

10 freshman
22 sophmores
13 juniors
2 seniors

$\Bigg\}$ how many committees?

A. $10 \cdot 22 \cdot 13 \cdot 2$

Ex Suppose a computer username consists of 3 letters followed by 5 numbers

How many usernames are possible?

$26^3 \cdot 10^5 = 26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10$

What if no repetitions are allowed?

$26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6$

Ex. How many batting lineups are possible on a team with 9 players.

$9! = 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$

1st batter   2nd batter   9th

<u>Permutations</u> = the number of ways of ordering a set of objects

(assume the objects are distinguishable)

Suppose given a set of n objects. There are $n!$ ways to order the objects.

$n \cdot (n-1) \cdot (n-2) \cdots (2) \cdot (1) = n!$

which object is first   which object is second

Permutations with indistinguishable objects, and combinations

HOMEWORK has been posted

p. 16-17: 1, 3, 5, 8, 9, 11, 13, 28

p. 18-19: 2, 3, 4, 8, 10

Permutation = a way to order a collection of objects

Example: A math contest with 20 participants

Q. How many rankings are possible? (assume no ties)

A. $20! = 20 \cdot 19 \cdot 18 \cdot 17 \cdots 3 \cdot 2 \cdot 1$

Example 10 books

different subjects
4 math
3 Chemistry
2 history
1 Language

We want all books on a particular subject are grouped together how many orderings are possible?

$24 = 4! = $ # orderings of math books

$6 = 3! = $ # orderings of chem books

$2 = 2! = $ # " history

$1 = 1! = $ # " language

$4! = $ # orderings of subjects

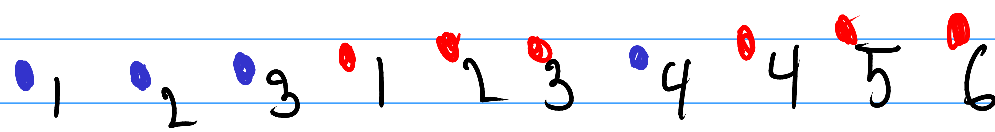
$$\text{Total} = 4! \cdot 4! \cdot 3! \cdot 2! \cdot 1!$$

Permutations with some indistinguishable objects

Ex 10 beads, 4 blue, 6 red are lined up in a row.

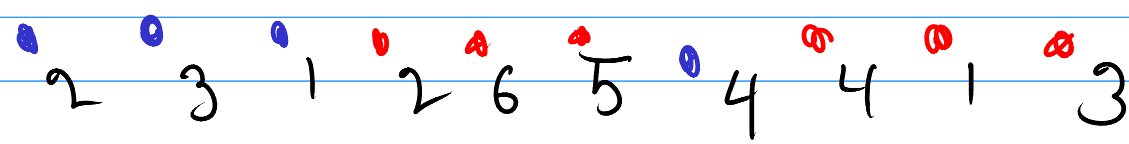
How many possible arrangements?

One idea: first solve the problem, supposing the beads are distinguishable



How many arrangements with subscript?

$10!$



Forget subscripts $\Rightarrow$ these two are the same.

How many times is this sequence of colors repeated?

permutations of blue subscripts = $4!$
permutations of red subscripts = $6!$

repeats = $4! \cdot 6!$

permutations with subscripts
= (# color sequences) · # repeats

$10! = (\text{\# color sequences}) \cdot 4! \cdot 6!$

color sequences = $\dfrac{10!}{4!\,6!}$

Example Anagrams of words with doubled letters

COMMITTEE How many anagrams?

$$\frac{9!}{2! \cdot 2! \cdot 2!}$$

n objects of r types

n_1 of type 1 all alike

n_2 of type 2 "

$\vdots$ $\vdots$

n_r of type r

$$\# \text{ arrangements} = \frac{n!}{n_1!\, n_2! \cdots n_r!}$$

Combinations

We have a collection of n objects, and we want to choose r of them

In how many ways can this choice be made?

Now, the objects are distinguishable, But we don't care about the order in which the objects are chosen.

"n choose 1" $= n$

"n choose 2" $= n \cdot (n-1)/2$

$\{1, 2, 3, 4, \ldots, n\}$

But $(1, 2)$ and $(2, 1)$ are the same collection of elements of this set

"n choose r" $= \dfrac{(\text{\# ordered subsets of size } r)}{(\text{\# permutations of a set of size } r)}$

$(\text{\# ordered subsets of size } r)$

$$= \underset{\text{1st}}{n} \cdot \underset{\text{2nd}}{(n-1)} \cdot \underset{\text{3rd}}{(n-2)} \cdots \underset{r\text{-th}}{\overset{(n-(r-1))}{(n-r+1)}}$$

$$= n \cdot (n-1) \cdot (n-2) \cdots (n-r+1) \cdot \frac{(n-r) \cdots (1)}{(n-r) \cdots (1)}$$

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

permutations of set of size r

$= r!$

"n choose r" $= \frac{n!}{(n-r)!} \cdot \frac{1}{r!}$

$$= \frac{n!}{(n-r)!\,r!} =: \binom{n}{r}$$ Binomial coefficient

Ex How many 5-card poker hands are possible?

"52 choose 5" $= \binom{52}{5} = \frac{52!}{47!\,5!}$

Ex 12 people divided into 3 committees of sizes 3, 4, and 5

How many ways?

$$\binom{12}{3} \cdot \binom{9}{4} \cdot \binom{5}{5}$$

1st comm.

$\binom{5}{5} = \frac{5!}{0!\,5!} = 1$

NOTE: $0! = 1$

$1! = 1$

$n \cdot (n-1)! = n!$

Prove $\binom{n}{r} = \binom{n}{n-r}$

Proof 1 use formula

$$\binom{n}{n-r} = \frac{n!}{\underbrace{(n-(n-r))}_{r}!\,(n-r)!}$$

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

Proof 2 combinatorial:

$\binom{n}{r}$ = # subsets of size r

$\binom{n}{n-r}$ = # subsets of size $n-r$

Given subset of size r, look at the complement (what's leftover). This is a set of size $n-r$.

Combinatorial Proof and the Binomial Theorem

Combinatorial Proof = Proof technique

count something in two different ways, get two different formulas, which must then be equal.

Recall $\binom{n}{r} = \frac{n!}{(n-r)!\,r!}$ = # subsets of size r in a set of total size n.

Theorem

$$\binom{n}{r} = \boxed{\binom{n-1}{r-1} + \binom{n-1}{r}} \quad (1 \leq r \leq n)$$

Proof

Right hand side

$$\frac{(n-1)!}{((n-1)-(r-1))!\,(r-1)!} + \frac{(n-1)!}{(n-1-r)!\,r!}$$

$$\frac{r}{r}\cdot\frac{(n-1)!}{(n-r)!\,(r-1)!} + \frac{(n-1)!}{(n-r-1)!\,r!}\cdot\frac{(n-r)}{(n-r)}$$

$$= \frac{r\cdot(n-1)!}{(n-r)!\,r!} + \frac{(n-1)!\cdot(n-r)}{(n-r)!\,r!}$$

using $r! = r\cdot(r-1)!$
$(n-r)! = (n-r)\cdot(n-r-1)!$

$$= \frac{r\,(n-1)! + (n-1)!\,(n-r)}{(n-r)!\,r!}$$

$$= \frac{n\,(n-1)!}{(n-r)!\,r!} = \frac{n!}{(n-r)!\,r!} = \binom{n}{r} \text{ QED.}$$

Combinatorial Proof is "conceptual"

Q: $S = \{1, 2, \ldots, n\}$

How many subsets of size r?

A1. $\binom{n}{r}$

A2. # subsets of size r

= # subsets of size r that contain 1

+ # subsets of size r that don't contain 1

subsets of size r containing 1

choose r−1 elements from $\underbrace{\{2,3,\ldots,n\}}_{n-1 \text{ of these}}$

$\binom{n-1}{r-1}$

subsets of size r not containing 1

choose r elements from $\{2,3,\ldots,n\}$

$\binom{n-1}{r}$

subsets of size $r = \binom{n-1}{r-1} + \binom{n-1}{r}$

Therefore $\underbrace{\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}}_{\text{Pascal's Triangle identity}}$

$$\binom{0}{0}$$

$$\binom{1}{0} \quad \binom{1}{1}$$

$$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$$

$$\vdots \qquad \vdots$$

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
$$

left-right symmetry $\Leftrightarrow \binom{n}{r} = \binom{n}{n-r}$

The Binomial Theorem is a bit of abstract algebra.

x, y variables $\quad x^5y^2$ - monomial

$x^4 + y^3$ - binomial

$x + x^2 + xy + y^2$ - polynomial

<u>Theorem</u>

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

$$(x+y)^3 = \binom{3}{0} x^0 y^3 + \binom{3}{1} x^1 y^2 + \binom{3}{2} x^2 y^1 + \binom{3}{3} x^3 y^0$$

$= 1y^3 + 3xy^2 + 3x^2y + 1x^3$

a row of pascal's triangle

2 Proofs

One proof uses induction

Another proof use combinatorics

Combinatorics Proof

$(x+y)^2 = (x+y)(x+y) = x\cdot x + x\cdot y + y\cdot x + y\cdot y$

$(x+y)^3 = (x+y)(x\cdot x + x\cdot y + y\cdot x + y\cdot y)$

8 terms

$= x\cdot x\cdot x + x\cdot x\cdot y + x\cdot y\cdot x + x\cdot y\cdot y$

$+ y\cdot x\cdot x + y\cdot x\cdot y + y\cdot y\cdot x + y\cdot y\cdot y$

$= x^3 + 3x^2y + 3xy^2 + y^3$

$$(x+y)^n = \underbrace{(x+y)(x+y)(x+y)\cdots(x+y)}_{n \text{ factors}}$$

Expand this out keep in track of the order in which the factors appear

To get a term can do, for example

Take x from the first factor
Take x from the second factor
Take y from the third factor
⋮
Take y from the nth factor

Gives a term $x \cdot x \cdot y \cdots \cdot y$

Total number of terms 2^n:

How many have k x's and $n-k$ y's

Each way of arranging k x's and $n-k$ y's will appear exactly once as a term

Permutations of n objects, k indistinguishable, $n-k$ indistinguishable

$$\frac{n!}{k!\,(n-k)!} = \binom{n}{k}$$ = # of times you get k x's and $n-k$ y's

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

starting chapter 2

HW 2 p. 18-19: 13, 14

chapter 2 problems 1, 2, 3, 4, 5

theoretical exercises 1, 2, 3, 6, 7, 9

Axioms of Probability

One attempt to define probability

The frequency with which an event occurs.

Take an experiment, run it n times, and let $n(E)$ be the number of times event E happens.

$$\text{"} P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \text{"}$$

This doesn't work as a definition, but we'll use a different definition, that will allow us to prove a very similar statement.

(The Strong Law of Large Numbers)

Sample Spaces and Event

Abstractly

Sample space — is a set, the set of possible outcomes of an experiment

Event — is a subset of the sample space

Probability — is a number associated to each event

Experiment	Sample space
Flipping a coin	$S = \{H, T\}$ (H = heads, T = tails) S is a set with two elements
Flipping a coin twice in a row	$S = \{(H,H), (H,T), (T,H), (T,T)\}$ 4 elements
outcome of a race with 7 competitors	$S = \{$ all $7!$ permutations of the set of competitors $\}$

Time it takes for a lab rat to finish a maze.

(set-builder notation)

$$S = \{t : 0 \le t < \infty\}$$

= set of all nonnegative numbers.

5 is in S but -1 is not

Events – a subset of the sample space

Subset: E is a subset of S, if every element of E is also an element of S.

"x is an element of S" $\Longleftrightarrow$ $x \in S$

means is an element of.

E is a subset of S

means (If $x \in E$ then $x \in S$.)

notation $E \subset S$

If we perform the experiment, get outcome $x \in S$.

If $x \in E$, then we say event E has occurred.

* Two coin flips $S = \{(H,H), (H,T)$
$(T,H), (T,T)\}$

event
$E = \{(H,H), (H,T)\}$

E is a subset of S, it represents the event of getting heads on the first flip.

* Race with 7 people $S = \{$all $7!$ perms$\}$

$E = \{$permutations starting with $5\}$

is the event that runner number 5 wins.

Q A die is rolled twice.

What is sample space?

$S = \{(1,1), (1,2), (1,3), \ldots$

$(2,1), (2,2), \ldots$

$\vdots$

$(6,1), (6,2), \ldots (6,6)\}$

all pairs of numbers between 1 and 6.

Event that the sum of rolls is 7?

$E = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

Event that the sum of the rolls is 12?

$E = \{(6,6)\}$

The notion of probability assigns a number $P(E)$ to each event subject to the following axioms

Axiom 1 $\quad 0 \leq P(E) \leq 1$

Axiom 2 $\quad P(S) = 1$

Axiom 3 $\quad$ If $E_1, E_2, \ldots$ is a sequence of events satisfying $E_i \cap E_j = \emptyset$ (for all $i \neq j$)

Then $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

Operations on sets

Fix a sample space S (we never mix sample spaces)

Let E and F be two events

($E \subset S$ and $F \subset S$)

<u>Union</u>: $E \cup F$ is the event consisting of outcomes in E or in F or in <u>Both</u>.

$x \in E \cup F \iff x \in E$ or $x \in F$ or both

$E \cup F$ occurs if E occurs or F occurs or both occur.

<u>Intersection</u>: EF (also written $E \cap F$) consists of outcomes both in E and in F

$x \in EF \iff x \in E$ <u>and</u> $x \in F$

EF occurs if E occurs and F occurs.

Two coin flips

$E = \{(H,H), (H,T)\}$ first flip heads

$F = \{(T,H), (H,H)\}$ second flip heads

$E \cup F = \{(H,H), (H,T), (T,H)\}$

first or second flip heads

$EF = \{(H,H)\}$ both flips heads

Two die rolls

E = sum of rolls is 6

F = sum of rolls is 12

$EF = \{\ \} = \emptyset$ the empty set

the empty event

Set operations, Axioms of Probability.

Recall from Last time:

S = sample space = set of possible outcomes

E an event = a subset of sample space

Union $E \cup F$ = "E or F or both"

Intersection EF = "E and F"

Empty set $\emptyset$ = set with no elements

it represents an event which is logically impossible

e.g. Flip coin get both Heads and Tails

$$\{H\} \cap \{T\} = \emptyset$$

Two events E and F are <u>mutually exclusive</u> if it is logically impossible for them to occur simultaneously

$$EF = \emptyset$$

Complement E^c = all outcomes in S that are not in E.

$S = \{H, T\}$

$E = \{H\}$ $\qquad$ $E^c = \{T\}$

<u>Notation</u>

$$E_1 \cup E_2 \cup E_3 \cup \cdots \cup E_n$$

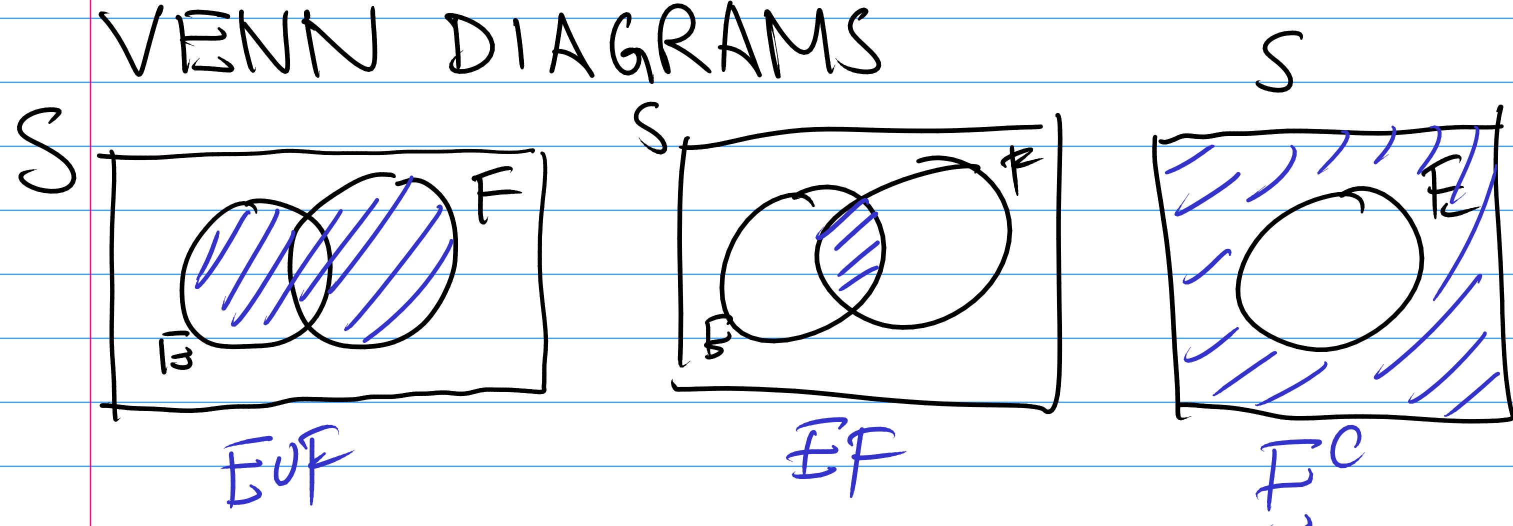
$$= \bigcup_{i=1}^{n} E_i$$

$$E_1 \cap E_2 \cap E_3 \cdots \cap E_n = \bigcap_{i=1}^{n} E_i$$

Also have $\quad \bigcup_{i=1}^{\infty} E_i \quad \bigcap_{i=1}^{\infty} E_i$

(Analogous to Σ notation $\quad \sum_{i=1}^{\infty} a_i$)

VENN DIAGRAMS



Laws for Set operations (Boolean Logic)

Commutative law $E \cup F = F \cup E \qquad EF = FE$

Associative $(E \cup F) \cup G = E \cup (F \cup G)$

$$(EF)G = E(FG)$$

Distribution law $(E \cup F)G = (EG) \cup (FG)$

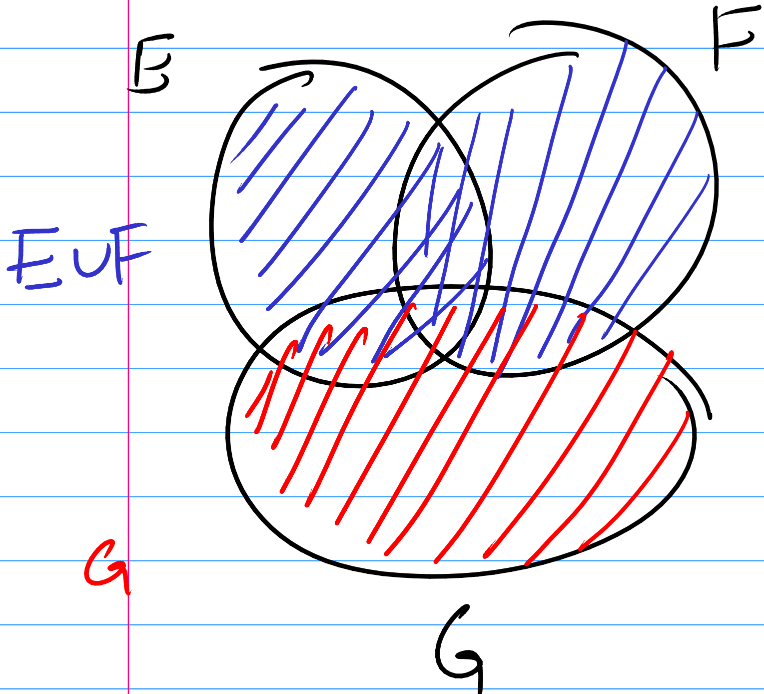
$$(EF) \cup G = (E \cup G)(F \cup G)$$

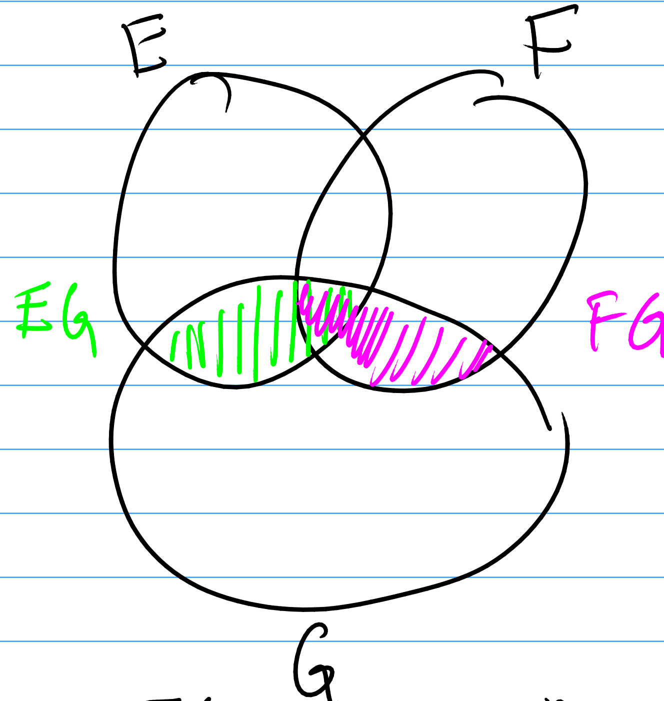
De Morgan's law

$$\begin{cases} (E \cup F)^c = E^c \cap F^c \\ \left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c \end{cases}$$

$$\begin{cases} (E \cap F)^c = E^c \cup F^c \\ \left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c \end{cases}$$

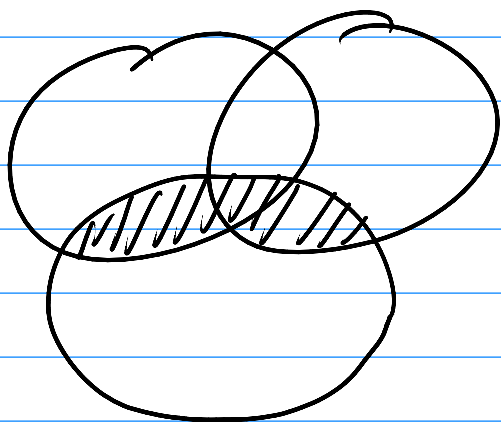
Prove $(E \cup F)G = (EG) \cup (FG)$



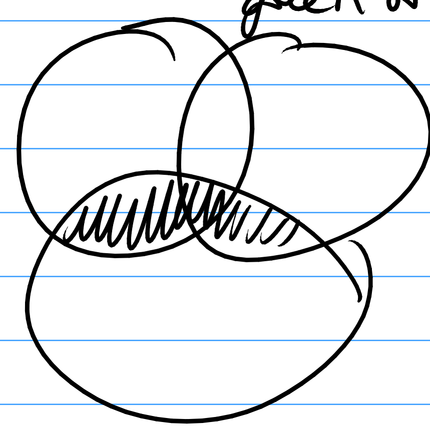


$(E \cup F)G$ = both red and blue

$EG \cup FG$ = either green or purple

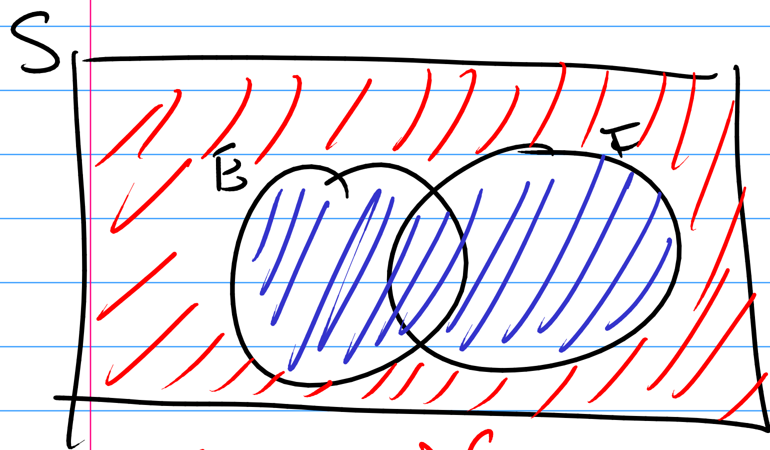


$=$

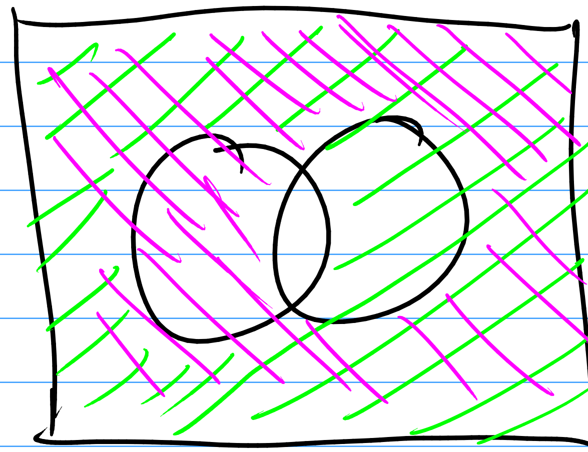


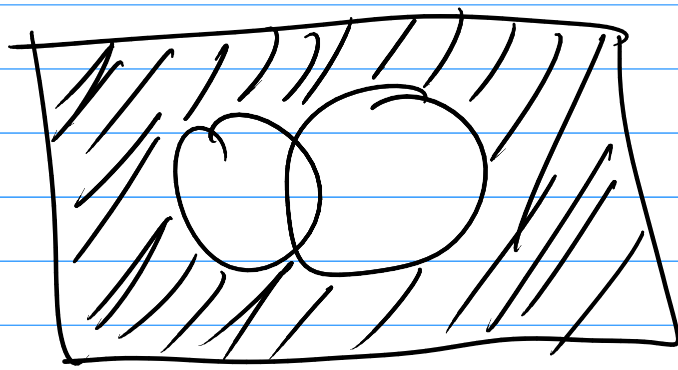
De Morgan's

$$(E \cup F)^c = E^c \cap F^c$$



$(E \cup F)^c$





Axioms of Probability

S = sample space

A measure of probability in S is an assignment of a number

$$P(E) = \text{"the probability of } E\text{"}$$

to each event E contained in the sample space.

<u>Axiom 1</u> $\quad 0 \leq P(E) \leq 1$

<u>Axiom 2</u> $\quad P(S) = 1$

<u>Axiom 3</u> $\quad$ If $E_1, E_2, E_3, \ldots$ is a sequence of events which are <u>mutually exclusive</u>

$$(E_i E_j = \emptyset \text{ unless } i = j)$$

Then
$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

"For mutually exclusive events, the prob of the union is the sum of the probabilities"

Remark 1 Axiom 3 applies to finite sequences of events

Remark 2 Axiom 3 asserts convergence of the infinite sum.

<u>Prop</u> $P(\emptyset) = 0$

Use axiom 3 $E_1 = S$ $E_2 = \emptyset$

$\ast$ $P(S \cup \emptyset) = P(S) + P(\emptyset)$

(Check $S, \emptyset$ mutually exclusive

$S\emptyset = \emptyset$ ✓)

$S \cup \emptyset = S$ (like adding nothing)

$$\underset{\substack{\text{Axiom 2} \\ \|\\ 1}}{P(S)} = \underset{\substack{\|\\ 1}}{P(S)} + P(\emptyset)$$

$1 = 1 + P(\emptyset)$

$P(\emptyset) = 0$ ∎

Axiom 2 forces S to be nonempty

Examples

$S = \{H, T\}$

$P(\{H\}) = \frac{1}{2}$

$P(\{T\}) = \frac{1}{2}$

$P(\emptyset) = 0 \qquad P(\{H,T\}) = 1$

Axioms 1, 2, 3 satisfied ✓

More generally: finite sample space

$S = \{x_1, x_2, \ldots, x_n\}$

Pick n numbers p_i, $i=1,\ldots,n$ such that $0 \le p_i \le 1$ and $\sum_{i=1}^{n} p_i = 1$

<u>Define</u>: $P(\{x_i\}) = p_i$

$\boxed{\{x_i\} \text{ simple event} = \text{event with exactly one element}}$

<u>Define</u> $P(E) = \sum_{i \text{ such that } x_i \in E} p_i$

(Forced by Axiom 3)

This satisfies Axioms 1, 2, 3 ✓.

Infinite sample space

$S = \{0, 1, 2, \ldots\}$ = non negative integers

$$P(\{i\}) = e^{-1} \frac{1}{i!} \qquad e = 2.71828\ldots$$

E event

$$P(E) = \sum_{i \in E} e^{-1} \frac{1}{i!}$$

Check Axioms 1, 2, 3.

Basic Properties, Inclusion-Exclusion

New OFFICE HOURS M 1-2, 3-4

W 9:30-10:30

S sample space

$E \subset S$ an event

$P(E)$ = probability of E

Axiom 1 $0 \leq P(E) \leq 1$

Axiom 2 $P(S) = 1$

Axiom 3 $E_1, E_2, \ldots, E_n, \ldots$ are mutually exclusive

$(E_i E_j = \emptyset$ unless $i = j)$

Then $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

Prop $P(\emptyset) = 0$ Did this last time.

Prop $P(E^c) = 1 - P(E)$

Proof:

$S = E \cup E^c$ Check Axiom 3 applies

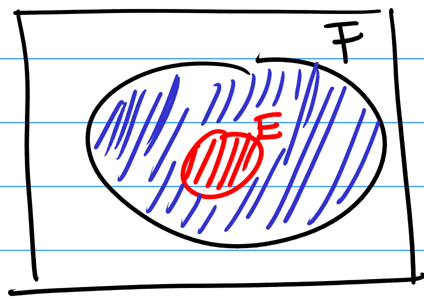
$EE^c = \emptyset$

$1 \overset{Ax2}{=} P(S) \overset{Ax3}{=} P(E) + P(E^c)$ QED.

Prop If $E \subset F$
(whenever E occurs, F also occurs)

THEN $P(E) \leq P(F)$

Proof:



$F = E \cup FE^c$

mutually exclusive

$(EFE^c = (EE^c)F = \emptyset F = \emptyset)$

Axiom 3 $\quad P(F) = P(E) + P(FE^c)$

By axiom 1 $\quad P(FE^c) \geq 0 \quad \Rightarrow \quad P(F) \geq P(E)$

QED

[E.g. rolling die $S = \{1, \ldots, 6\}$

$P(\{2\}) \leq P(\{\text{even}\}) = P(\{2,4,6\})$

Inclusion–Exclusion Identity

→ compute probability of a union of events, when the events are not assumed to be mutually exclusive.

$P(E \cup F)$

Rolling 6-sided die

$$P(\{\text{even}\}) = P(\{2,4,6\}) = \frac{1}{2}$$

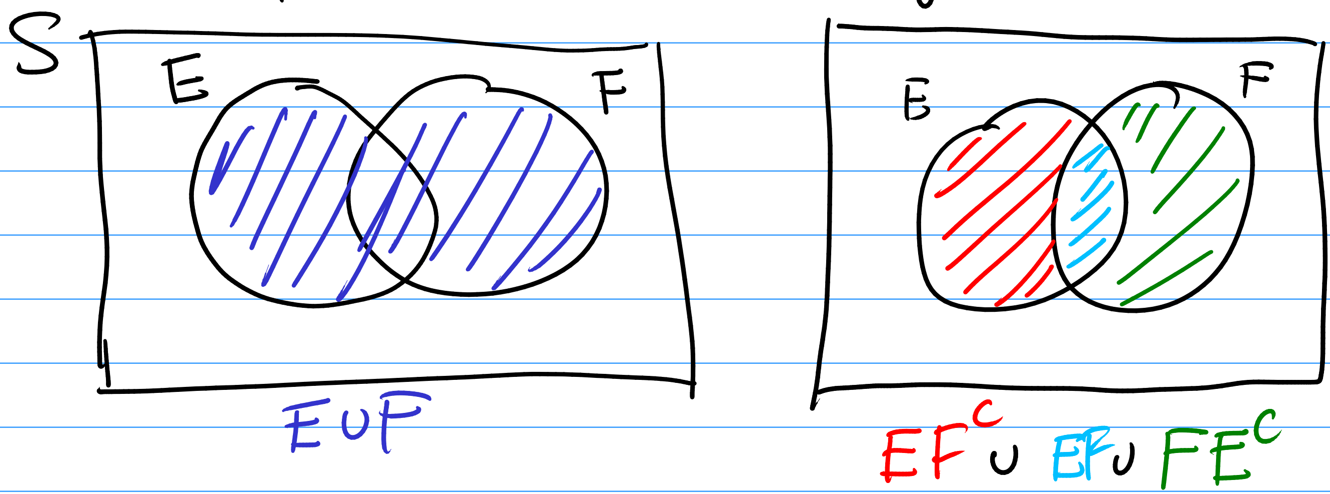
$$P(\{\text{prime}\}) = P(\{2,3,5\}) = \frac{1}{2}$$

$$P(\{\text{even}\} \cup \{\text{prime}\}) = P(\{2,3,4,5,6\}) = \frac{5}{6}$$

(inclusion-exclusion for 2 events)

Prop $P(E \cup F) = P(E) + P(F) - P(EF)$

Break up $E \cup F$ into mutually exclusive events



$$P(E \cup F) \overset{\text{Axiom 3}}{=} P(EF^c) + P(EF) + P(FE^c)$$

$$\begin{cases} P(E) = P(EF^c) + P(EF) \\ P(F) = P(EF) + P(FE^c) \end{cases}$$

$$P(E) + P(F) = P(EF^c) + \underbrace{2P(EF)}_{\text{intersection is overcounted}} + P(FE^c)$$

$$P(E) + P(F) - P(EF) = P(EF^c) + P(EF) + P(FE^c) = P(E \cup F)$$

Example: I'm trying to find gifts my mom. I find two.
I estimate:

$P(L_1) = P(\{\text{she likes first gift}\}) = 0.5$

$P(L_2) = P(\{\text{she like second gift}\}) = 0.4$

$P(L_1 L_2) = P(\{\text{she likes both}\}) = 0.3$

$P(\{\text{she like neither}\}) = P(L_1^c L_2^c)$

$$P(L_1 \cup L_2) = P(L_1) + P(L_2) - P(L_1 L_2)$$
$$= 0.6$$

$$P(L_1^c L_2^c) \underset{\substack{\uparrow \\ \text{De Morgan's law}}}{=} P((L_1 \cup L_2)^c) = 1 - P(L_1 \cup L_2) = 0.4$$

n-event inclusion exclusion:

$$P(E_1 \cup E_2 \cup \dots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2})$$

$\binom{n}{2}$ pairwise intersections

$$+ \sum_{i_1 < i_2 < i_3} P(E_{i_1} E_{i_2} E_{i_3})$$

$\binom{n}{3}$ triple intersections

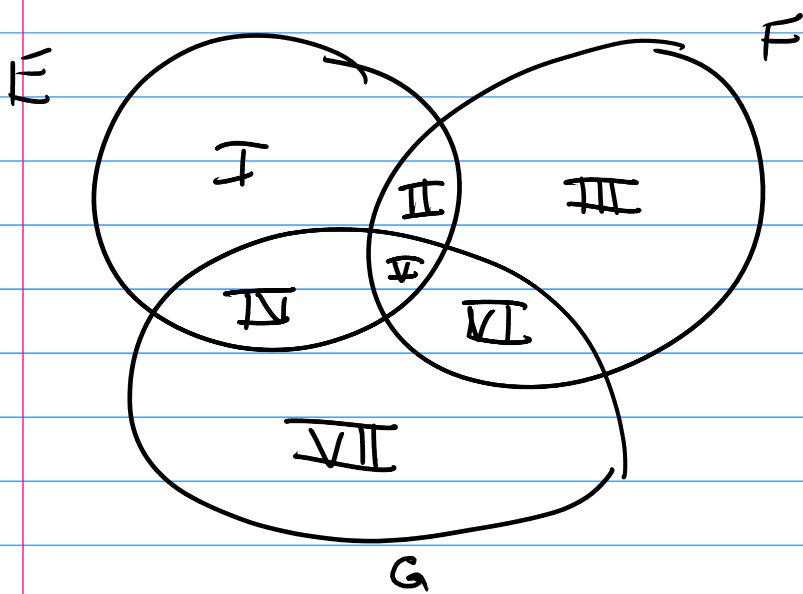
$$- \sum_{i_1 < i_2 < i_3 < i_4} P(E_{i_1} E_{i_2} E_{i_3} E_{i_4})$$

$\binom{n}{4}$ quadruple intersections

cont →

$+\cdots+(-1)^{n+1}P(E_1E_2\cdots E_n)$

3 sets $P(E\cup F\cup G)$



$$P(E)+P(F)+P(G) = P(\mathrm{I})+P(\mathrm{III})+P(\mathrm{VII})$$
$$+2P(\mathrm{II})+2(\mathrm{IV})+2P(\mathrm{VI})$$
$$+3P(\mathrm{V})$$

$$-P(EF)-P(EG)-P(FG) = -P(\mathrm{II})-P(\mathrm{IV})-P(\mathrm{VI})$$
$$-3P(\mathrm{V})$$

$$+P(EFG) = P(\mathrm{V})$$

Add it up $\Rightarrow$ get each region exactly once.

Equally Likely Outcomes.

Ch 2 problems : 8, 13, 18, 22, 25, 32, 39, 41, 54

Ch 2 theoretical : 10, 11, 18, 19

Ch 3 theoretical : 3.2

Today examples

Sample space $S = \{x_1, \dots, x_N\}$

$$P(\{x_1\}) = P(\{x_2\}) = \cdots = P(\{x_N\}) = \frac{1}{N}$$

In this situation, we say the sample space has equally likely outcomes.

For more general event $E \subset S$

$$P(E) = \frac{\#\text{outcomes in } E}{N} = \text{proportion of sample space in } E$$

So we can compute Probabilities just by counting.

Ex roll 2 dice. What is probability that the sum is 7?

$$S = \{ (1,1) \quad (1,2) \quad \cdots \quad (1,6)$$
$$(2,1) \quad (2,2) \quad \cdots \quad (2,6)$$
$$\vdots$$
$$(6,1) \quad \cdots \quad (6,6) \}$$

$N = 6 \cdot 6 = 36$

Assumption: each outcome is equally likely.

E = sum is $7 = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

6 outcomes in E.

$$P(E) = \frac{6}{36} = \frac{1}{6}$$

Ex Group of 11 people 6 men, 5 women

3 people "randomly" select of the 11.

$P(1 \text{ man}, 2 \text{ women})$

S = set of $\binom{11}{3}$ subsets of the 11 people of size 3.

By "randomly selected" we mean each subset is equally likely.

$$\underbrace{\binom{6}{1}}_{\text{men}}\underbrace{\binom{5}{2}}_{\text{women}} = \#\text{ of subsets with 1 man, 2 women}$$

$$P(1\text{ man}, 2\text{ women}) = \frac{\binom{6}{1}\binom{5}{2}}{\binom{11}{3}} = \frac{4}{11}$$

Or I could do: $S = \{\text{ordered selection of 3 people}\}$

$$N = \underbrace{11}_{\substack{\text{choices}\\ \text{for 1st}}} \cdot \underbrace{10}_{\text{2nd}} \cdot \underbrace{9}_{\text{3rd}} = 990$$

$\#\, M, W, W = 6 \cdot 5 \cdot 4 = 120$

$\#\, W, M, W = 5 \cdot 6 \cdot 4 = 120$

$\#\, W, W, M = 5 \cdot 4 \cdot 6 = 120$

$$P(1\text{ man}, 2\text{ women}) = \frac{120+120+120}{990} = \frac{4}{11}$$

ordered subsets equally likely $\Longleftrightarrow$ unordered subsets equally likely

$3! = 6$ ordered subsets $\longleftrightarrow$ 1 unordered subset

The Occupancy Problem: n cells, r particles/balls

$$\underbrace{\boxed{r_1}\boxed{r_2}\boxed{r_3}\boxed{\cdots}\boxed{r_n}}_{n \text{ cells}}$$

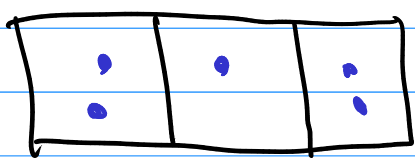
The particles are distributed "RANDOMLY", and we end up with some number in each cell.

r_i in the ith cell

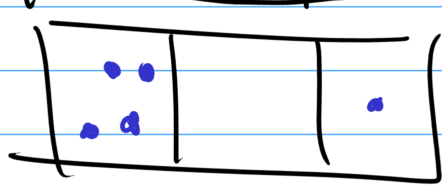
call $(r_1, r_2, \cdots, r_n)$ the "occupancy vector"

What is the probability of getting a particular occupancy vector?

$n = 3$ cells $\quad r = 5$ particles



$(r_1, r_2, r_3) = (2, 1, 2)$



$(r_1, r_2, r_3) = (4, 0, 1)$

Put each of r particles in one of n cells, randomly (each cell equally likely)

$\underbrace{n \cdot n \cdot n \cdots n}_{r} = n^r$ points in sample space

Assume each of these is equally likely $\frac{1}{n^r}$

How many ways to get $(r_1, r_2, \ldots, r_n)$?

$$\binom{r}{r_1}\binom{r-r_1}{r_2}\binom{r-r_1-r_2}{r_3}\cdots$$

$$= \frac{r!}{\cancel{(r-r_1)!}\, r_1!} \frac{\cancel{(r-r_1)!}}{(r-r_1-r_2)!\,(r_2)!} \cdots$$

$$= \frac{r!}{r_1!\, r_2! \cdots r_n!}$$

$$P((r_1, r_2, \ldots, r_n)) = \frac{r!}{r_1!\, r_2! \cdots r_n!} \frac{1}{n^r}$$

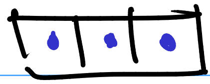
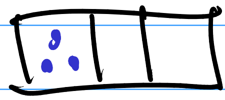
Particle which obey this probability law are said to have MAXWELL-BOLTZMANN statistics.

Note: actual elementary particles satisfy instead

BOSE-EINSTEIN or FERMI-DIRAC

BOSE-EINSTEIN	FERMI-DIRAC
Statistics (bosons)	statistics (Fermions)
Each occupancy vector is equally likely	each $r_i = 0$ or 1 (Pauli exclusion principle)
$\binom{r+n-1}{n-1}^{-1}$	$\binom{n}{r}^{-1}$

$n = r = 3$

	Maxwell Boltzmann	Bose Einstein
	$\frac{2}{9}$	$\frac{1}{10}$
	$\frac{1}{27}$	$\frac{1}{10}$

Example (Birthday Paradox)

If n people are in a room, what is probability that 2 have the same birthday

(relation to previous: person $\leftrightarrow$ particle, birthday $\leftrightarrow$ cell)

Assume each person has each birthday with equal probability. (365 days No leap years)

$(365)^n$ points in sample space

How many ways to have no two people with same b-day?

$$\underbrace{(365)}_{\text{1st person}}\underbrace{(364)}_{\text{2nd person}}\cdots\underbrace{(365-n+1)}_{n\text{ th person}}$$

$$P(\text{no two have same}) = \frac{(365)\cdots(365-n+1)}{(365)^n}$$

$P(\text{two have same}) = 1 - P(\text{no two have same b-day})$

if $n \geq 23$

$P(\text{two have same}) > \frac{1}{2}$

More examples with equally likely outcomes

Last time: Occupancy problem, Birthday Paradox

More about choice of sample space:

<u>Example</u> $n+m$ balls n red, m blue

Arrange the balls in row, randomly what is the probability of getting a particular sequence of colors?

$P(\bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet) = ?$ $\left(\begin{matrix} n=3 \\ m=4 \end{matrix}\right)$

$(n+m)!$ total orderings

groups of n indistinguishable
m indistinguishable objects

$\dfrac{(n+m)!}{n!\,m!}$ number of possible color sequence

If we regard the outcome of the experiment as being just the sequence of colors

Sample space has $\dfrac{(n+m)!}{n!\,m!}$ points

If each is equally likely,

$$P(\{\text{color sequence}\}) = \left[\frac{(n+m)!}{n!\,m!}\right]^{-1}$$

OR we could regard the experiment as taking one ball at a time and putting it in one of the $n+m$ possible positions.

$S = \{\text{orderings of } n+m \text{ balls}\}$
has $(n+m)!$ points

Assume each is equally likely $P(\{x\}) = \dfrac{1}{(n+m)!}$

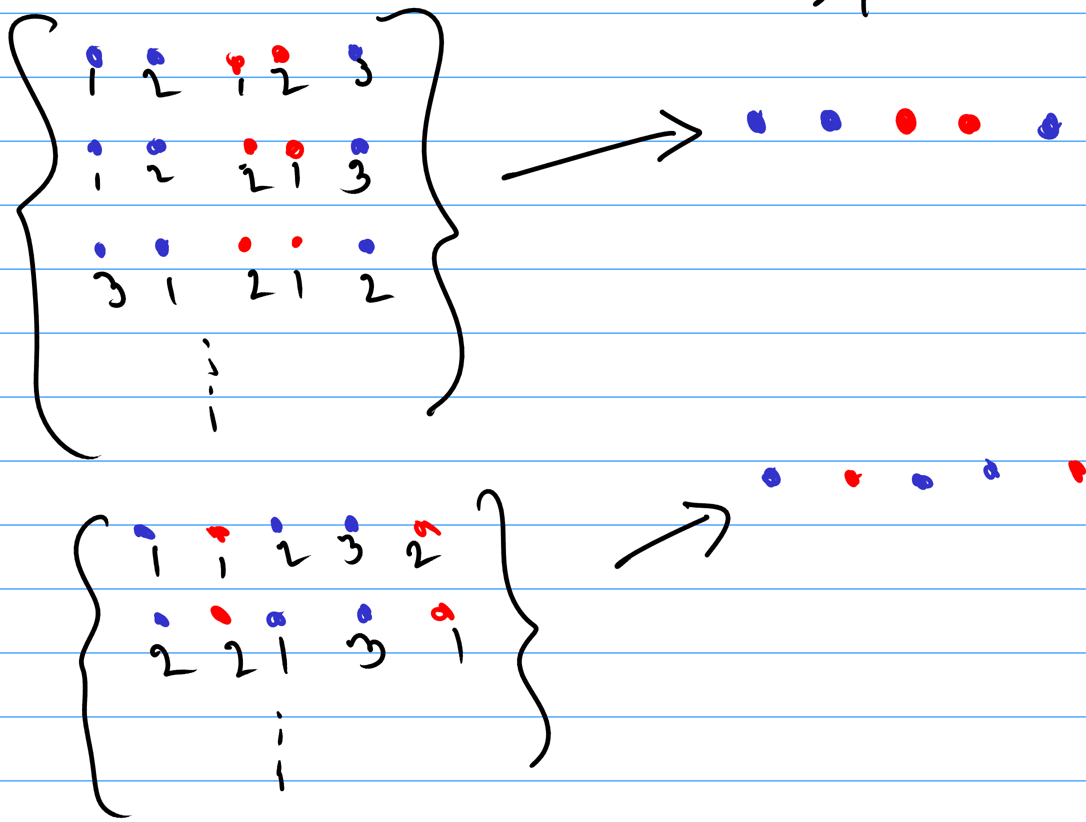
E = particular color sequence
consists of many points in sample space.

Given a particular color sequence, can order red balls in any way and blue balls in any way, to get a point in the sample space

points in $E = n!\,m!$

$$P(E) = \frac{n!\,m!}{(n+m)!} = \left[\frac{(n+m)!}{n!\,m!}\right]^{-1}$$

$S_{w/order}$ $S_{color\ sequences}$



points going to each outcome in $S_{color\ sequences}$

is always the same namely $n!m!$

Example of inclusion-exclusion

Sports club N members
three sports tennis squash badminton

36 play tennis 22 tennis & squash
28 squash 12 tennis & badminton
18 badminton 9 squash & badminton

4 all three

So how many members play some racket sport?

Consider experiment where we randomly select a member of the club.

$S = \{ \text{members of the club} \}$

$P(\{\text{each member}\}) = \frac{1}{N}$

$T =$ set of members plays tennis $\qquad P(T) = \frac{\#\text{ of members that play}}{N}$

$S =$ members plays squash

$B =$ " " badminton

At least one sport $\Leftrightarrow T \cup S \cup B$

$P(T \cup S \cup B) = \frac{\#\text{ play at least one sport}}{N}$

$$P(T \cup S \cup B) = P(T) + P(S) + P(B) - P(TS) - P(TB) - P(SB) + P(TSB)$$

$P(T) = \frac{36}{N}$ $\quad P(S) = \frac{28}{N}$ $\quad P(B) = \frac{18}{N}$

$P(TS) = \frac{22}{N}$ $\quad P(TB) = \frac{12}{N}$ $\quad P(SB) = \frac{9}{N}$

$P(TSB) = \frac{4}{N}$

$$\frac{36 + 28 + 18 - 22 - 12 - 9 + 4}{N} = \frac{\text{\# play at least one}}{N}$$

100 people go to a party put coats in a closet
when they leave, they grab a coat randomly
what is the probability that no person
selects their own coat?

$= 1 - P(\text{at least one person selects own coat})$

E_i = person i selects own coat

atleast one selects own $= \bigcup_{i=1}^{100} E_i$

$$P\left(\bigcup_{i=1}^{100} E_i\right) = \sum_{i=1}^{100} P(E_i) - \sum_{i<j} P(E_i E_j)$$

$$+ \sum_{i<j<k} P(E_i E_j E_k) - \text{4-way intersection} \cdots$$

$$P(E_i) = \frac{99!}{100!}$$ i gets own

$$P(E_i E_j) = \frac{98!}{100!}$$ i & j get own

$$P(E_{i_1} E_{i_2} \cdots E_{i_n}) = \frac{(100-n)!}{100!}$$ n people get own

$$P(E_{i_1} E_{i_2} \cdots E_{i_n}) = \frac{(100-n)!}{100!}$$ $\Big\}$ term like this appears $\binom{100}{n}$ times

$$\binom{100}{n} \frac{(100-n)!}{100!} = \frac{100!}{(100-n)!\,n!} \frac{(100-n)!}{100!} = \frac{1}{n!}$$

$$P(\cup E_i) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} \cdots - \frac{1}{100!}$$

$$1 - P(\cup E_i) = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} \cdots + \frac{1}{100!}$$

$$\approx 0.368$$

Compare $$\frac{1}{e} = \sum_{n=0}^{\infty} (-1)^n \frac{1}{n!}$$

(The matching problem)

Conditional Probability

Office hours M 1-2, 3-4 W 9:30-10:30

$P(E|F)$ = probability that E occurs given that $\underbrace{F}_{\text{condition}}$ occurs

Importance (i) compute probability w/ partial information

(ii) break up problems into conditional ones which may be easier.

(iii) Reasoning about hypotheses/evidence (Bayes's Formula)

(iv) Can define "independent events".

Suppose we are dealt 2 cards from a 52 card deck:

[Cards 13 values 2 3 4 ⋯ 10 J Q K A

4 suits Spades, hearts, diamonds, clubs]

eg. A spades

Suppose dealt 2 cards $P(2 \text{ aces})$?

$\binom{52}{2}$ = total # 2 card hands

= # points in S

$\binom{4}{2}$ = pairs of aces

$$P(2\text{ aces}) = \frac{\binom{4}{2}}{\binom{52}{2}} = \frac{\left(\frac{4\cdot 3}{2}\right)}{\left(\frac{52\cdot 51}{2}\right)} = \frac{4\cdot 3}{52\cdot 51}$$

Suppose get cards one at a time

$P(2\text{ aces} \mid \text{first card is an Ace})$

Prob. of getting 2 aces given that first card is an ace.

Thinking of drawing second card as a new experiment.

3 aces out of 51 cards left

$$P(2\text{ aces} \mid \text{first is Ace}) = \frac{3}{51}$$

$$P(1\text{st card is an Ace}) = \frac{4}{52}$$

See $P(2\text{ aces}) = P(\text{1st card is Ace}) \cdot P(\text{2 aces} \mid \text{1st card is A})$

$$\frac{4\cdot 3}{52\cdot 51} = \frac{4}{52} \cdot \frac{3}{51}$$

$$P(\text{2 aces} \mid \text{1st card is A}) = \frac{P(\text{2 aces})}{P(\text{1st card is Ace})}$$

Formalize

F = 1st card is A

E = 2nd card is A

"2aces" = EF

$$\textcircled{*} \quad P(E \mid F) = \frac{P(EF)}{P(F)}$$

We promote $\textcircled{*}$ to a definition

If E and F are events, we define $P(E|F)$ by $\textcircled{*}$

$P(E \mid F)$ = "Prob E given F"

"Prob E conditional on F"

Suppose x is an outcome. If F occurs then x is in F.

If we want E to also occur, we need x in E also

so ultimately x is in EF

That why we take $P(EF)$

$$1 = P(F|F) = \frac{P(FF)}{c} = \frac{P(F)}{c} \quad \text{so } c = P(F)$$

So $P(F)$ is a normalization.

Another interpretation: Given that F is known to occur: then we can replace the sample space S with the subset F.

(reduced sample space)

<u>Ex</u> Urn with r red and b blue balls

n Balls chosen in order w/o replacement

$(n \leq r+b)$

Suppose k of n chosen are blue

what is P(1st ball chosen is blue)

We work in reduced samplespace

B_k = event that k blue balls are chosen.

Each of the outcomes in B_k is equally likely (need to think)

Among the n balls chosen, the first is equally likely to be any of those n, and there are k chances for it to be blue so $P(\text{1st is blue} \mid B_k) = \frac{k}{n}$

Working w/ full sample space

B = first ball chosen is blue

B_k = k blue balls are chosen

$$P(B \mid B_k) = \frac{P(BB_k)}{P(B_k)}$$

$$P(BB_k) = P(B)\,P(B_k \mid B)$$

$$P(B \mid B_k) = \frac{P(B_k \mid B)\,P(B)}{P(B_k)}$$

Trick: reversing the order of the conditional probability

can do these parts combinatorially

$$P(B) = \frac{b}{r+b} \qquad P(B_k) = \frac{\binom{b}{k}\binom{r}{n-k}}{\binom{r+b}{n}}$$

$$P(B_k \mid B) = \frac{\binom{b-1}{k-1}\binom{r}{n-k}}{\binom{r+b-1}{n-1}}$$

$$P(B \mid B_k) = \frac{P(B_k \mid B) \cdot P(B)}{P(B_k)} = \frac{k}{n}$$

Bayes's formula

S sample space $\quad E$ and F events

"conditioning on F"

$$E = EF \cup EF^c$$

Proof
$$E = ES = E(F \cup F^c)$$
$$= EF \cup EF^c$$

And EF and EF^c are mutually exclusive

$$P(E) = P(EF) + P(EF^c)$$

$$P(E|F) = \frac{P(EF)}{P(F)}$$

$$= P(E|F)P(F) + P(E|F^c)P(F^c)$$

"$P(E)$ is weighted average of $P(E|F)$ and $P(E|F^c)$"

Recall trick "reversing the conditional probability"

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

Bayes's formula

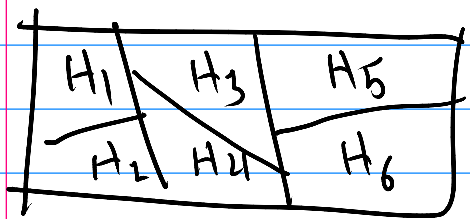
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)\,P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

More generally let $H_1, \ldots, H_n$ be a collection of mutually exclusive, exhaustive events

Mutually exclusive : $H_i H_j = \emptyset \quad (i \neq j)$

Exhaustive : $H_1 \cup H_2 \cup \cdots \cup H_n = S$

S



Consider H_i to be "hypotheses"

Now take any event E, which we regard as some "Evidence"

Bayes's formula
$$P(H_k \mid E) = \frac{P(E \mid H_k)\, P(H_k)}{\sum_{i=1}^{n} P(E \mid H_i)\, P(H_i)}$$

$P(E \mid H_i)$ - prob of evidence given hypothesis

$P(H_i)$ — "prior probability" of each hypothesis

Insurance company divides people into two class $\begin{cases} \text{accident prone} & A \\ \text{not accident prone} & A^c \end{cases}$

$P(B|A) = P($ prone person has accident in 1 year$) = .4$

$P(B|A^c) = P($ not accident prone has accident in 1 year$) = .2$

$P(A) = 30\%$ of population is accident prone

$S = \{$set of people$\}$ $A = \{$accident prone$\}$

$B = \{$people who have an accident in one year$\}$

Q: $P(B)$?

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$
$$= (.4)(.3) + (.2)(.7) = .26$$

Q: Suppose a person has an accident B what is probability that this person is accident prone?

$$P(A|B) = \frac{P(B|A)P(A)}{\underbrace{P(B|A)P(A) + P(B|A^c)P(A^c)}_{.26}}$$

$$= \frac{(.4)(.3)}{.26} = \frac{6}{13}$$

We have some prior probabilities

$C = \{\text{cheating}\}$ $\quad C^c = \{\text{not cheating}\}$

$E =$ way that they played

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

Under what conditions is $P(C|E) > P(C)$

Need
$$\frac{P(E|C)}{P(E|C)P(C) + P(E|C^c)P(C^c)} > 1$$

$$P(E|C) > P(E|C)P(C) + P(E|C^c)P(C^c)$$

$$P(E|C)\cancel{[1 - P(C)]} > P(E|C^c)\cancel{P(C^c)}$$

$$P(E|C) > P(E|C^c) \Rightarrow \text{increases confidence in } C$$

Example 3 cards

RR red on both sides
RB red / black
BB black on both sides

Card randomly selected and placed down on table, so we only get to see one side.

Suppose we see red facing up. What is prob that the other side is black?

Hypothesis = RR or RB or BB

Evidence = {see red} or {see black}

$$P(RB \mid \{\text{see red}\}) = \frac{P(\text{see red} \mid RB)\, P(RB)}{\left(P(\text{see red} \mid RB)\, P(RB) + P(\text{see red} \mid RR)\, P(RR) + P(\text{see red} \mid BB)\, P(BB) \right)}$$

$$= \frac{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)}{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{3}\right) + 0} = \frac{1}{3}$$

Independent Events

Events are independent if the occurance of one event doesn't affect the probability of the other event.

In general $P(E|F) = \dfrac{P(EF)}{P(F)} \neq P(E)$

But if equality holds

$P(E|F) = P(E)$ or equivalently $P(EF) = P(E)P(F)$

Then we say E and F are independent.

Examples 1 Multiple trials of given experiment

The outcomes of different trials are independent

Roll a die n times (n trials)

$E = \{\text{even \# on first roll}\}$ $F = \{\text{get odd \# on 10th roll}\}$

Example 2 Roll 2 dices

E_1 = sum is 6
E_2 = sum is 7
F = first die is 4

E_1 and E_2 are not independent (in fact mut. exclusive)

(1,5) (2,4) (3,3) (4,2) (5,1)

E_1 and F $\qquad P(E_1) = \frac{5}{36}$

$P(F) = \frac{1}{6}$ $\qquad P(E_1F) = \frac{1}{36}$

$\frac{5}{36} \cdot \frac{1}{6} \overset{?}{\neq} \frac{1}{36}$ NO not indep

E_2 and F $\qquad P(E_2) = \frac{6}{36} = \frac{1}{6}$

(1,6), (2,5) (3,4) (4,3) (5,2) (6,1)

$P(E_2F) = \frac{1}{36}$

$\frac{1}{6} \cdot \frac{1}{6} \overset{?}{=} \frac{1}{36}$ Yes these are independent.

Prop If E and F are independent so are E and F^c
also E^c and F
also E^c and F^c

$E = EF \cup EF^c$ $\qquad P(E) = P(EF) + P(EF^c)$

not exclusive

$P(E) = P(E)P(F) + P(EF^c)$

$[1 - P(F)]\,P(E) = P(EF^c)$

$P(F^c)\,P(E) = P(EF^c)$ $\square$

When are E, F, and G independent?

Roll 2 dice $\quad$ E = sum is 7

F = first die is 4

G = second die is 3.

$$\begin{tikzcd}
E \arrow[rr, leftrightarrow, "\text{indep}"] \arrow[dr, leftrightarrow, "\text{indep}"'] & & F \arrow[dl, leftrightarrow, "\text{indep.}"] \\
& G &
\end{tikzcd}$$

$$P(E \mid FG) = 1$$

$$\frac{P(EFG)}{P(FG)} = 1$$

$$P(EFG) = P(FG)$$

$$\neq P(E)P(FG)$$

Definition

3 events E F and G are independent (as a set of 3)

If $\quad P(EFG) = P(E)\,P(F)\,P(G)$

and $\quad P(EF) = P(E)P(F)$

$P(EG) = P(E)P(G)$

$P(FG) = P(F)P(G)$

n events $E_1, E_2, \ldots, E_n$ are independent

if for any subcollection of the events

$$E_{i_1}, E_{i_2}, \ldots, E_{i_r}$$

we have

$$P(E_{i_1} E_{i_2} \cdots E_{i_r}) = P(E_{i_1})\,P(E_{i_2}) \cdots P(E_{i_r})$$

i.e., the product rule holds for any subcollection of the events.

E.g. of E, F, G independent

E independent from FG, $F \cup G$

NB. Mutually exclusive $EF = \emptyset \Rightarrow E$ and F are not independent

If $E \subset F \Rightarrow E$ and F are not independent

<u>Main example</u> Bernoulli Trials

A Bernoulli trial is an experiment w/ 2 outcomes
outcomes = {success, failure}

$$P(\{\text{success}\}) = p \qquad 0 \leq p \leq 1$$

$$P(\{\text{failure}\}) = 1 - p$$

We do n trials. What is the probability of at least 1 success?

$$\{\text{at least 1 success}\}^c = \{\text{all failures}\} = F_1 F_2 \cdots F_n = \bigcap_{i=1}^{n} F_i$$

$$F_i = \{\text{failure on } i\text{th trial}\}$$

$$P(F_1 F_2 \cdots F_n) = P(F_1) P(F_2) \cdots P(F_n) = (1-p)^n$$

Because trials are independent.

$$P(\text{at least 1 success}) = 1 - P(\text{all fail}) = 1-(1-p)^n$$

Do n trials

Q: $P(\text{exactly } k \text{ successes})$

$P(\text{a particular sequence of } k \text{ successes and } n-k \text{ failures})$

(e.g. $P(SFSFSS) = p^4(1-p)^2$)

$$= p^k(1-p)^{n-k}$$

\# of sequence of k successe and $n-k$ $\quad \binom{n}{k}$

$$P(\text{exactly } k \text{ success}) = \binom{n}{k} p^k (1-p)^{n-k}$$

$n=4$

$$SFSS = F_1^c F_2 F_3^c F_4^c$$

these are independent

$$FFSS = F_1 F_2 F_3^c F_4^c$$

these are independent

these intersections are mutually exclusive

$P(\cdot \mid F)$ is a probability measure

Exam on Friday. Wednesday will be review.

Office hours M 1-2, 3-4, W 9:30-10:30

Let F be some fixed event

define $Q(E) = P(E \mid F)$

$Q(\cdot) = P(\cdot \mid F)$ satisfies all properties of a probability measure

<u>Prop</u> $0 \le P(E \mid F) \le 1$ $\qquad 0 \le Q(E) \le 1$

$P(S \mid F) = 1$ $\qquad Q(S) = 1$

if $E_1, E_2, \ldots$ are mutually exclusive

$$P\left(\bigcup_{i=1}^{\infty} E_i \,\middle|\, F\right) = \sum_{i=1}^{\infty} P(E_i \mid F)$$

$$Q\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} Q(E_i)$$

Anything that follows from axioms, has a "conditional" version

$P(E^c \mid F) = 1 - P(E \mid F)$

[Assume $P(F) \neq 0$]

$$P(E_1 \cup E_2 | F) = P(E_1|F) + P(E_2|F) - P(E_1E_2|F)$$

<u>Proof of Prep</u>

$$0 \leq P(E|F) \leq 1$$

$$P(E|F) = \frac{P(EF)}{P(F)} \geq 0$$

$$P(EF) \leq P(F) \quad b/c \quad EF \subset F$$

$$\frac{P(EF)}{P(F)} \leq 1$$

$$P(S|F) = \frac{P(SF)}{P(F)} = \frac{P(F)}{P(F)} = 1$$

$$P\left(\bigcup_{i=1}^{\infty} E_i \middle| F\right) = \frac{P\left(\left(\bigcup_{i=1}^{\infty} E_i\right) F\right)}{P(F)}$$

$$\left(\bigcup_{i=1}^{\infty} E_i\right) F = \bigcup_{i=1}^{\infty} (E_iF) \quad \text{distributive law}$$

$$E_iF \, E_jF \subseteq E_iE_j = \emptyset \quad b/c \ E_i \text{ mut. exc.}$$

$$\frac{P\left(\bigcup_{i=1}^{\infty} E_iF\right)}{P(F)} = \sum_{i=1}^{\infty} \frac{P(E_iF)}{P(F)} = \sum_{i=1}^{\infty} P(E_i|F)$$

Independent events E_1, E_2 $P(E_1E_2) = P(E_1)P(E_2)$

Conditionally independent events

$$P(E_1E_2 | F) = P(E_1 | F) P(E_2 | F)$$

Examples Insurance company

A = accident prone
A_1 = has accident in 1st year
A_2 = has accident in 2nd year

$P(A) = .3$ $P(A_i | A) = .4$
$P(A_i | A^c) = .2$

Assume A_1, A_2 conditionally independent

Previously we saw $P(A_i) = .26$

New Q: $P(A_2 | A_1)$ Accident in 2nd year given accident in first year.

Condition on A, whether person is accident.
Also, take all probabilities conditional on A_1

$$P(A_2|A_1) = P(A_2 | AA_1) P(A|A_1) + P(A_2 | A^cA_1) P(A^c | A_1)$$

$$P(A_2 | AA_1) \underset{\uparrow}{=} P(A_2 | A) = .4$$

because different years independent

$$P(A_2 \mid A^c A_1) = P(A_2 \mid A^c) = .2$$

$$P(A \mid A_1) = \frac{P(A_1 \mid A)\, P(A)}{P(A_1)} = \frac{(.4)(3)}{.26} = \frac{6}{13}$$

$$P(A^c \mid A_1) = 1 - P(A \mid A_1) = \frac{7}{13}$$

$$P(A_2 \mid A_1) = (.4)\frac{6}{13} + .2\,\frac{7}{13} \approx .29$$

Laplace's rule of succession:

$k+1$ coins in a box ith coin comes up heads with probability i/k. $(i = 0, \ldots, k)$

We flip coin n tries, and get n Heads what is prob. that $(n+1)$st flip is heads?

C_i = ith coin is selected. $P(C_i) = \frac{1}{k+1}$

F_n = first n flips are heads.

H = $(n+1)$st flip is heads.

$P(F_n \mid C_i) = (i/k)^n$ (Bernoulli trials with $p = i/k$)

$P(H \mid C_i) = i/k$

The different flips are conditionally independent, given C_i

$$P(H|F_n) = \sum_{i=0}^{k} P(H|F_n C_i)\, P(C_i|F_n)$$

$$P(H|F_n C_i) = P(H|C_i) = i/k$$

because H and F_n are independent given C_i

Bayes Formula

$$P(C_i|F_n) = \frac{P(F_n|C_i)\,P(C_i)}{P(F_n)}$$

$$= \frac{P(F_n|C_i)\,P(C_i)}{\sum_{j=0}^{k} P(F_n|C_j)\,P(C_j)} = \frac{(i/k)^n \frac{1}{k+1}}{\sum_{j=0}^{k} (j/k)^n \frac{1}{k+1}}$$

$$= \frac{(i/k)^n}{\sum_{j=0}^{k} (j/k)^n}$$

$$P(H|F_n) = \sum_{i=0}^{k} (i/k) \frac{(i/k)^n}{\sum_{j=0}^{k} (j/k)^n} = \frac{\sum_{i=0}^{k} (i/k)^{n+1}}{\sum_{j=0}^{k} (j/k)^n}$$

if k is large

$$P(H|F_n) \approx \frac{n+1}{n+2}$$

Review for Exam 1

HW: Ch 3 Theoretical
3.26, 3.28, 3.29 (large k limit)

Ch 4 problems:
4.1, 4.3, 4.5, 4.6

→ use the following approx

Laplace's rule of succession
$$\frac{\sum_{i=0}^{k}\left(\frac{i}{k}\right)^{n+1}}{\sum_{j=0}^{k}\left(\frac{j}{k}\right)^{n}} \approx \frac{n+1}{n+2}$$

$$\frac{1}{k}\sum_{i=0}^{k}\left(\frac{i}{k}\right)^{N} \xrightarrow[k\to\infty]{} \int_0^1 x^N\,dx = \frac{1}{N+1}$$

Ch 1 Combinatorics: Permutations / combinations and variations on such.

1: Suppose you need to answer 7 out of 10 questions
How many ways?
$$\binom{10}{7} = \frac{10!}{3!\,7!}$$

Need to take at least 3 out of first 5

$$\underset{\substack{\text{3 out}\\\text{of first}\\5}}{\binom{5}{3}}\underset{\substack{\text{4 out}\\\text{of}\\\text{second 6-10}}}{\binom{5}{4}} + \binom{5}{4}\binom{5}{3} + \binom{5}{5}\binom{5}{2}$$

2. Consider sequences of n digits, each digit 0-9

$$10^n = \underbrace{10 \cdot 10 \cdot 10 \cdots 10}_{n}$$

No 2 consecutive digits equal?

$$10 \cdot 9^{n-1} = \underset{\text{first}}{10} \cdot \underbrace{\underset{\text{second}}{9} \cdot \underset{\text{third}}{9} \cdots \cdot 9}_{n-1}$$

If 0 appears as a digit exactly i times: $i=0,\ldots,n$

$i=0$: 9^n

$i=1$: $\binom{n}{1} 9^{n-1}$

i general: $\binom{n}{i} 9^{n-i}$

3. Combinatorial Proof of

$$\binom{n}{r} = \binom{n}{n-r} \quad ?$$

Suppose have n people: split into two groups, one of size r, other of size $n-r$

$\binom{n}{r}$ ways choose people in first group

$\binom{n}{n-r}$ ways choose people in second group

Hence $\binom{n}{r} = \binom{n}{n-r}$

$\binom{n}{r}$ = # of subset of n of size r

4. Lottery 40 balls numbered 1-40 in an urn, draw unordered set of 8.

$\binom{40}{8}$ possible outcomes: equally likely.

Ticket w/ 8 numbers

$$P(\text{winning}) = 1 \Big/ \binom{40}{8}$$

smaller prize if you match 7

P(match exactly 7)

E_1 = match first 7   # outcomes in $E_1 = 33$

E_2 = match first 7 but not eighth $\# E_2 = 32$

$P(E_2) = \dfrac{32}{\binom{40}{8}}$

$P(\text{match exactly } 7) = \binom{8}{7} 32 \Big/ \binom{40}{8}$

$\binom{8}{7} = 8$

5. E, F, G $P(E) = .22$ $P(EF) = .11$

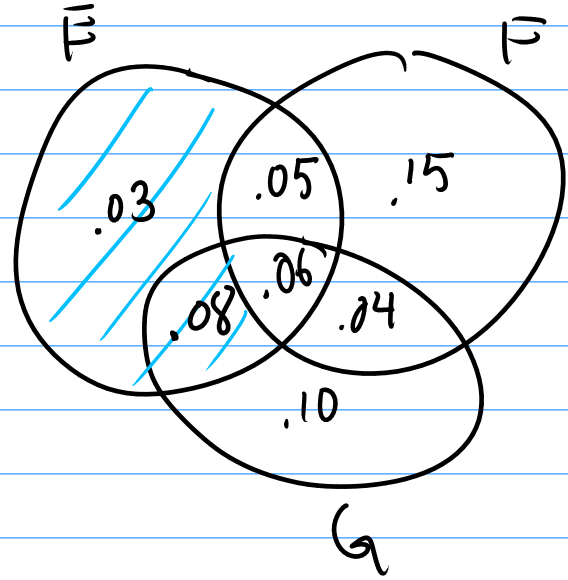
$P(F) = .30$ $P(EG) = .14$ $P(EFG) = .06$

$P(G) = .28$ $P(FG) = .10$

$P(EF^c)$?

$= P(E) - P(EF)$

$= .22 - .11 = .11$



6. 2 cards are drawn from 52

$B = \{\text{both cards are aces}\}$ $\quad P(B) = \dfrac{\binom{4}{2}}{\binom{52}{2}}$

suppose we know First card is the ace of spades

$P(B \mid \text{first card is A of Spades}) = \dfrac{3}{51}$

$= \dfrac{P(B \cap \{\text{first is A of Spades}\})}{P(\{\text{first is A of spades}\})} = \dfrac{3/52\cdot 51}{1/52} = \dfrac{3}{51}$

7. You ask neighbor to water plant

$P(\text{neighbor waters}) = .9$

$P(\text{plant dies} \mid \text{watered}) = .15$

$P(\text{plant die} \mid \text{unwater}) = .8$

<u>Q1</u>: $P(\text{Plant dies}) = P(\text{die} \mid \text{water})\,P(\text{water}) + P(\text{die} \mid \text{nowater})\,P(\text{nowater})$

$= (.15)(.9) + (.8)(.1)$

Q2: Given Plant dies Prob. that it was watered

$$P(\text{water} \mid \text{die}) = \frac{P(\text{die} \mid \text{water})\, P(\text{water})}{P(\text{die})}$$

$$= \frac{P(\text{die} \mid \text{water})\, P(\text{water})}{P(\text{die} \mid \text{water})\, P(\text{water}) + P(\text{die} \mid \text{no water})\, P(\text{no water})}$$

Random Variables

Exam Stats in terms of Raw Scores

1st Q	48
Median	67
3rd Q	86
Max	100

$$C = \frac{1}{2}(R - 100) + 100$$

Range is compressed to 50–100

Random Variable ← This idea for rest of course.

Interested in a <u>function</u> of the outcome of some probabilistic process.

Eg.
1) Roll two dice X = sum of dice
2) Flip 3 coins X = # of heads
3) Give a student a midterm X = score
4) Stock market index

(random variable)

Can assign probabilities to values of a R.V.

Flip 3 fair coins $Y = \#$ of heads

Y	outcomes
0	TTT
1	HTT, THT, TTH
2	HHT, HTH, THH
3	HHH

Event $\{Y=0\} = \{TTT\}$

$\{Y=1\} = \{HTT, THT, TTH\}$

$\{Y=2\} = \ldots$

$\{Y=3\} = \ldots$

$P(Y=0) = P(\{TTT\}) = 1/8$

$P(Y=1) = P(\{HTT, THT, TTH\}) = 3/8$

$P(Y=2) = 3/8$

$P(Y=3) = 1/8$

$$\sum_{i=0}^{3} P(Y=i) = P\left(\bigcup_{i=0}^{3} \{Y=i\}\right) = P(S) = 1$$

because $\{Y=i\}$ are mutually exclusive.

In other words, can define events by conditions on the value of a random variable.

$\boxed{Y \text{ a R.V.}}$

$$\{Y=0\} \cup \{Y>0\} = \{Y \geq 0\}$$

$$\{Y \geq 0\} \cap \{Y \leq 0\} = \{Y=0\}$$

$$\{Y \geq 0\} \cap \{Y<0\} = \emptyset$$

$$\{a \leq Y \leq b\}$$

Ex Urn contains 11 balls

3 white, 3 red, 5 black

+\$1   -\$1   0

We draw 3   X = amount we win or lose

All Possible values of X = -3, -2, -1, 0, 1, 2, 3

(An example of a $\boxed{\text{discrete}}$ R.V.)

Because ONLY FINITELY MANY POSSIBLE VALUES.

$$P(X=0) = \Bigg[\underbrace{\binom{5}{3}}_{\text{All black}} + \underbrace{\binom{3}{1}}_{\text{white}}\underbrace{\binom{3}{1}}_{\text{red}}\underbrace{\binom{5}{1}}_{\text{black}}\Bigg]\binom{11}{3}^{-1}$$

$\binom{11}{3}$ TOTAL OUTCOMES

$P(\text{we win money}) = P(X > 0)$

$= P(X = 1, 2, \text{or } 3) = P(X=1) + P(X=2) + P(X=3)$

Discrete Random Variables

HW 6: Ch 4 Problems: 4.13, 4.18, 4.19, 4.20, 4.28

Theoretical: 4.2, 4.3, 4.4, 4.7, 4.8

Random Variable = function on the sample space.

$P(X = a)$

A Discrete Random Variable is one which can take on at most countably many values

<u>Ex</u> If X has finitely many possible values, then X is <u>discrete</u>

<u>Ex</u> If X has every integer as a possible value then X is discrete

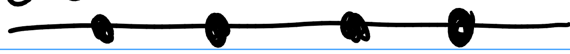
<u>Not discrete</u> X = amount of time it takes finish a race

possible values: $\{t : 0 \leq t < \infty\} = [0, \infty)$

all real #'s in this interval

Then X is a continuous random variable

discrete:



X

continuous

X

Thing about discrete R.V.: In general there are gaps between the possible values, and the possible values can be indexed by integers

possible values of $X = \{x_1, x_2, x_3, \dots\}$

PROBABILITY MASS FUNCTION of a discrete random variable X:

Let a be a possible value of X

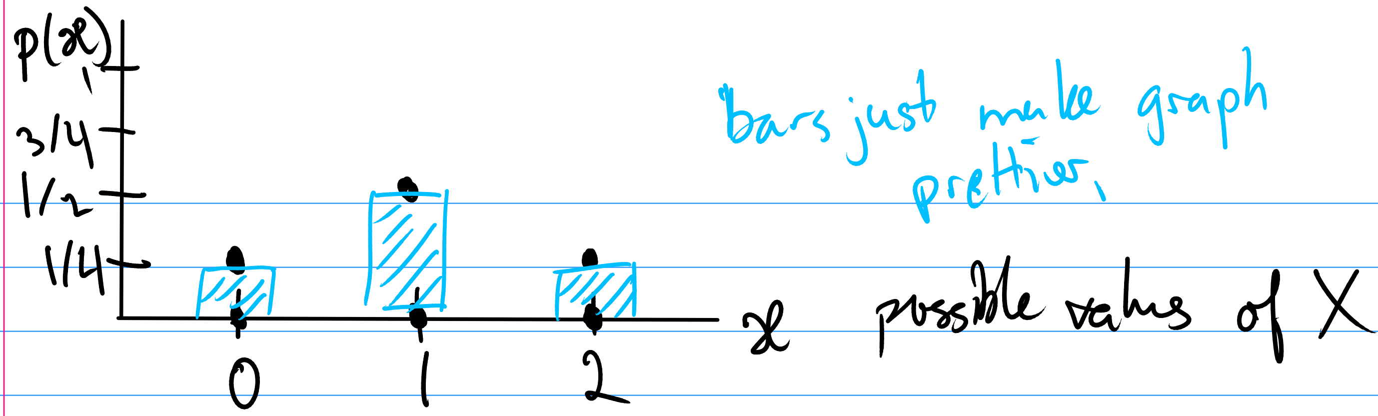
define $p(a) = P(X = a)$

Ex flip two coins $X = \#$ of heads

$S = \{HH, HT, TH, TT\}$
$X = \quad 2 \quad 1 \quad 1 \quad 0$

possible values 0 1 2

$P(X=0) = \frac{1}{4} \qquad P(X=1) = \frac{1}{2} \qquad P(X=2) = \frac{1}{4}$



bars just make graph prettier,

and $p(x)$ is zero at all other points

Basic properties of P.M.F.

possible values of $X = \{x_1, x_2, x_3, \dots\}$

$1 \geq p(x_i) \geq 0$ (by Axiom 1)

$P(x) = 0$ if x is not possible value

$$\sum_{i=1}^{\infty} p(x_i) = 1 \quad \text{(by Axiom 3)}$$

$$\sum_{i=1}^{\infty} p(x_i) = \sum_{i=1}^{\infty} P(X = x_i) = P\left(\bigcup_{i=1}^{\infty} \{X = x_i\}\right)$$

$$= P(S) = 1$$

Everything having to do with discrete R.V. is expressed it terms of (finite or infinite) sum

For continuous R.Vs, these sums are replaced by 'integrals'.

For many questions about R.V.s, we don't care about sample space or the interpretation in terms of some experiment, we just care about the probability mass function of the R.V.

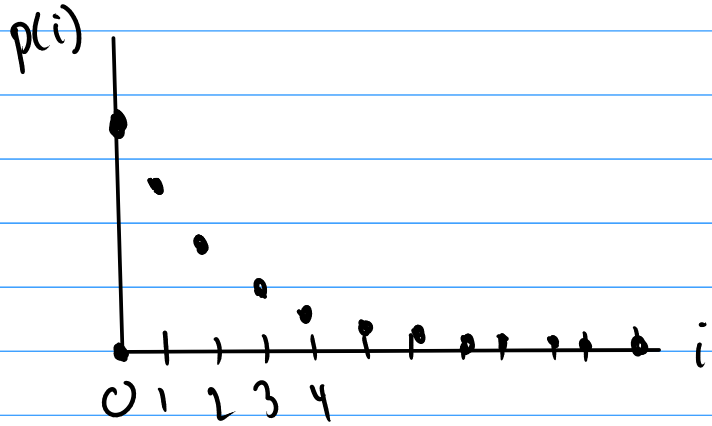
<u>Example</u> Some Random Variable X has

P.M.F $\quad p(i) = c\lambda^i / i!$

Possible values $i = 0, 1, 2, \ldots$

λ is a given constant

c is an (as yet undetermined) normalization constant.



<u>Q</u> Relationship btwn c & λ

Need

$$\sum_{i=0}^{\infty} p(i) = 1$$

$$1 = \sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{\infty} c\frac{\lambda^i}{i!} = c\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = ce^{\lambda}$$

$$e^x = \sum_{i=0}^{\infty}\frac{x^i}{i!}$$ Taylor Series for the exponential function.

In other words $c = \frac{1}{e^{\lambda}} = e^{-\lambda}$

$$p(i) = e^{-\lambda} \lambda^i / i!$$

Compute probabilities using P.M.F.

$$P(X=0) = p(0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda}$$

$$P(X=1) = p(1) = e^{-\lambda} \frac{\lambda^1}{1!} = \lambda e^{-\lambda}$$

$$P(X=2) = p(2) = e^{-\lambda} \frac{\lambda^2}{2!} = \lambda^2 e^{-\lambda}/2$$

$$P(X>2) = 1 - P(X=0) - P(X=1) - P(X=2)$$

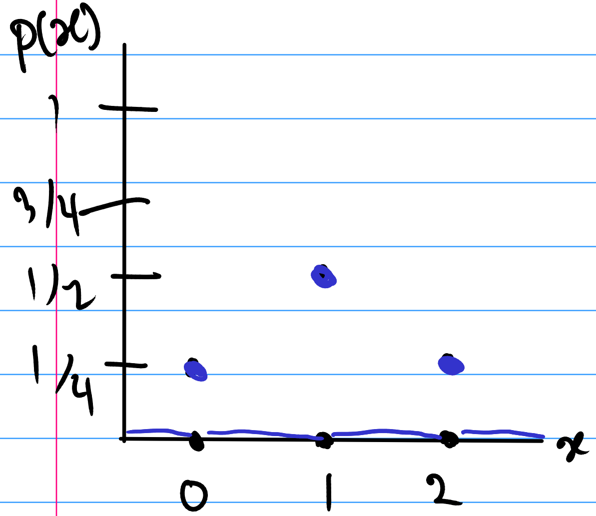
$$= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \lambda^2 e^{-\lambda}/2$$

Another answer: $P(X>2) = \sum_{i=3}^{\infty} p(i)$

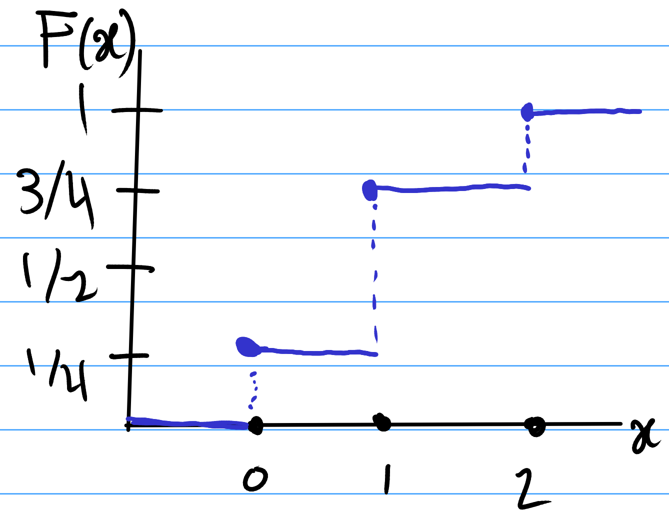
To any Random Variable X can associate another functi[on]
call Cumulative distribution function

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

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



p(x)
1
3/4
1/2
1/4
0
1
2
x



F(x)
1
3/4
1/2
1/4
0
1
2
x

EXPECTED VALUE OR EXPECTATION

So far Random Variables X

Probability mass function $p(x) = P(X = x)$

Cumulative distribution function $F(x) = P(X \leq x)$

If X is a discrete Random variable with probability mass function $p(x)$ then the expectation of X is

$$E[X] = \sum_{x : p(x) > 0} x\, p(x) = \sum_{x \text{ a possible values}} x\, P(X = x)$$

use brackets because it depends on all possible values of X

Ex 1 $X = 0$ or 1 each with prob. $1/2$

$$p(0) = \frac{1}{2} = p(1)$$

$$E[X] = 0\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right) = \frac{1}{2}$$

Ex 2 $p(0) = \frac{1}{3}$ $p(1) = \frac{2}{3}$ $E[X] = 0\frac{1}{3} + 1\frac{2}{3} = \frac{2}{3}$

$E[X]$ is an average or mean value of X

more precisely, a <u>weighted average</u> where each possible value is weighted by the probability that it will occur.

<u>Ex</u> X = roll of a fair 6-sided die

$$x = 1, 2, 3, 4, 5, 6$$

$$E[X] = \sum_{i=1}^{6} i\left(\frac{1}{6}\right) = \frac{1}{6}\sum_{i=1}^{6} i = \frac{1}{6}21 = \frac{7}{2} = 3.5$$

<u>Ex</u> Here's a game

$$P(\text{win}) = \frac{99}{100} \qquad P(\text{lose}) = \frac{1}{100}$$

$$\left.\begin{array}{l} \text{win} \Rightarrow \text{gain } \$1 \\ \text{lose} \Rightarrow \text{lose } \$1000000 \end{array}\right\} \text{payout}$$

X = amount we win or lose

$$E[X] = 1\left(\frac{99}{100}\right) + (-1000000)\frac{1}{100} \approx -10000$$

THEORETICAL EXAMPLE

Let A be an event

I is an <u>indicator</u> variable:

$I = 1$ if A occurs (i.e. for any outcome in A)

$I = 0$ if A^c occurs

$E[I] = ?$

possible values: 0, 1

PMF: $p(0) = P(I=0) = P(A^c) = 1 - P(A)$
$\quad\; p(1) = P(I=1) = P(A)$

$$E[I] = 0\,p(0) + 1\,p(1) = p(1) = P(A)$$

Reprove inclusion-exclusion:

E, F events $\quad P(E \cup F) = P(E) + P(F) - P(EF)$

I_E indicator variable for E

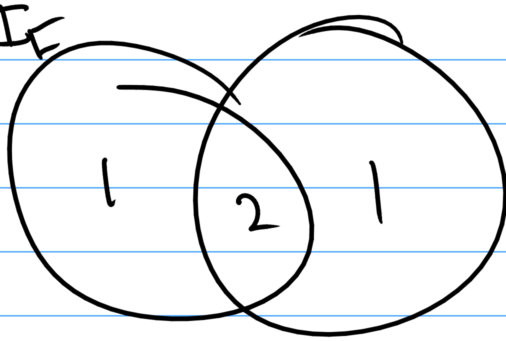
I_F " " " F

$$I_E + I_F = \begin{cases} 0 & \text{if neither } E \text{ nor } F \text{ occurs} \\ 1 & \text{if exactly one of } E \text{ or } F \text{ occurs} \\ 2 & \text{if both occur.} \end{cases}$$

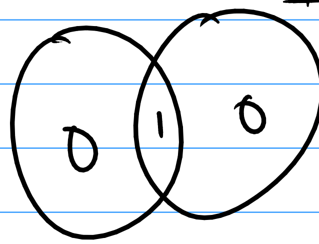
$$I_{EF} = \begin{cases} 1 & \text{if both } E, F \text{ occur} \\ 0 & \text{otherwise} \end{cases}$$

$$I_{E\cup F} = \begin{cases} 1 & \text{if } E \text{ or } F \text{ or both} \\ 0 & \text{otherwise.} \end{cases}$$

$I_E + I_F$

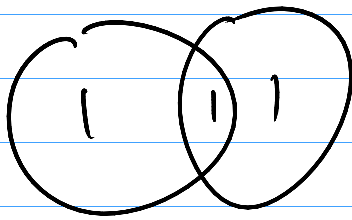


$-$



I_{EF}

$=$



$= I_{E\cup F}$

$$I_{E\cup F} = I_E + I_F - I_{EF}$$

$$E[I_{E\cup F}] = E[I_E] + E[I_F] - E[I_{EF}]$$

$$P(E\cup F) = P(E) + P(F) - P(EF)$$

100 ball in Urn
20 blue
30 green
50 red

Ball is drawn at random and we win amount equal to the number of balls of that color.

X = amount we win

$E[X]$? Possible values: $X = 20, 30, 50$

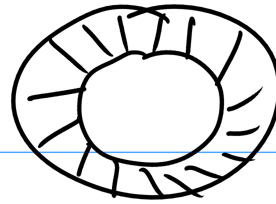
PMF
$p(20) = 20/100 = .2$
$p(30) = 30/100 = .3$
$p(50) = 50/100 = .5$

$$E[X] = 20(.2) + 30(.3) + 50(.5)$$
$$= 4 + 9 + 25 = 38$$

Urn contains just one ball of each color
payouts the same Y = winnings

$$E[Y] = 20\left(\tfrac{1}{3}\right) + 30\left(\tfrac{1}{3}\right) + 50\left(\tfrac{1}{3}\right)$$
$$= \frac{100}{3} = 33\tfrac{1}{3}$$

Roulette



Version 1 wheel w/ 18 red places, 18 black places] # 1-36

X Bet on color red/black → payout 1 to 1 ratio

Y Bet on particular # → payout 35 to 1 ratio

X: lose → $X=-1$ $P(\text{lose}) = \frac{1}{2}$

win → $X=1$ $P(\text{win}) = \frac{1}{2}$

$$E[X] = (-1)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{2}\right) = 0$$

This game is "fair"

Y: lose → $Y = -1$ $P(\text{lose}) = 35/36$

win → $Y = 35$ $P(\text{win}) = 1/36$

$$E[Y] = (-1)\left(\frac{35}{36}\right) + (35)\left(\frac{1}{36}\right) = 0$$

Also fair

Change the wheel 18 red, 18 black } #1-36

two green places 0, 00

Total of 38 places.

payouts are the same as before

X bet on red $X = -1, 1$

$$E[X] = (-1)\left(\frac{20}{38}\right) + 1\left(\frac{18}{38}\right) = -\frac{2}{38} = -\frac{1}{19}$$

Y = bet on #5 $Y = -1, 35$

$$E[Y] = (-1)\left(\frac{37}{38}\right) + (35)\left(\frac{1}{38}\right) = -\frac{2}{38} = -\frac{1}{19}$$

So lose about $\$\frac{1}{19} = 5.3¢$ on the dollar, per game

PROPERTIES OF EXPECTATION and VARIANCE

Recall Random variable X

Probability mass function $p(x) = P(X=x)$

Expecation $E[X] = \sum_{x} x\, p(x)$

where x ranges through all possible values of X.

Next concept:

Def If X is a random variable with $\mu = E[X]$
(μ = expectation or mean of X)

The variance is

$$\mathrm{Var}[X] = E[(X-\mu)^2] = E[(X-E[X])^2]$$

$\mathrm{Var}[X] = E[Y]$ $\quad Y = (X-\mu)^2$

or in other words if $g(t) = (t-\mu)^2$

then $Y = g(X)$

Ex X a random variable taking values $-1, 0, 1$

$P\{X=-1\} = .2$

$P\{X=0\} = .5$

$P\{X=1\} = .3$

$E[X] = (-1)(.2) + 0(.5) + 1(.3)$
$= .1$

$Var(X)$: $\quad Y = (X-.1)^2$

possible values of $X-.1 = \begin{cases} .9 \\ -.1 \\ -1.1 \end{cases}$

possible values of $Y = \begin{cases} (.9)^2 = .81 \\ (-.1)^2 = .01 \\ (-1.1)^2 = 1.21 \end{cases}$

$P(Y=.81) = .3$
$P(Y=.01) = .5$
$P(Y=1.21) = .2$

$Var[X] = E[Y]$
$= (.81)(.3) + (.01)(.5) + 1.21(.2)$

$Z = X^2$ possible values of Z $0, 1$

$P(Z=0) = P(X=0) = .5$

$P(Z=1) = P(X=1) + P(X=-1) = .3 + .2 = .5$

$E[Z] = 0(.5) + 1(.5) = .5$

What about variable X takes values $0, 1, 2, 3, 4, \ldots$

$p(i)$ probability mass function of X

$$Y = \sin\left(\frac{2\pi}{13} X\right)$$

PMF of Y? $\quad p(y) = P(Y = y) = P\left(\sin\left(\frac{2\pi}{13}X\right) = y\right)$

To find expectation of Y it is not necessary to find the probability mass function of Y. In fact you only need PMF of X.

ANOTHER (more fundamental) DEFINITION OF EXPECTATION

Assume S = sample space is discrete

each outcome σ in S has some prob $P\{\sigma\}$

$$P(E) = \sum_{\sigma \text{ in } E} P\{\sigma\}$$

Think of a random variable as a function from S to the set of real numbers

$X : S \to \mathbb{R} \qquad X(\sigma) =$ value of X on outcome σ

$$E[X] = \sum_{\omega} X(\omega)\, P\{\omega\}$$ sum over all outcomes.

Why is it the same as "PMF definition"?

Group terms according to the value of X

$$\sum_{\omega} X(\omega) P\{\omega\} = \sum_{x} \sum_{\substack{\omega \text{ such} \\ \text{that} \\ X(\omega)=x}} X(\omega) P\{\omega\}$$

$$= \sum_{x} \sum_{\substack{\omega \text{ such that} \\ X(\omega)=x}} x\, P\{\omega\} = \sum_{x} x \sum_{\substack{\omega \\ \text{s.t.} \\ X(\omega)=x}} P\{\omega\}$$

$$= \sum_{x} x\, P(X=x) = \sum_{x} x\, p(x)$$

Proposition If Random variable Y is a function of X

$Y = g(X)$ and $p(x)$ is PMF of X.

$$E[Y] = \sum_{\substack{x \\ \text{possible} \\ \text{value} \\ \text{of } X}} g(x)\, p(x)$$

$\boxed{\text{Don't need PMF of } Y}$

Proof $E[Y] = \sum_{\sigma} Y(\sigma)\, P\{\sigma\}$

$$= \sum_{\sigma} g(X(\sigma))\, P(\sigma)$$

$$= \sum_{x} \sum_{\substack{\sigma \text{ s.t.} \\ X(\sigma)=x}} g(x)\, P(\sigma)$$ group terms according to value of X

$$= \sum_{x} g(x) \sum_{\substack{\sigma \text{ s.t.} \\ X(\sigma)=x}} P(\sigma) = \sum_{x} g(x)\, P(X=x)$$

$$= \sum_{x} g(x)\, p(x)$$

$P(X=-1) = .2$ $\quad Y = X^2$

$P(X=0) = .5$

$P(X=1) = .3$

$$E[Y] = (-1)^2(.2) + 0^2(.5) + (1)^2(.3)$$
$$= .5$$

COMMON APPLICATIONS

$Y = X^n \quad E[X^n] = \sum_{x} x^n\, p(x)$

$E[X^n]$ is called the <u>n th moment</u> of X

$$E[aX+b] = \sum_x (ax+b)p(x)$$

$$= a\sum_x xp(x) + b\sum_x p(x)$$

$$= aE[X] + b(1)$$

$$= aE[X] + b$$

IMPORTANT PROPERTY ~~NOT IN CHAP 4~~ in Section 4.9

If X and Y are any R.V.s on the same sample space then

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

$$E[X+Y] = \sum_\theta (X(\theta)+Y(\theta))\,P(\theta)$$

$$= \sum_\theta X(\theta)P(\theta) + \sum_\theta Y(\theta)P(\theta)$$

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

$E[X] = \mu$

$Var[X] = E[(X-\mu)^2]$

ANOTHER FORMULA (often more convenient)

$Var[X] = E[X^2] - \mu^2 = E[X^2] - (E[X])^2$

$(X-\mu)^2 = X^2 - 2\mu X + \mu^2$

$E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu \underbrace{E[X]}_{\mu} + \mu^2$

$= E[X^2] - \mu^2$

Variance / Binomial Random Variables

Next HW:

Problems: 4.30, 4.33, 4.35, 4.38, 4.41, 4.57, 4.59, 4.61

Theoretical Exercises: 4.16, 4.19

Reference for previous lecture §4.4 §4.9

$$E[X] = \sum_{\theta} X(\theta)\, P\{\theta\}$$

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

<u>Ex</u> X = fair die roll $X = 1, 2, 3, 4, 5, 6$

$$E[X] = 3.5$$

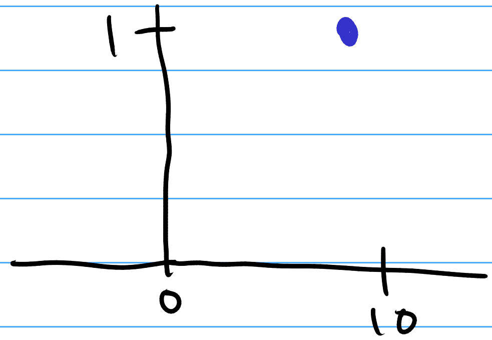
$Var[X]$? $E[X^2] = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right)$

$$= 91\left(\frac{1}{6}\right)$$

$$Var[X] = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \approx 2.9$$

Variance measures the "width" of the range over which X is distributed

$W = 10$ with prob 1



$$E[W] = 10$$

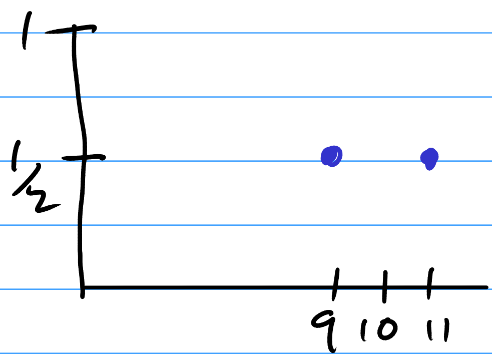
$$\mathrm{Var}[W] = E[(W-10)^2]$$

$$(W-10)^2 = 0 \quad \text{with prob } 1$$

$$\mathrm{Var}[W] = 0$$

$Y = 11$ with Prob. $\frac{1}{2}$
$Y = 9$ with prob $\frac{1}{2}$

$$E[Y] = 10$$



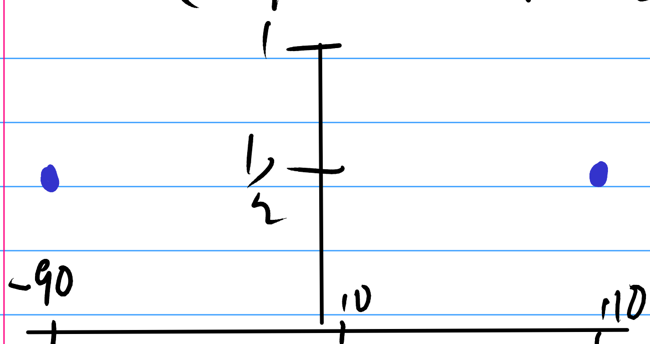
$$\mathrm{Var}[Y] = E[(Y-10)^2]$$

$$(Y-10) = \begin{cases} 1 & \text{prob } \frac{1}{2} \\ -1 & \text{prob } \frac{1}{2} \end{cases}$$

$$\mathrm{Var}[Y] = (1)^2 \frac{1}{2} + (-1)^2 \frac{1}{2} = 1$$

$$Z = \begin{cases} 110 & \text{with prob. } \frac{1}{2} \\ -90 & \text{with prob } \frac{1}{2} \end{cases}$$

$$E[Z] = 110 \frac{1}{2} + (-90)\frac{1}{2} = 10$$



$$(Z-10)^2 = (100)^2 = 10000$$

with prob 1

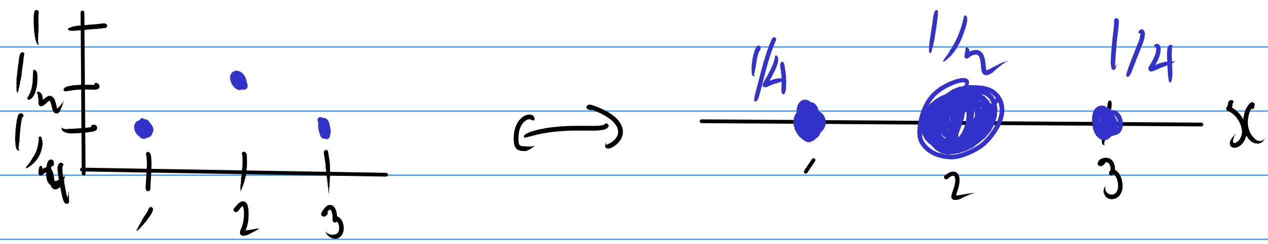
$$\mathrm{Var}[Z] = 10000$$

Standard Deviation $SD[X] = \sqrt{Var[X]}$

$SD[W] = 0$ $\quad SD[Y] = 1$ $\quad SD[Z] = 100$

So variance measures how widely dispersed the values of X are.

<u>Analogy</u>: If think of the probability mass function as describing the masses of certain bodies arranged on a line:



$E[X] \leftrightarrow$ center of mass

$Var[X] \leftrightarrow$ moment of inertia about the center of mass

Bernoulli trials and Binomial Random Variables

Recall Bernoulli trial $\quad P(\text{success}) = p$
$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad P(\text{failure}) = 1-p$

Suppose X is a random variable which is 1 on success and 0 on failure

Prob. mass function $\quad p(0) = 1-p$
$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad p(1) = p$

X is "Bernoulli Random variable with parameter p"

Now suppose we do n independent bernoulli trials

$$X = \# \text{ of successes obtained in } n \text{ trials}$$

X takes values $0, 1, \ldots, n$

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

X is called "Binomial random variable with parameters (n, p)"

$\#$ of trials $\quad$ prob of success per trial

$$1 \overset{?}{=} \sum_{i=0}^{n} p(i) = \sum_{i=0}^{n} \binom{n}{i} \underset{\overset{\shortparallel}{x}}{p^i} \underset{\overset{\shortparallel}{y}}{(1-p)^{n-i}} = (x+y)^n = (p+1-p)^n$$
$$= 1^n = 1$$

Examples

Basic (seen before) Five fair coins are flipped

$X = \#$ heads

Then X is a binomial Random variable w/ parameters $(n=5, p=\frac{1}{2})$

$$P(X=2) = \binom{5}{2}\left(\tfrac{1}{2}\right)^2\left(\tfrac{1}{2}\right)^3 = \frac{10}{32}$$

Example Sell cases of wine. each case contains 20 bottles

Prob of bad bottle $= .05$

Money-back gaurantee that case will contain no more than 1 bad bottle.

$X = \#$ of bad bottles in a case (20 bottles)

$P(\text{have to give money back}) = P(X \geq 2)$

$= 1 - P(X=0) - P(X=1)$

X is binomial R.V. with parameters

$(n=20,\ p=.05)$

Bernoulli trial = check whether a bottle is bad

P(success) = P(bottle is bad) = .05

P(failure) = P(bottle is good) = .95

$P(X\geq 2) = 1 - P(X=0) - P(X=1)$

$= 1 - \binom{20}{0}(.05)^0(.95)^{20} - \binom{20}{1}(.05)^1(.95)^{19}$

$= .26$

Binomial Cont'd / Poisson R.V.s

Recall: Binomial Random variable

X takes values $0, 1, 2, \ldots, n$

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

This represents the # of successes in n independent Bernoulli trials with p = probability of success

What is expectation and variance?

Involves recursive relationship between the moments of X

Trick: If X is a binomial RV w/ parameters (n,p) and Y is a binomial RV w/ parameters $(n-1,p)$

Moments $E[X^k] = np\,E[(Y+1)^{k-1}]$ **

$E[X] = np\,E[(Y+1)^0] = np \cdot 1 = np$

$E[X^2] = np\,E[(Y+1)^1] = np(E[Y]+1)$

$= np((n-1)p+1) = n^2p^2 + np(1-p)$

$$\mathrm{Var}[X] = E[X^2] - (E[X])^2$$

$$= n^2p^2 + np(1-p) - (np)^2$$

$$= np(1-p)$$

So: $E[X] = np \Rightarrow E\left[\frac{X}{n}\right] = p$

$\frac{X}{n}$ is average number of successes in n trials

Expected proportion of successes is p.

Shape binomial probability mass function

with parameters (n, p)

$p(k) = P(X=k)$, then $p(k)$ first increases

monotonically, then decreases monotonically

it obtains its maximum value when

$k = \lfloor (n+1)p \rfloor$ ← rounded down

Poisson Random Variable X with parameter λ

X has possible values $0, 1, 2, 3, \ldots$ up to ∞

$$p(i) = P(X=i) = e^{-\lambda} \frac{\lambda^i}{i!}$$

Recall $$\sum_{i=0}^{\infty} p(i) = e^{-\lambda} \underbrace{\left(\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \right)}_{\text{taylor series for } e^{\lambda}} = e^{-\lambda}\left(e^{\lambda}\right) = 1$$

Poisson random variable also known as "law of rare events"

E.g. # misprints per page in a book

people who live to 100

wrong telephone numbers dialed in day

α-decays per unit time per unit mass in a radioactive material.

Poisson distribution approximates binomial distribution when n is large and p is small and np is of moderate size. $(\lambda = np)$

X is binomial w/ parameters (n, p)

set $\lambda = np$ hence $p = \frac{\lambda}{n}$

$$P(X=i) = \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{(n-i)!\, i!} \left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{n-i}$$

$$= \frac{(n)(n-1)(n-2)\cdots(n-i+1)}{n^i} \; \frac{\lambda^i}{i!} \; \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^i}$$

Take $n \to \infty$ $\quad \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx \frac{n^i}{n^i} = 1$

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

$$\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}$$ because $\lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n = e$

$$P(X=i) \approx e^{-\lambda} \frac{\lambda^i}{i!}$$

<u>Examples</u> Suppose # typos per page in a book is a Poisson RV with parameter $\lambda = \frac{1}{2}$

$$P(X \geq 1) = 1 - P(X=0) = 1 - e^{-\lambda} = 1 - e^{-1/2} = .393$$

$$E[X] = \sum_{i=0}^{\infty} i\, e^{-\lambda} \frac{\lambda^i}{i!} = \sum_{i=1}^{\infty} \frac{e^{-\lambda} \lambda^i}{(i-1)!}$$

$$= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

(let $j = i - 1$)

$$= \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Bernoulli, Poisson and Random Processes

(cf. pp. 152–154)

Recall: Poisson Random Variable X

parameter λ

Possible values $0, 1, 2, \ldots$ up to ∞

Prob. Mass function $p(X=i) = e^{-\lambda}\dfrac{\lambda^i}{i!}$

$E[X] = \lambda$

$Var[X] = \lambda \qquad SD[X] = \sqrt{\lambda}$ standard deviation

$$E[X] = \sum_{i=0}^{\infty} i\, e^{-\lambda}\frac{\lambda^i}{i!} = \sum_{i=1}^{\infty} e^{-\lambda}\frac{\lambda^i}{(i-1)!}$$

$$= \lambda \sum_{i=1}^{\infty} e^{-\lambda}\frac{\lambda^{i-1}}{(i-1)!} = \lambda \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!} \quad (j=i-1)$$

$$= \lambda e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

$$E[X^2] = \sum_{i=0}^{\infty} i^2 e^{-\lambda}\frac{\lambda^i}{i!} = \lambda \sum_{i=1}^{\infty} i\, e^{-\lambda}\frac{\lambda^{i-1}}{(i-1)!}$$

$$= \lambda \sum_{j=0}^{\infty}(j+1)\, e^{-\lambda}\frac{\lambda^j}{j!}$$

$$= \lambda \left[\underbrace{\sum_{j=0}^{\infty} j e^{-\lambda} \frac{\lambda^j}{j!}}_{\substack{\lambda \\ \text{by previous argument}}} + \underbrace{\sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!}}_{\substack{1 \text{ because sum of} \\ \text{all values of PMF}}} \right]$$

$$= \lambda(\lambda + 1) = E[X^2]$$

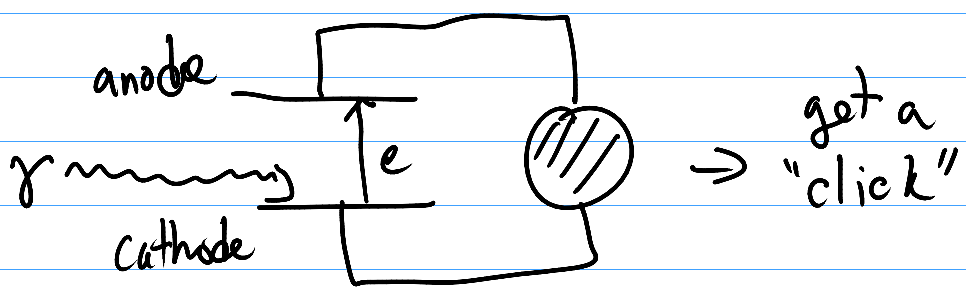
$$\mathrm{Var}[X] = E[X^2] - (E[X])^2 = \lambda(\lambda+1) - \lambda^2$$

$$= \lambda$$

$$E[X^k] = \lambda E[(X+1)^{k-1}]$$

<u>Poisson Random Process</u>

<u>Example</u> Geiger Counter



Start counting clicks at time 0

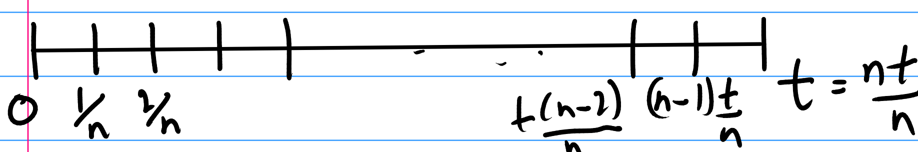
$N(t)$ = # of clicks in the time interval $[0,t)$
from 0 to t

For each t, $N(t)$ is a Random variable

But the whole family of Random variables depends on another parameter t, which can be any positive real number

So the collection of random variables $N(t)$ describes a continuous-time random process
(stochastic process)

Want to compute $P(N(t)=k)$



① Memorylessness: If two intervals of time are disjoint, then behavior of the process in those two intervals is independent.

② Some assumptions about the rate of "clicks"

$P(\geq 1 \text{ click in interval of length } h) = \lambda h + o(h)$

(depends only on the length of the interval)
roughly proportional to length of interval

λ = expected rate of clicks
units $1/\text{time}$

(Note not the same λ as before)

③ Can't get two clicks at exactly the same t

$P(\geq 2 \text{ clicks in interval of length } h) = o(h)$

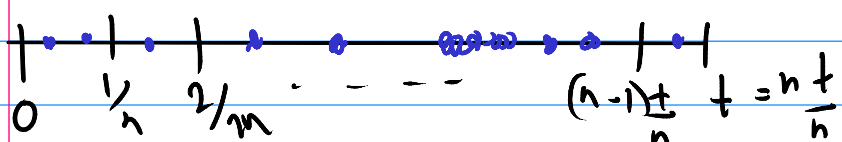
Landau Notation: $f(h) = o(h)$ as $h \to 0$

means $\displaystyle\lim_{h \to 0} \frac{f(h)}{h} = 0$

$f(h)$ goes to zero faster than h

so in the $h \to 0$ limit, $f(h)$ is negligible.

Subdivide the interval into bits of length $\frac{t}{n}$



take $(n \to \infty)$

$P(N(t) = k) = P(k \text{ intervals contain exactly } 1 \text{ click, } n-k \text{ intervals contain exactly } 0 \text{ clicks})$

$+ P(N(t) = k \text{ and some subinterval contains } \geq 2 \text{ clicks})$

$\boxed{\text{Step 1.}}$

The second term is negligible

$P(\text{some subinterval contains} \geq 2 \text{ clicks})$

$= P\left(\bigcup_{i=1}^{n} \{i\text{th subinterval contains} \geq 2 \text{ clicks}\}\right)$

$$\leq \sum_{i=1}^{\infty} P(i\text{th subinterval contains} \geq 2 \text{ clicks})$$

(This is a corollary of inclusion-exclusion called Boole's inequality)

$$= \sum_{i=1}^{n} o\left(\frac{t}{n}\right) = n\, o\left(\frac{t}{n}\right) = t\, n\, o\left(\frac{1}{n}\right) \to 0 \text{ as } n \to \infty$$

$$\shortparallel$$

$$t\, \frac{o(1/n)}{1/n} \to 0 \text{ as } \frac{1}{n} \to 0$$

$$P(N(t)=k) \approx P\left(\begin{matrix} k \text{ subintervals contain 1 click} \\ n-k \text{ contain 0 clicks} \end{matrix}\right)$$

approximate this probability by the binomial distribution

$$= \binom{n}{k}\left(\lambda \frac{t}{n} + o\left(\frac{t}{n}\right)\right)^{k}\left(1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right)\right)^{n-k}$$

Like Bernoulli trials with $p = \lambda \frac{t}{n} + o\left(\frac{t}{n}\right)$

$$np = n\left(\lambda \frac{t}{n} + o\left(\frac{t}{n}\right)\right) = \lambda t + n\, o\left(\frac{t}{n}\right) \to \lambda t$$

as $n \to \infty$

$$P(N(t)=k) = e^{-(\lambda t)} \frac{(\lambda t)^k}{k!}$$

Expected # of clicks in interval of length t is $E[N(t)] = \lambda t$

$$\text{Expected Rate} = E\left[\frac{N(t)}{t}\right] = \frac{1}{t}E\left[N(t)\right] = \lambda$$

Note: $o(h)$ notation will not be on the test!!!

More discrete probability distributions

NEXT Homework

Ch 4 Problems 4.63, 4.70, 4.71, 4.75, 4.78
Theoretical 4.27, 4.29
Ch 5 Problems 5.1

Sources of random variables – Random Processes

Bernoulli trials $P(\text{success}) = p$ $P(\text{failure}) = 1-p$

$N(n)$ = # successes in n trials (binomial R.V. parameters (n, p))

Y = # of trials required to get first success
(geometric Random variable)

$Z(r)$ = # of trials required to get first r successes
(negative binomial random variable)

Binomial Random Variable X

- Parameters $n = \#$ of trials, $p =$ prob. of success
- Possible values $0, 1, 2, \ldots, n$
- Prob. Mass function $P(X=i) = \binom{n}{i} p^i (1-p)^{n-i}$
- $E[X] = np$
- $Var[X] = np(1-p)$
- SLOGAN: X represents the number of successes in n Bernoulli trials, where each trial has probability p of success
- EXAMPLE: A die is rolled 10 times
  let $X = \#$ times a 6 appears

then X is a binomial random variable with
$n = 10$, $p = 1/6$

- **Remark:** Binomial Random variable with $n=1$ is called Bernoulli random variable

Geometric random variable X

- Parameters p
- Possible values $1, 2, 3, \ldots, \infty$
- Probability mass function

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

EXAMPLE Roll a die until it comes up 6
$X = \#$ rolls required

$$P(X=3) = P(\text{not } 6, \text{not } 6, 6) = \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$$

$$P(X=i) = P(\text{not } 6 \ (i-1) \text{ times}, 6) = \left(\frac{5}{6}\right)^{i-1} \left(\frac{1}{6}\right)$$

Remark on the name

$$\sum_{i=1}^{\infty} P(X=i) = \sum_{i=1}^{\infty} (1-p)^{i-1} p = p \sum_{i=1}^{\infty} (1-p)^{i-1}$$

$$\left(\sum_{i=1}^{\infty} a r^{i-1} = a \frac{1}{1-r} \right)$$

$$= p(1) \frac{1}{1-(1-p)} = p \cdot \frac{1}{p} = 1$$

- $E[X] = \frac{1}{p}$
- $Var[X] = \frac{1-p}{p^2}$

Example Sampling with replacement:

10 balls in an urn. One is white, 9 black
Balls are drawn, then replaced

X = # tries required to get the white ball

X is geometric random variable with parameter $p = \frac{1}{10}$

What is the probability that more than k tries are needed?

$$P(X \geq k) = \sum_{i=k}^{\infty} (1-p)^{i-1} p \qquad \text{first term} = (1-p)^{k-1} p$$

$$\text{ratio} = (1-p)$$

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

Negative binomial random variable

- Parameters $r = $ # successes required, p
- Possible value $r, r+1, r+2, \ldots, \infty$
- Probability Mass function

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

in order to have $X=i$, need exactly $r-1$ successes in $i-1$ trials, and need success on rth trial

$$\binom{i-1}{r-1} p^{r-1} (1-p)^{(i-1)-(r-1)} \cdot p$$

- Expectation $E[X] = \frac{r}{p}$
- Variance $Var[X] = \frac{r(1-p)}{p^2}$

- Geometric Random variable is negative binomial with $r=1$

Another process sampling without replacement

N balls in urn   m are "red"
N−m are "blue"

Draw n balls without replacement

X = # of red balls in this sample of size n.

X is called Hypergeometric Random Variable

- Parameters n, m, N
- Possible values $0, 1, \ldots, n$
- Probability mass function

$$P(X=i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

- Expectation $E[X] = n\frac{m}{N}$

(same as with replacement!)

- Variance $\mathrm{Var}[X] = n\frac{m}{N}\left(1-\frac{m}{N}\right)\left(1-\frac{n-1}{N-1}\right)$

Take $m, N \to \infty$ in such a way that $p = \frac{m}{N}$ remains fixed, then the hypergeometric random variable approaches a binomial random variable

Continuous Random Variables (§5.1)

Next Midterm 3/23

Material up to lecture 22 (today)

- No book, No calculator
- Yes Notes
  - 1 US Letter (8.5" x 11") sheet
  - 2-sided
  - Handwritten by you (collaboration allowed)
  - No Xeroxing
  - Cram as much in as you want
  - No Magnifying glasses

⚠ Might see things like $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$

Continuous Random Variables are just like discrete R.V. <u>except</u>

every sum becomes an <u>integral</u>

Continuous R.V. can take a continuous range of values

Eg. Possible values of X

could be $(0,1) = \{x \in \mathbb{R} \mid 0 < x < 1\}$

$[0,1] = \{x \mid 0 \leq x \leq 1\}$

$[0,1) = \{x \mid 0 \leq x < 1\}$

$(-\infty, \infty)$ = all real numbers

X = price of a stock

X = time that a machine works before breakdown

X = error in an experimental measurement.

How to describe probability associated with X?

[Recall for discrete R.V., we have probability mass function $p(i) = P(X=i)$

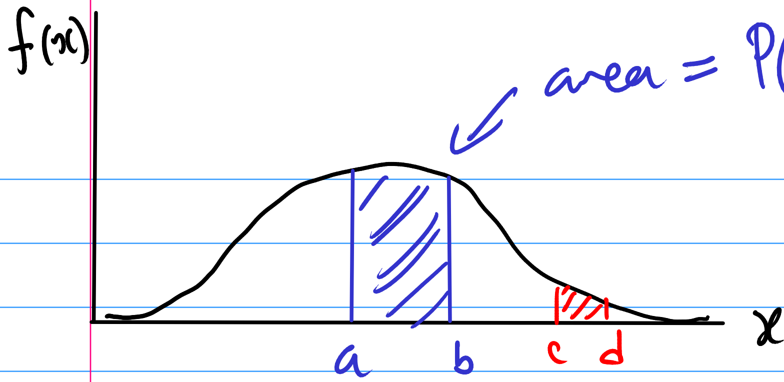
$$P(a \leq X \leq b) = \sum_{a \leq i \leq b} P(X=i)$$]

For continuous R.V. we have a function $f(x)$

such that $P(a \leq X \leq b) = \int_a^b f(x)\,dx$

$f(x)$ is called the <u>probability density function</u> of X

area = $P(a \le X \le b)$

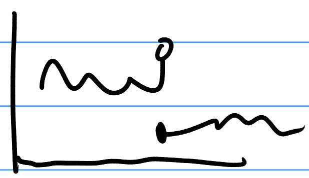


Because integral = area under curve

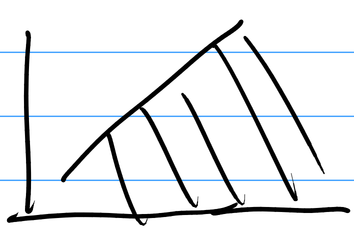
$$P(a \le X \le b) > P(c \le X \le d)$$

Basic properties of f

- $f(x) \ge 0$ everywhere
- $\int_{-\infty}^{\infty} f(x)\,dx = P(-\infty \le X \le \infty) = 1$
- $f(x)$ need not be a continuous function but it will usually be piecewise continuous



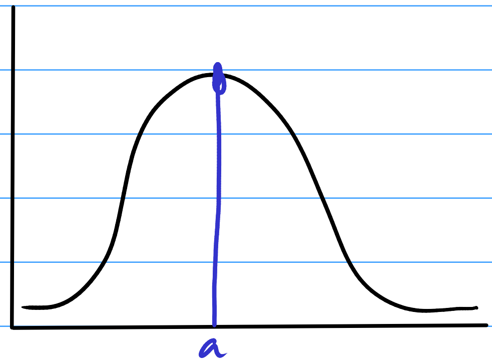
- $\int_{-\infty}^{\infty} f(x)\,dx$ exists $\Rightarrow$ $f(x) \to 0$ as $x \to \pm\infty$



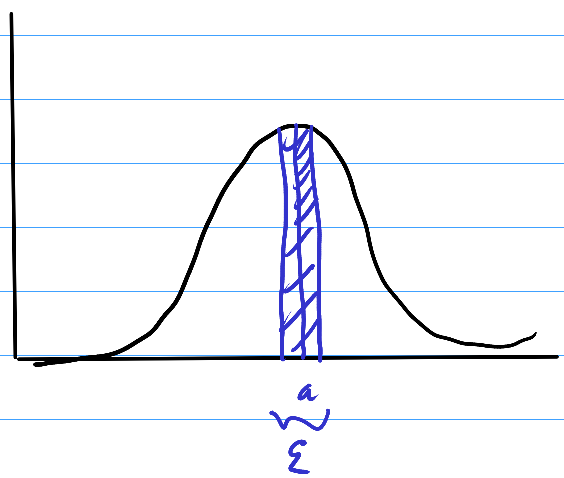
no

yes

$$P(X=a) = \int_a^a f(x)\,dx = 0 \neq f(a)$$



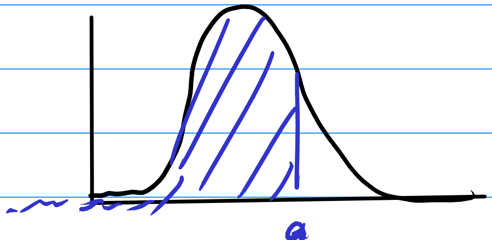
$$P\left(a-\frac{\varepsilon}{2} \leq X \leq a+\frac{\varepsilon}{2}\right) = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f(x)\,dx \approx \text{width} \times \text{height}$$
$$= \varepsilon f(a)$$



Also still have Cumulative distribution function

[for discrete RV $F(a) = \sum_{i \leq a} P(X=i)$]

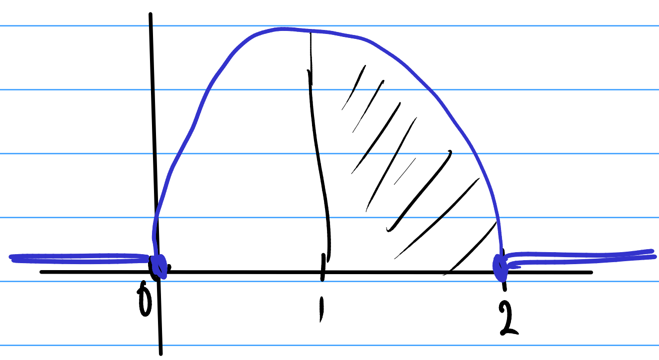
$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)\,dx$$



$P(a \le X \le b)$
$P(a \le X < b)$
$P(a < X \le b)$
$P(a < X < b)$
} all equal

Example X continuous R.V. with density

$$f(x) = \begin{cases} C(4x-2x^2) & \text{if } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$



C is a normalization

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx$$

$$= \int_{-\infty}^{0} 0\,dx + \int_{0}^{2} C(4x-2x^2)\,dx + \int_{2}^{\infty} 0\,dx$$

$$= \int_{0}^{2} C(4x-2x^2)\,dx$$

$$= C\left[2x^2 - \frac{2x^3}{3}\right]_0^2 = C \cdot \frac{8}{3} \longrightarrow C = \frac{3}{8}$$

$$P(X>1) = \int_1^\infty f(x)\,dx = \int_1^2 f(x)\,dx \left(+ \int_2^\infty 0\,dx \right)$$

$$\int_1^2 \frac{3}{8}(4x-2x^2)\,dx = \frac{1}{2}$$

Relationship between PDF (density) and CDF (cum dist.)

$$F(a) = \int_{-\infty}^a f(x)\,dx$$

$$\frac{d}{da}F(a) = \frac{d}{da}\int_{-\infty}^a f(x)\,dx \overset{FTC}{=} f(a)$$

Continuous Random Variables: expectation and variance.

- Reminder test on Friday up to lecture 22
- One 8.5" x 11" sheet of notes 2-sided
- Must be handwritten: no Xeroxes

Recall X is a continuous random variable

Possible values of X is some interval

$$(a,b) = \{x \mid a < x < b\}$$

$$[a,b] = \{x \mid a \le x \le b\}$$

or $[a,b) = \{x \mid a \le x < b\}$

etc.

or $(-\infty, \infty)$ = all real numbers

X has a probability density function (PDF) $f(x)$ such that

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

Cumulative distribution function $F(a)$

$$F(a) = P(X \le a) = P(-\infty \le X \le a)$$
$$= \int_{-\infty}^{a} f(x)\,dx$$

Example Amount of time a computer functions before breaking is a continuous random variable X

Let's say PDF of X is given by

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \ge 0 \quad \text{in months} \\ 0 & x < 0 \end{cases}$$

λ is a normalization constant: choose it so that

$$P(-\infty \le X \le \infty) = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} \lambda e^{-x/100}\,dx = \lambda \int_0^{\infty} e^{-x/100}\,dx$$

$$\left[\quad u = -\frac{x}{100} \quad du = -\frac{dx}{100} : \int_0^{a} e^{-x/100}\,dx = \int_0^{-\infty} e^{u}\,du\,(-100) \right.$$

$$= -100\int_0^{-\infty} e^u\,du = -100\left[e^u\right]_0^{-\infty} = -100\left(0 - e^0\right) = 100]$$

$$\left[\int e^{au}\,du = \frac{1}{a}e^{au} + C\right]$$

$$1 = \lambda(100) \qquad \lambda = \frac{1}{100}$$

$$f(x) = \begin{cases} \frac{1}{100}e^{-x/100} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Q: $P(50 < X < 150) = \displaystyle\int_{50}^{150} \frac{1}{100}e^{-x/100}\,dx$

$$= \left[\frac{1}{100}(-100)e^{-x/100}\right]_{50}^{150} = \left[-e^{-x/100}\right]_{50}^{150}$$

$$= -e^{-150/100} - \left(-e^{-50/100}\right) = e^{-1/2} - e^{-3/2}$$

Expectation (Mean or Average Value)

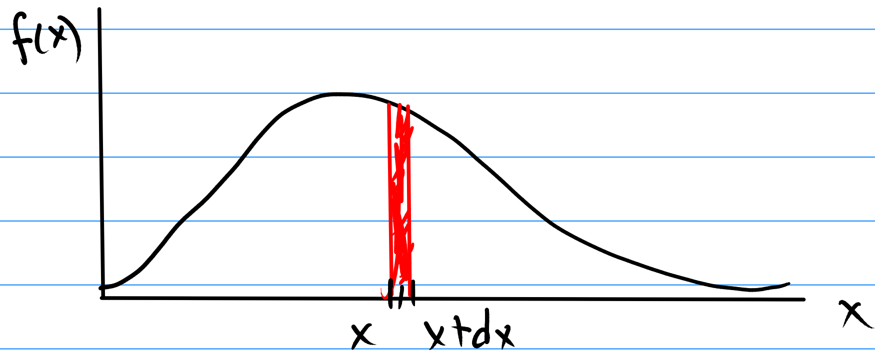
For discrete R.V. X has possible values x_i

$$P(a \le X \le b) = \sum_{a \le x_i \le b} P(X = x_i)$$

$$E[X] = \sum_{x_i} x_i \, P(X = x_i)$$

For continuous Random variable

$$f(x)\,dx \approx P(x \le X \le x + dx)$$



Breakup real line into segments of width $\frac{1}{n}$

index the segments as $\left[\frac{i}{n}, \frac{i+1}{n}\right]$ $i = -\infty, \ldots, \infty$

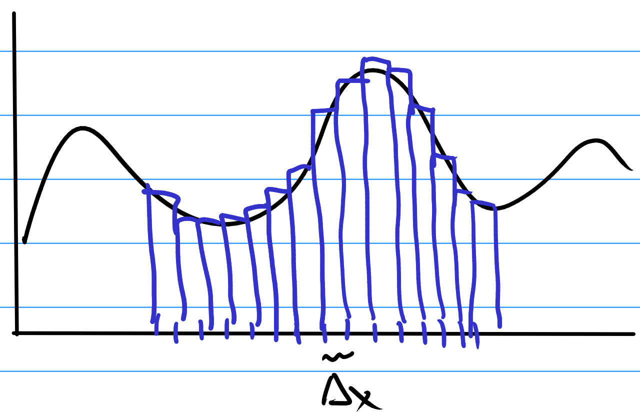
$$E[X] \approx \sum_{i=-\infty}^{\infty} \left(\frac{i}{n}\right) P\left(\frac{i}{n} \le X \le \frac{i}{n} + \frac{1}{n}\right)$$

$$\approx \sum_{i=-\infty}^{\infty} \left(\frac{i}{n}\right) f\left(\frac{i}{n}\right) \frac{1}{n}$$

We recognize this as a riemann sum

as $n \to \infty$, this sum converges to the integral

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx$$



$\sum f(x_i)\, \Delta x_i$

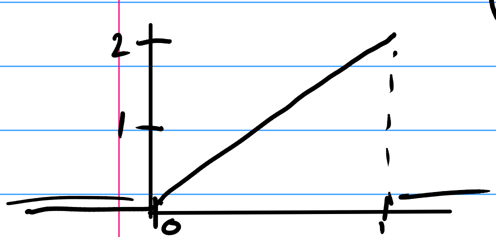
Variance: definition interms of expectation is same as before

$\mu = E[X]$ then $\mathrm{Var}[X] = E\left[(X-\mu)^2\right]$

$$= E[X^2] - \mu^2$$

Ex X has probability density function

$$f(x) = \begin{cases} 2x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$



Can check this is normalised properly

Find $E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x f(x) dx$

$$= \int_0^1 x(2x)\, dx = \int_0^1 2x^2 dx = \left[2\frac{x^3}{3}\right]_0^1$$

$$= \frac{2}{3} - 0 = \frac{2}{3}$$

$\mathrm{Var}[X] = E[X^2] - \left(\frac{2}{3}\right)^2$

Fact (see next lecture for justification)

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$= \int_0^1 x^2 (2x) dx = \frac{1}{2}$$

Review for Exam 2

- 1 sheet of notes (2-sided) allowed
- Must be handwritten by you.

Next HW: Problems 5.2, 5.3, 5.4, 5.6, 5.7, 5.8

Random Variable = numerical function of the outcome of a probabilistic experiment.

Discrete RV = possible values can be indexed by integers
(or set of possible values is finite)

Possible values $X = x_1, x_2, x_3, \ldots$

For Discrete RV, have probability mass function (PMF)

$$p(x) = P\{X = x\}$$

This function contains most important information about X.

$1 \geq p(x) \geq 0 \qquad \sum_{\substack{x \text{ possible} \\ \text{values} \\ \text{of } X}} p(x) = 1$

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

Expectation value = mean or average value

$$E[X] = \sum_{\substack{x \text{ possible} \\ \text{values}}} x\, p(x) = \sum_{x} x\, P\{X = x\}$$

Eg: In gambling examples X = winnings in a game

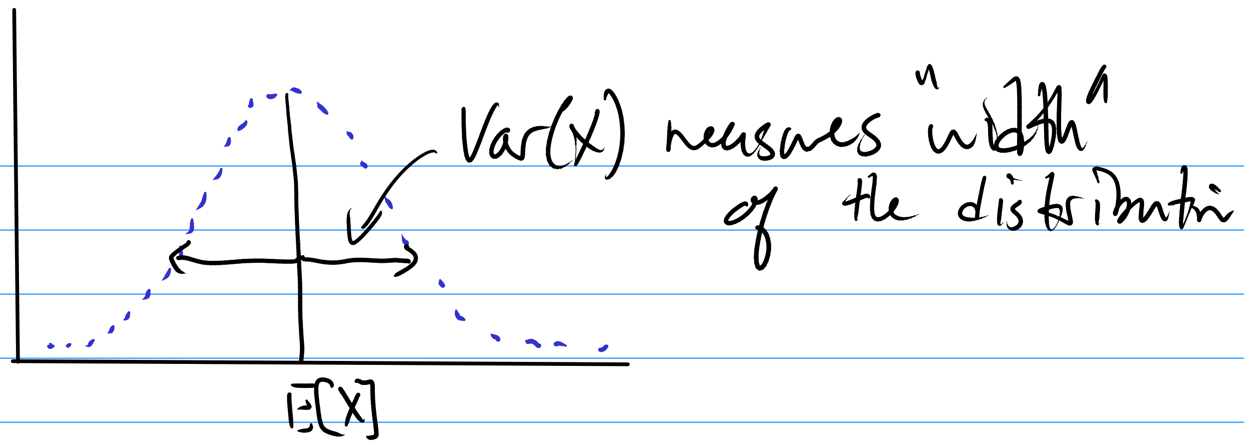
$E[X]$ = average winnings per game.

Variance $\quad Var(X) = E\left[(X - E[X])^2\right]$

compute with this one $\rightarrow \quad = E[X^2] - (E[X])^2$

Variance measures how far X varies from its expectation value.

Small variance $\Rightarrow$ unlikely that X will deviate far from $E[X]$



Properties of $E[X]$

Alternative formula for expectation

Sample space S' outcome $\sigma \in S'$

X is a function of σ

$$E[X] = \sum_{\sigma \in S'} X(\sigma)\, P\{\sigma\}$$

If $p(x) = P\{X = x\}$ is PMF of X

and $Y = g(X)$ where $g(x)$ is some function

$$E[Y] = \sum_{\substack{x \text{ possible} \\ \text{value} \\ \text{of } X}} g(x)\, p(x) = \sum_x g(x)\, P\{X = x\}$$

$$E[aX+b] = a\,E[X] + b \quad (a, b \text{ constants})$$

$$E[X^n] = \sum_x x^n P\{X = x\}$$

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

$$= \sum_x x^2 P\{X=x\} - \left(\sum_x x P\{X=x\}\right)^2$$

If X and Y are two random variables

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

Particular Random variables

Binomial
Poisson
Geometric
Negative binomial
Hypergeometric

all related to Bernoulli trials

sampling w/o replacement

Poisson process

Bernoulli trials (independent $P(\text{success}) = p$)

$X = \#$ of success in n trials

PMF $\quad P\{X=i\} = \binom{n}{i} p^i (1-p)^{n-i} \quad i=0,\ldots,n$

$E[X] = np$

$Var(X) = np(1-p)$

$\boxed{\text{Binomial w/ params } (n,p)}$

Geometric RV:

$Y = \#$ of trials to first success

$P\{Y=i\} = (1-p)^{i-1} p \quad i=1,2,3,\ldots$

$E[Y] = \frac{1}{p} \quad Var(Y) = \frac{1-p}{p^2}$

Negative Binomial

$Z = \#$ of trials for r successes

$P\{Z=n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r} \quad n=r, r+1, \ldots$

$E[Z] = \frac{r}{p} \quad Var(Z) = \frac{r(1-p)}{p^2}$

Poisson Process expected rate λ of events

$X(t)$ = # of event in interval of time of length t

$$P\{X(t)=k\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

$$E[X(t)] = \lambda t \qquad Var(X(t)) = \lambda t$$

Or we could take λ to be the expected # of events.

The Poisson random variable with parameter λ

$$P\{X=k\} = e^{-\lambda}\frac{\lambda^k}{k!}$$

Binomial to Poisson approximation

If n is large, p is small, but np is moderate

then Binomial RV w/ parameters (n,p)

$\approx$ Poisson RV w/ parameter $\lambda = np$

That is

$$P\{X=i\} = \binom{n}{i} p^i (1-p)^{n-i} \approx e^{-\lambda} \frac{\lambda^i}{i!}$$

(where $\lambda = np$)

Hypergeometric N balls m red
N−m blue

Select n balls w/o replacement

$X = \#$ of red in sample

$$P\{X=i\} = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

$$E[X] = \frac{nm}{N}$$

Continuous RV X has probability density function $f(x)$

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

An Example of expectation

Exam 2 Stats

	Raw	Curved
Mean	78.1	85.1
1st Quartile	61.25	73.6
Median	81.5	87.4
3rd Quartile	96.75	97.8

$$\text{curved} = 100 - \left(\frac{15}{22}\right)(100 - \text{raw})$$

Continuous Random Variable X w/ probability density function $f(x)$

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

$$E[X] = \int_{-\infty}^{\infty} x f(x)\,dx$$

Example: Distributor sends parts to factory: shipping time is a random variable X w/ PDF $f(x)$.

If parts are early by s days, cost cs to store them.

If parts are late by s days, cost ks

due lost time.

When to send parts to minimize these costs?

t = # of days before needed that we send the parts

$t=10 \iff$ send parts 10 days before needed

$C_t(X)$ cost if we send parts t days before and it takes X days to arrive

Early : $(X-t) \leq 0 \quad C_t(X) = c(t-X)$

Late : $(X-t) \geq 0 \quad C_t(X) = k(X-t)$

$X = t \quad C_t(X) = 0$

X is a positive R.V. $\quad f(x) = 0$ for $x < 0$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

(analogous to $E[g(X)] = \sum_x g(x) P(X=x)$ for discrete)

$$E[C_t(X)] = \int_0^{\infty} C_t(x) f(x)\, dx$$

$$= \int_0^t c(t-x) f(x)\,dx + \int_t^\infty k(x-t) f(x)\,dx$$

$$= ct\int_0^t f(x)\,dx - c\int_0^t x f(x)\,dx + k\int_t^\infty x f(x)\,dx - kt\int_t^\infty f(x)\,dx$$

$\frac{d}{dt}$[this]

$$\frac{d}{dt}\, ct\int_0^t f(x)\,dx = c\int_0^t f(x)\,dx + ct\,\frac{d}{dt}\int_0^t f(x)\,dx$$

$\int_0^t f(x)\,dx = F(t)$ cumulative dist. func.; $\frac{d}{dt}\int_0^t f(x)\,dx = f(t)$

$$\frac{d}{dt}\int_t^\infty x f(x)\,dx = \frac{d}{dt} - \int_\infty^t x f(x)\,dx = -t f(t)$$

$$\frac{d}{dt}\left[t\int_t^\infty f(x)\,dx\right] = \int_t^\infty f(x)\,dx + t\,\frac{d}{dt}\int_t^\infty f(x)\,dx$$

$$= [1 - F(t)] - t f(t)$$

$$\frac{d}{dt}(\text{stuff}) = cF(t) + ct f(t) - ct f(t)$$

$$- kt f(t) - k[1 - F(t)] + kt f(t)$$

$$= cF(t) - k[1 - F(t)]$$

$$= (c+k) F(t) - k$$

min $\quad 0 = \frac{d}{dt}(\text{stuff}) = (c+k)F(t) - k$

need $F(t) = \frac{k}{k+c}$

solution $t = F^{-1}\left(\frac{k}{k+c}\right)$

$F^{-1}(u)$ is inverse of cumulative distribution function

also known as quartile function of X.

→ Properties of Expectation
→ Uniform Random Variables

Next HW
Problems: 5.12, 5.13, 5.15, 5.16, 5.18
5.23, 5.27
Theoretical: 5.2, 5.3, 5.9

If X has PDF $f(x)$

$$E[X] = \int_{-\infty}^{\infty} x f(x)\,dx$$

Theorem | If $Y = g(X)$ (g some function)
then $E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

Corollary (1) $g(x) = x^n$
$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x)\,dx$$

eg. $\mathrm{Var}(X) = E[X^2] - (E[X])^2$

(2) $E[aX+b] = aE[X] + b$ a, b constant

Also true: $E[X+Y] = E[X] + E[Y]$

[Another definition of expectation

$E[X] = \sum_{\theta} X(\theta) P(\theta)$ sum over $\theta \in S$ outcomes

$E[X] = \int_S X\,dP$ integral w.r.t. probability measure]

Def Y is a nonnegative R.V. $(Y \geq 0)$

if $P(Y<0) = 0$

$$0 = P(Y<0) = \int_{-\infty}^{0} f(y)\,dy$$

implies $f(y) = 0$ (since $f(y) \geq 0$ always)
for $-\infty < y < 0$

Lemma Let Y be a nonnegative R.V.

$$E[Y] = \int_0^\infty P\{Y>y\}\,dy$$

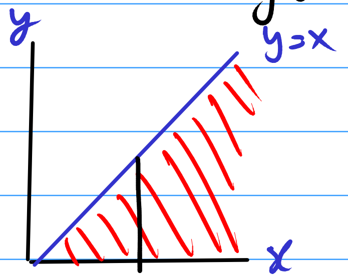
Remark $P\{Y>y\} = 1 - P\{Y \leq y\} = 1 - F(y)$
(CDF of Y)

Proof $$P\{Y>y\} = \int_y^\infty f(x)\,dx$$

$$\int_0^\infty P\{Y>y\}\,dy = \int_0^\infty \int_y^\infty f(x)\,dx\,dy$$

(iterated integral / double integral)
(read from inside out)
(x then y)

reverse order of integration (y then x)
Draw the region over which we are integrating



$$\int_0^\infty \int_0^x f(x)\,dy\,dx$$

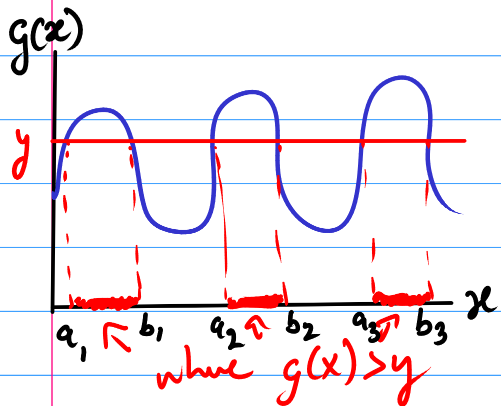
$$\int_0^\infty \int_0^x f(x)\,dy\,dx = \int_0^\infty \Big[y f(x)\Big]_0^x dx$$

$$= \int_0^\infty x f(x)\,dx = \int_0^\infty y f(y)\,dy = E[Y] \text{ by definition } \blacksquare$$

<u>Theorem</u> If g is a non negative function of x
Then $E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

<u>Proof</u> $E[g(x)] = \int_0^\infty P\{g(X) > y\}\,dy$ by Lemma

$$P\{g(X)>y\} = \int_{x \text{ such that } g(x)>y} f(x)\,dx$$

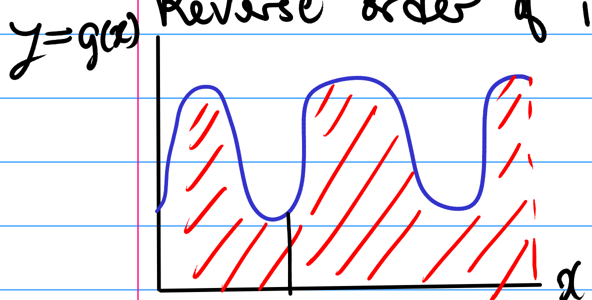


$$P\{g(X)>y\} = \int_{a_1}^{b_1} f(x)dx + \int_{a_2}^{b_2} f(x)dx + \int_{a_3}^{b_3} f(x)dx$$

$$= \int_{\{x \mid g(x)>y\}} f(x)\,dx$$

$$E[g(X)] = \int_0^\infty \int_{\{x \mid g(x)>y\}} f(x)\,dx\,dy$$

Reverse order of integration



$$\int_{-\infty}^{\infty} \int_0^{g(x)} f(x)\,dy\,dx$$

$$\int_{-\infty}^{\infty}\int_0^{g(x)} f(x)\,dy\,dx = \int_{-\infty}^{\infty}\Big[y f(x)\Big]_{y=0}^{y=g(x)} dx$$

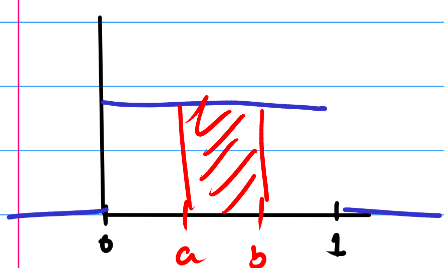
$$= \int_{-\infty}^{\infty} g(x) f(x)\,dx \quad \text{what we wanted.}$$

Uniform Random Variable

(Continuous analog of "equally likely outcomes")

Def Uniform RV on $(0,1)=\{x \mid 0<x<1\}$ has PDF

$$f(x) = \begin{cases} 1 & 0<x<1 \\ 0 & \text{otherwise} \end{cases}$$



$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 1\,dx = 1$$

X = uniform on $(0,1)$

$$P(0<X<1) = 1$$

And if $0<a<b<1$

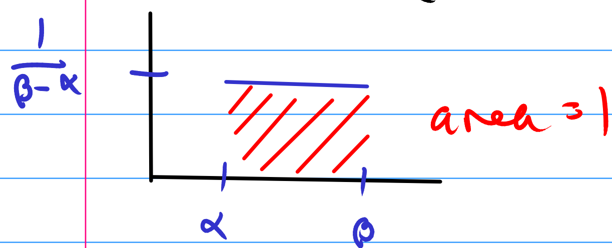
$$P(a\le X<b) = \int_a^b 1\,dx = b-a$$

only depends on length of the interval (a,b) not the position.

X is uniform on (α, β) $\alpha < \beta$

if X has PDF

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$



Uniform and Normal Random Variables

Note: Theoretical Exercise 5.2, NOT 5.1 is on Homework

* Q-drop deadline is next Monday, April 2.

X is uniform on (α, β) if it has PDF

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$

Cumulative distribution function CDF

$$F(a) = \int_{-\infty}^{a} f(x)\,dx$$

if $a \leq \alpha$ $\quad \int_{-\infty}^{a} f(x)\,dx = \int_{-\infty}^{a} 0\,dx = 0$

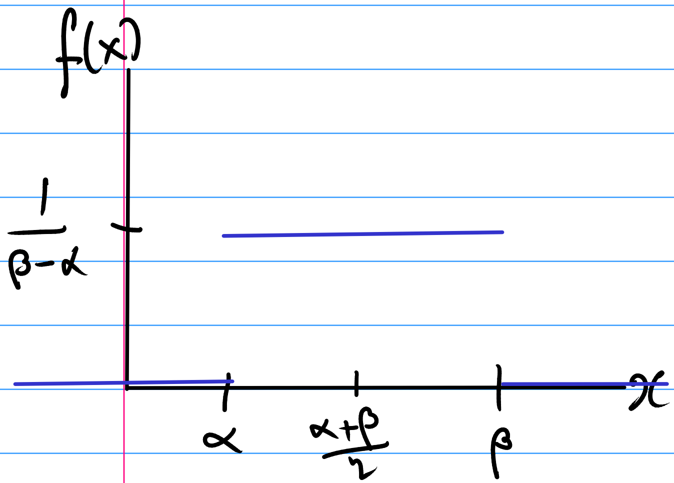
If $\alpha < a < \beta$

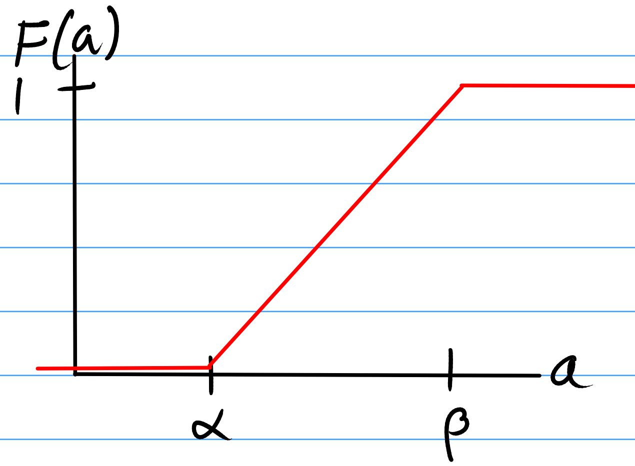
$$\int_{-\infty}^{a} f(x)\,dx = \int_{\alpha}^{a} f(x)\,dx = \int_{\alpha}^{a} \frac{1}{\beta-\alpha}\,dx$$

$$= \left[\frac{1}{\beta-\alpha} x\right]_{x=\alpha}^{x=a} = \frac{a-\alpha}{\beta-\alpha}$$

If $\beta \leq a$ $\quad \int_{-\infty}^{a} f(x)\,dx = \int_{\alpha}^{\beta} \frac{1}{\beta-\alpha}\,dx = \frac{\beta-\alpha}{\beta-\alpha} = 1$

$$F(a) = \begin{cases} 0 & a \le \alpha \\ \dfrac{a-\alpha}{\beta-\alpha} & \alpha < a < \beta \\ 1 & \beta \le a \end{cases}$$





$$E[X] = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{\alpha}^{\beta} x \frac{1}{\beta-\alpha}\,dx$$

$$= \frac{1}{\beta-\alpha}\left[\frac{x^2}{2}\right]_{\alpha}^{\beta} = \frac{1}{\beta-\alpha}\,\frac{1}{2}(\beta^2-\alpha^2)$$

$$= \frac{1}{\cancel{\beta-\alpha}}\,\frac{1}{2}\,\cancel{(\beta-\alpha)}(\beta+\alpha) = \frac{1}{2}(\beta+\alpha)$$

$=$ mid point of (α, β)

Varince: $E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_{\alpha}^{\beta} x^2 \frac{1}{\beta-\alpha}\,dx$

$$= \frac{1}{\beta-\alpha}\left[\frac{x^3}{3}\right]_{\alpha}^{\beta} = \frac{1}{\beta-\alpha}\,\frac{1}{3}(\beta^3-\alpha^3)$$

Fact $(\beta^3-\alpha^3) = (\beta-\alpha)(\beta^2+\alpha\beta+\alpha^2)$

$$E[X^2] = \frac{1}{3}(\beta^2+\alpha\beta+\alpha^2)$$

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

$$= \frac{1}{3}(\beta^2+\alpha\beta+\alpha^2) - \left(\frac{1}{2}(\beta+\alpha)\right)^2$$

$$= \frac{1}{3}(\beta^2+\alpha\beta+\alpha^2) - \frac{1}{4}(\beta^2+2\alpha\beta+\alpha^2)$$

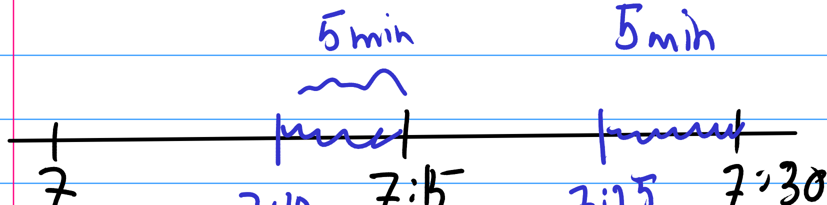
$$= \frac{1}{12}(\beta^2-2\alpha\beta+\alpha^2) = \frac{1}{12}(\beta-\alpha)^2$$

Observe $\beta-\alpha$ = width of (α,β)

<u>Example</u> Busses arrive at a stop at 7, 7:15, 7:30, ...

Passenger arrives at a time uniformly distributed in the interval (7:00, 7:30)

P(wait less than 5 minutes)



5+5 minutes out of 30 minutes

$$P = \frac{10}{30} = \frac{1}{3}$$

Normal Random Variable

a very "universal" distribution

→ approximates binomial distribution
(DeMoivre - Laplace)

→ model errors in scientific observations (Gauss)

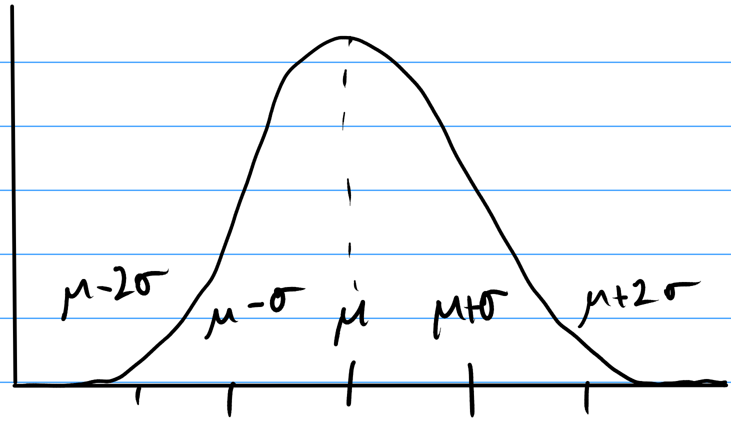
→ Models any sum of independent identically distributed RVs with finite mean and variance

(Central Limit Theorem)

Normal Random Variable with mean μ & variance σ^2
(standard deviation σ)

has PDF

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} \qquad (-\infty < x < \infty)$$



Bell curve

Why is $\dfrac{1}{\sqrt{2\pi}\,\sigma}\displaystyle\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\,dx = 1$?

$$\| $$

sub $y = \dfrac{x-\mu}{\sigma}$ $\qquad \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty} e^{-y^2/2}\,dy$

$dy = \dfrac{dx}{\sigma}$

Need to show $I = \displaystyle\int_{-\infty}^{\infty} e^{-y^2/2}\,dy = \sqrt{2\pi}$

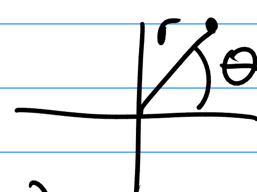
Look at $I^2 = \left(\displaystyle\int_{-\infty}^{\infty} e^{-y^2/2}\,dy\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2}\,dy\right)$

$$= \left(\int_{-\infty}^{\infty} e^{-y^2/2}\,dy\right)\left(\int_{-\infty}^{\infty} e^{-x^2/2}\,dx\right)$$

shuffle
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-y^2/2}\,e^{-x^2/2}\,dy\,dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,dy\,dx$$

change to polar coordinates



$r^2 = x^2 + y^2 \qquad \theta = \tan^{-1}\left(\dfrac{y}{x}\right)$

$dy\,dx = r\,dr\,d\theta$

$$= \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2}\, r\, dr\, d\theta$$

$$\int_0^{\infty} r e^{-r^2/2}\, dr \qquad u = r^2/2 \qquad du = r\,dr$$

$$= \int_0^{\infty} e^{-u}\, du = \left[-e^{-u}\right]_{u=0}^{u=\infty} = 0 - (-e^{-0}) = 1$$

$$I^2 = \int_0^{2\pi} 1\, d\theta = 2\pi \Rightarrow I = \sqrt{2\pi} \qquad \blacksquare$$

<u>Q:</u> Why we just find the antiderivative of $e^{-y^2/2}$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$$

then by FTC, $\frac{d}{dx}\Phi = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

<u>Fact:</u> $\Phi(x)$ cannot be expressed in terms of "elementary functions" (Liouville)

elementary functions

$\frac{x^3+x^2+2x+5}{3x+\sqrt{x}}$ algebraic function

e^x, $\ln x$, $\sin x$, $\arctan x$ elementary transcendental functions

$\Phi(x)$, $\Gamma(x)$, $\zeta(s)$, $J_n(x)$, ...

special functions

To deal with $\Phi(x)$, either

(a) lengthy computations by hand)

b) computer / calculator

c) Table of values (p. 201)

<u>Standard</u> normal Random variable

= Normal RV with mean 0 and variance 1

PDF $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

CDF $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} \, dy$

Normal variables and DeMoivre-Laplace Limit theorem

X normal R.V. with mean μ variance σ^2
has PDF

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

Z standard normal R.V. (has mean 0, variance 1)
has PDF

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

CDF of Z is called Φ

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$$

This is a special function

Q: What about CDF of X w/ μ, σ

$$F_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-(x-\mu)^2/2\sigma^2} dx$$

A: $F_X(x)$ may be computed in terms of Φ

For any random variable X with

$$E[X] = \mu \quad \text{and} \quad Var(X) = \sigma^2$$

Define $Z = \frac{X-\mu}{\sigma}$

$E[Z] = 0$

$$E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{E[X]-\mu}{\sigma} = 0$$

$Var(Z) = 1$

$$Var(Z) = E[Z^2] = E\left[\frac{(X-\mu)^2}{\sigma^2}\right] = \frac{1}{\sigma^2} E[(X-\mu)^2]$$

$$= \frac{1}{\sigma^2} Var(X) = 1$$

Conversely if Z has $E[Z]=0$ and $Var(Z)=1$

the $X = \sigma Z + \mu$ is a random variable

with $E[X] = \mu \quad Var(X) = \sigma^2$

If X is any random variable then

$Z = \frac{X-\mu}{\sigma}$ will be called the standardization of X.

If Z is a standard normal random variable (with mean 0 and variance 1)

then $X = \sigma Z + \mu$ is a normal random variable with mean μ and variance σ^2

Conversely: if X is normal with mean μ and variance σ^2, then

$Z = \frac{X-\mu}{\sigma}$ is normal mean 0 variance 1

Proof Z normal $\Rightarrow$ X normal

CDF of $X = F_X(x) = P\{X \leq x\}$

$$= P\{\sigma Z + \mu \leq x\}$$

$$= P\{Z \leq \frac{x-\mu}{\sigma}\}$$

$$= F_Z\left(\frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

So $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ (very useful)

Density function of X $f_X(x)$

$$f_X(x) = \frac{d}{dx} F_X(x) = \Phi'\left(\frac{x-\mu}{\sigma}\right) \cdot \frac{1}{\sigma}$$

$$\Phi'(y) = f_Z(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2/2} \cdot \frac{1}{\sigma}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

So X is normal with mean μ and variance σ^2

Computing probabilities for normal variables with $\Phi(x)$:

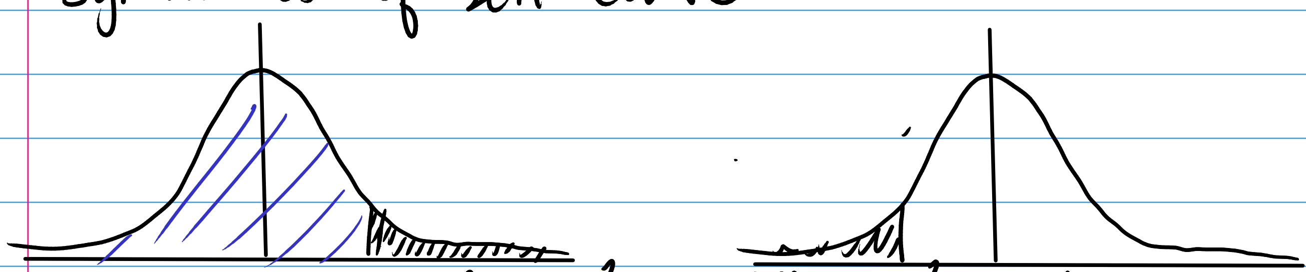
For any variable X with PDF f_X, CDF F_X

$$P\{a \leq X \leq b\} = \int_a^b f_X(x)\,dx = F_X(b) - F_X(a)$$

Z standard normal R.V.

$$P\{a \leq Z \leq b\} = \Phi(b) - \Phi(a)$$

Symmetries of bell curve:



$\Phi(x)$ $\times$ $P\{Z > x\} = P\{Z < -x\} = \Phi(-x)$

$P\{Z > x\} = 1 - \Phi(x)$

Also $\Phi(-x) = 1 - \Phi(x)$

(Table of values on p. 201 has $\Phi(x)$ for $x > 0$. Use this relation to find $\Phi(x)$ for $x < 0$)

Ex X normal w/ $\mu = 3$, $\sigma^2 = 9$, $\sigma = 3$

Q $P\{2 < X < 5\} = P\left\{\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right\}$

$= P\left\{-\frac{1}{3} < Z < \frac{2}{3}\right\}$

$= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right)$

$= \Phi\left(\frac{2}{3}\right) - \left[1 - \Phi\left(\frac{1}{3}\right)\right]$

$= \Phi\left(\frac{2}{3}\right) + \Phi\left(\frac{1}{3}\right) - 1 = .7454 + .6293 - 1$

$= .3779$

Q: if X is normal w/ μ, σ

$P\{\mu < X < \mu + \sigma\} = P\left\{\frac{\mu - \mu}{\sigma} < Z < \frac{\mu + \sigma - \mu}{\sigma}\right\}$

$= P\{0 < Z < 1\} = \Phi(1) - \Phi(0)$

$= .3413$

DeMoivre-Laplace limit theorem

(Approximating binomial RV. by Normal R.V.)

S_n = binomial w/ parameters n, p

(S_n = # of success in n trials)

P(success) = p.

Mean = $E[S_n] = np$ $Var(S_n) = np(1-p)$

Standardize

$$\frac{S_n - np}{\sqrt{np(1-p)}}$$ has mean 0 and verimce 1

DeMoive & Laplace say

$$\frac{S_n - np}{\sqrt{np(1-p)}} \sim \text{standard normal } (\mu=0 \ \sigma=1) \quad \text{as } n \to \infty$$

if n large

$$P\left\{a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right\} \approx P\{a \le Z \le b\} = \Phi(b) - \Phi(a)$$

As a rule approximatn is good of $np(1-p) \geq 10$

Example Fair $(p = .5)$ coin is flipped 40

$X = \#$ heads

Use normal approx to compute $P(X=20)$

Wrong: $P\{X=20\} = P\left\{\frac{X-20}{\sqrt{10}} = \frac{20-20}{\sqrt{10}}\right) = P(Z=0)$

$= \Phi(0) - \Phi(0) = 0$!

CONTINUITY = use half-integers in inequality

$P(X=20) = P\{19.5 < X < 20.5\}$

$= P\left\{\frac{19.5-20}{\sqrt{10}} < Z < \frac{20.5-20}{\sqrt{10}}\right\}$

$= P\{-.16 < Z < .16\}$

$\approx .1272$

exact $\binom{40}{20}\left(\frac{1}{2}\right)^{40} = .1254$

Exponential & Gamma Random Variables

HW Problems 5.32, 5.34, 5.39
Theoretical 5.13, 5.30, 5.31

Ch6 Problems 6.1, 6.7, 6.9

Sequel to lecture on Poisson process

X = exponential R.V. with rate parameter λ
has density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(So X is a nonnegative R.V.)

Cumulative distribution function

$$F_X(a) = \int_0^a \lambda e^{-\lambda x}\,dx = \left[-e^{-\lambda x}\right]_0^a$$

$$= 1 - e^{-\lambda a} \quad (\text{if } a \geq 0)$$

and $F_X(a) = 0$ if $a < 0$

Compute $E[X^n] = \int_{-\infty}^{\infty} x^n f(x)\,dx$

$$= \int_0^\infty x^n f(x)\,dx = \int_0^\infty x^n \lambda e^{-\lambda x}\,dx$$

integration by parts $\quad u = x^n \quad du = nx^{n-1}$

$dv = \lambda e^{-\lambda x} \quad v = -e^{-\lambda x}$

$$\int_0^\infty x^n \lambda e^{-\lambda x} dx = \left[x^n (-e^{-\lambda x}) \right]_0^\infty - \int_0^\infty n x^{n-1} (-e^{-\lambda x}) dx$$

$$= (0 - 0) + \int_0^\infty n x^{n-1} e^{-\lambda x} dx$$

$$= \frac{n}{\lambda} \int_0^\infty x^{n-1} \lambda e^{-\lambda x} dx$$

$$= \frac{n}{\lambda} E[X^{n-1}]$$

$$\therefore \quad E[X^n] = \frac{n}{\lambda} E[X^{n-1}]$$

$$E[X] = E[X^1] = \frac{1}{\lambda} E[X^0] = \frac{1}{\lambda}$$

$$E[X^2] = \frac{2}{\lambda} E[X] = \frac{2}{\lambda} \frac{1}{\lambda} = \frac{2}{\lambda^2}$$

$$Var(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Recall Poisson process: events are happening randomly, but with average rate λ.

$N(t)$ = # of events in interval of time of length t.

$$P\{N(t) = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad \text{Poisson dist.}$$

X = time we have to wait until the first event happens.

X is then an exponential random variable with rate parameter λ.

Proof $P\{X > t\} = P\{N(t) = 0\} = e^{-\lambda t}$

$$F_X(t) = P\{X \le t\} = 1 - P\{X > t\} = 1 - e^{-\lambda t}$$

$F_X(t) = 1 - e^{-\lambda t}$ Therefore X is exponential.

Ex Suppose open store, X = time for first customer to arrive. X is exponential with $\lambda = 10 \frac{\text{customers}}{\text{hour}}$.

$$P\{X > .2 \text{ hours}\} = e^{-\lambda(.2 \text{ hours})} = e^{-2}$$

$$E[X] = \frac{1}{\lambda} = \frac{1}{10} \text{ hour.}$$

Special property of exponential dist:
MEMORYLESSNESS

$$P\{X > s+t \mid X > t\} = P\{X > s\}$$

$$\frac{P\{X > s+t\}}{P\{X > t\}} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = \frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda t}} = e^{-\lambda s} = P\{X > s\}$$

(Exp dist is only continuous distribution with this property)

(geometric R.V. is discrete distribution with this property)

If X is time to wait for an event, it doesn't matter how long we already waited:

<u>Ex</u> Waiting to be served by customer service

$$E[X] = 5 \text{ min} = \frac{1}{\lambda} \quad \lambda = .2 \text{ per minute}$$

$$P\{X > 5 \text{ min}\} = e^{-.2 \cdot 5} = e^{-1} = .368$$

$$P\{X > 10 \text{ min} \mid X > 5 \text{ min}\} = P\{X > 5 \text{ min}\} = e^{-1}$$

$$P\{X > 65 \text{ min} \mid X > 60 \text{ min}\} = P\{X > 5 \text{ min}\} = e^{-1}$$

→ It doesn't matter when we start waiting

→ In Poisson time between two events or time from any point until next event is also exponentially distributed with same rate λ

Analogy

Bernoulli	Poisson
Binomial	Poisson
geometric	exponential
negative binomial	Gamma distribution

One thing Gamma distribution models is the amount of time we must wait for k events to occur in the poisson process.

Y_k = Gamma RV with parameters λ, k

$$f_{Y_k}(t) = \begin{cases} \lambda e^{-\lambda t} \dfrac{(\lambda t)^{k-1}}{(k-1)!} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

We can extend the definition to allow k (integer) to become α (real number)

Y_α = Gamma RV w/ paramters $\lambda, \alpha > 0$

$$f_{Y_\alpha}(x) = \begin{cases} \lambda e^{-\lambda x} \dfrac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Since $\int_0^\infty f_{Y_\alpha}(x)dx = 1 \Rightarrow \Gamma(\alpha) = \underbrace{\int_0^\infty e^{-x} x^{\alpha-1} dx}_{\text{This is definition of } \Gamma\text{-function}}$

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1), \quad \Gamma(1) = 1$$

$$\Gamma(k+1) = k!$$

Gamma dist with $\lambda = \frac{1}{2}$, $\alpha = \frac{n}{2}$

this is called χ_n^2 - distribution

Distribution of a function of a R.V. and joint distributions.

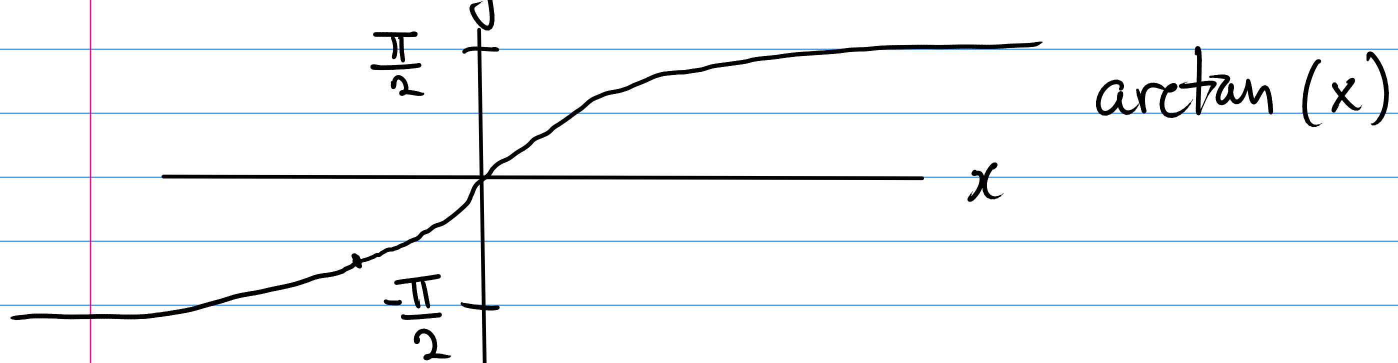
Suppose X and Y are continuous random variables

Suppose $Y = g(X)$ $\quad$ $g(x)$ is a function

Relationship between density/distribution functions of X and Y.

Special Case g is an increasing function

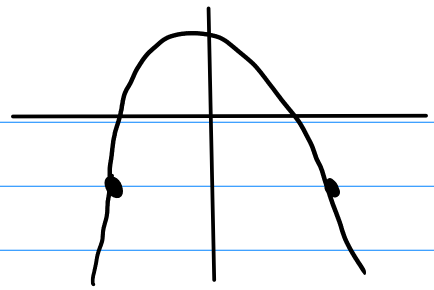
Assume $g'(x) > 0$ for all x



Two properties range of g is some interval (a,b)
a could be $-\infty$
b could be $+\infty$

for each $y \in (a,b)$, there is a unique x such that

$$g(x) = y \iff g^{-1}(y) = x$$



impossible b/c $g'(x) > 0$

$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\}$

if y is in range of $g = (a, b)$

$= P\{X \leq g^{-1}(y)\}$ inequality preserved b/c g increasing.

$= F_X(g^{-1}(y))$

if $y < a$ so is below range of g

$F_Y(y) = P\{g(X) \leq y\} = 0$

if $y > b$ so y is above range of g

$F_Y(y) = P\{g(X) \leq y\} = 1$

$$F_Y(y) = \begin{cases} 0 & y \leq a \\ F_X(g^{-1}(y)) & a < y < b \\ 1 & b \leq y \end{cases}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

$$(\text{if } a<y<b) = \frac{d}{dy} F_X(g^{-1}(y)) = F_X'(g^{-1}(y))(g^{-1})'(y)$$

$$= f_X(g^{-1}(y))(g^{-1})'(y)$$

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))(g^{-1})'(y) & a<y<b \\ 0 & \text{otherwise} \end{cases}$$

$Y = \arctan(X)$ $\qquad (a,b) = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$

Range

$g = \arctan$

$g^{-1} = \tan \qquad (g^{-1})' = \sec^2$

$$f_Y(y) = f_X(\tan(y)) \sec^2 y \qquad -\frac{\pi}{2} < y < \frac{\pi}{2}$$

This is probability analog of change-of-variables in an integral (u substitution)

Probability with 2 random variables

X, Y 2 random variables

Joint distribution function

$$F(a,b) = P\{X \le a, Y \le b\}$$

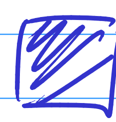
encodes info about probabilities of X, of Y and of events defined in terms of X and Y.

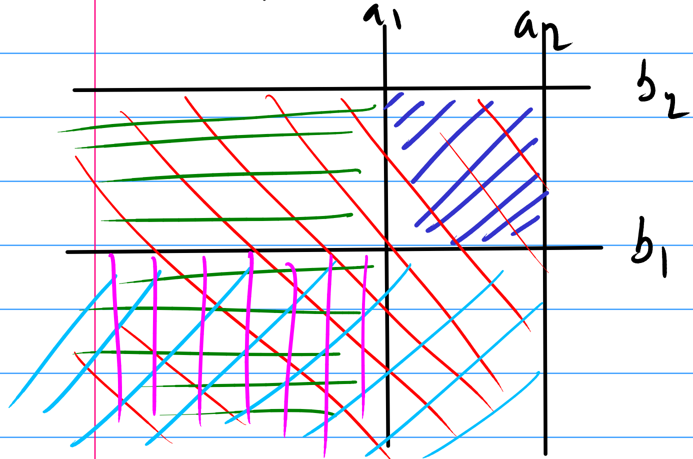
$$F_X(a) = P\{X \le a\} = P\{X \le a, Y < \infty\}$$

$$= \lim_{b \to \infty} F(a,b)$$

$$\equiv F(a, \infty)$$

$$F_Y(b) = \lim_{a \to \infty} F(a,b) \equiv F(\infty, b)$$

$P\{a_1 < X \le a_2, \; b_1 < Y \le b_2\} =$ 



$F(a_2, b_2)$

$-F(a_1, b_2)$

$-F(a_2, b_1)$

$+F(a_1, b_1)$

For X and Y discrete R.V.'s
joint probability mass function

$$p(x,y) = P\{X=x, Y=y\}$$

C = some set of pairs of possible values

$$P\{(X,Y) \in C\} = \sum_{(x,y)\in C} p(x,y)$$

$$p_X(x) = P\{X=x\} = P\{X=x, Y=\text{anything}\}$$

$$= \sum_{\substack{y \\ \text{possible} \\ \text{values of } Y}} p(x,y)$$

$$p_Y(y) = \sum_x p(x,y)$$

$$F(a,b) = P\{X \le a, Y \le b\} = \sum_{x \le a} \sum_{y \le b} p(x,y)$$

Roll two 4-sided dice

X = first die (1, 2, 3, 4)

Y = second die (1, 2, 3, 4)

Z = sum (2, 3, 4, 5, 6, 7, 8)

Joint PMF of X and Y

$x \backslash y$	1	2	3	4	p_X
1	1/16	1/16	1/16	1/16	1/4
2	1/16	1/16	1/16	1/16	1/4
3	1/16	1/16	1/16	1/16	1/4
4	1/16	1/16	1/16	1/16	1/4
p_Y	1/4	1/4	1/4	1/4	

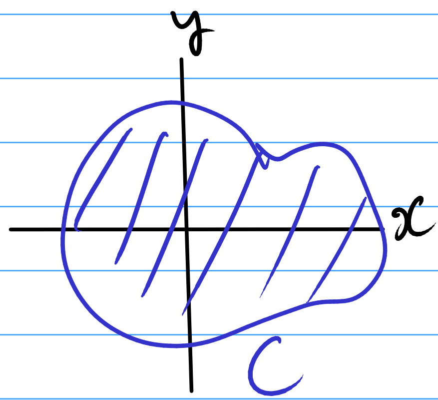
Joint PMF of X and Z

$x \backslash z$	2	3	4	5	6	7	8	p_X
1	1/16	1/16	1/16	1/16	0	0	0	1/4
2	0	1/16	1/16	1/16	1/16	0	0	1/4
3	0	0	1/16	1/16	1/16	1/16	0	1/4
4	0	0	0	1/16	1/16	1/16	1/16	1/4
p_Z	1/16	2/16	3/16	4/16	3/16	2/16	1/16	

Continuous case: X and Y are jointly continuous if there is a density function

$f(x,y)$ such that

$$P\{(X,Y)\in C\} = \iint_C f(x,y)\,dx\,dy$$



In particular $P\{a_1 < X \le a_2, b_1 < Y \le b_2\} = \int_{b_1}^{b_2}\int_{a_1}^{a_2} f(x,y)\,dx\,dy$

$$F(a,b) = \int_{-\infty}^{b}\int_{-\infty}^{a} f(x,y)\,dx\,dy$$

conversely $f(a,b) = \dfrac{\partial^2}{\partial a\,\partial b} F(a,b)$

Independent Random Variables

Recall : given a pair X, Y of random vars.
consider joint distribution

$$F(a,b) = P\{X \le a, Y \le b\}$$

Discrete case: $p(x,y) = P\{X=x, Y=y\}$

Continuous: $f(x,y)$

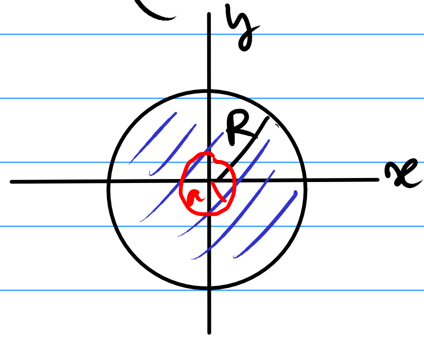
$$P\{(X,Y) \in C\} = \iint_C f(x,y)\,dx\,dy$$

Example of jointly continuous R.V.s

Define joint density function of X and Y to

$$f(x,y) = \begin{cases} c & \text{if } x^2+y^2 \le R^2 \\ 0 & \text{otherwise} \end{cases}$$

Picture



f is nonzero on disk of radius R.

find c: $1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy$

$$= \iint_{\{x^2+y^2 \le R^2\}} c\,dx\,dy + \cancel{\iint_{\{x^2+y^2 > R^2\}} 0\,dx\,dy}$$

$$= c \iint_{\{x^2+y^2 \le R^2\}} 1\,dx\,dy = \int_{-R}^{R}\int_{-\sqrt{R^2-y^2}}^{+\sqrt{R^2-y^2}} 1\,dx\,dy$$

$$= c\,\text{Area}\left(\{x^2+y^2 \le R^2\}\right) = c\pi R^2 = 1$$

$$c = \frac{1}{\pi R^2}$$

Q. What is prob. that a randomly chosen point lies at a distance $\le a$ from the center?

$$P\{X^2+Y^2 \le a^2\} = c \iint_{\{x^2+y^2 \le a^2\}} 1\,dx\,dy = c\,\text{Area}(\ )$$

$$= c \cdot \pi a^2 = \frac{\pi a^2}{\pi R^2} = \frac{a^2}{R^2}$$

Note added: This is only true if $a \le R$

if $a > R$, then $P\{X^2+Y^2 \le a^2\} = 1$

Recall A and B are independent events

if $P(A \cap B) = P(A)P(B)$

X and Y are independent random variables

if, for any two sets C and D of real numbers

$P((X \in C) \text{ and } (Y \in D)) = P(X \in C) \cdot P(Y \in D)$

i.e., $X \in C$ and $Y \in D$ are <u>independent</u> <u>events</u>.

If X and Y are independent

- $P\{a_1 < X \leq a_2, b_1 < Y \leq b_2\}$
  $= P\{a_1 < X \leq a_2\} \cdot P\{b_1 < Y \leq b_2\}$
- $P\{X \leq a, Y \leq b\} = P\{X \leq a\} \cdot P\{Y \leq b\}$

i.e. $F_{X,Y}(a, b) = F_X(a) \cdot F_Y(b)$

In fact X and Y are independent if

$F_{X,Y}(a,b)$ is the product of $F_X(a)$ and $F_Y(b)$

- In discrete case : joint prob. mass function

$p(x,y) = P\{X=x, Y=y\} = P\{X=x\} \cdot P\{Y=y\}$

$$(*) \quad p(x,y) = p_X(x)\, p_Y(y)$$

In fact $(*)$ is equivalent to X and Y being indep.

$$P\{X \in C, Y \in D\} = \sum_{y \in D} \sum_{x \in C} p(x,y)$$

$$= \sum_{y \in D} \sum_{x \in C} p_X(x)\, p_Y(y) \qquad \text{(assumption.)}$$

$$= \sum_{y \in D} \left(p_Y(y) \sum_{x \in C} p_X(x) \right)$$

$$= \left(\sum_{x \in C} p_X(x) \right) \left(\sum_{y \in D} p_Y(y) \right)$$

$$= P\{X \in C\} \cdot P\{Y \in D\}$$

Thus X and Y are independent $\boxtimes$

For Continuous Random Variables X, Y

X and Y are independent $\iff f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$

$\iff f_{X,Y}(x,y) = h(x)\, g(y)$ for some functions $h(x), g(y)$

Discrete example: $n+m$ Bernoulli trials w/ prob p of success

X = # success in first n trials

Y = # success in last m trials

X and Y are independent

$$P\{X=i, Y=j\} = P\{X=i\}P\{Y=j\}$$
$$= \binom{n}{i} p^i (1-p)^{n-i} \cdot \binom{m}{j} p^j (1-p)^{m-j}$$

Z = # of successes in all $n+m$ trials $(Z=X+Y)$

X and Z are not independent

$$P\{X=i, Z=i+j\} = P\{X=i, Y=j\} = P\{X=i\}P\{Y=j\}$$

$$P\{X=i\} \cdot P\{Z=i+j\} \neq P\{X=i\}P\{Y=j\}$$

Continuous example Alice and Bob are to meet at certain location

each arrives at a time between 12 and 1 uniformly in this interval, and independently of each other.

X = time in minutes after 12 that Alice arrives

Y = time in minutes " " " Bob "

Q Prob. that Alice arrives first?

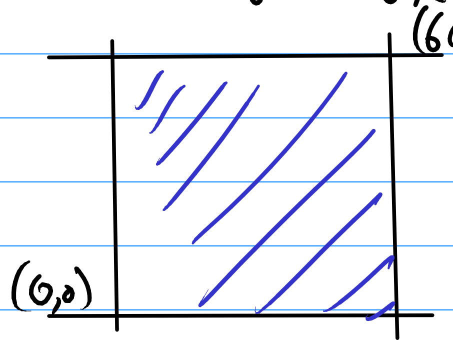
A: $P\{X < Y\} = \frac{1}{2}$ (intuitive by symmetry)

$$f_X(x) = \begin{cases} \frac{1}{60} & 0 < x < 60 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1}{60} & 0 < y < 60 \\ 0 & \text{otherwise} \end{cases}$$

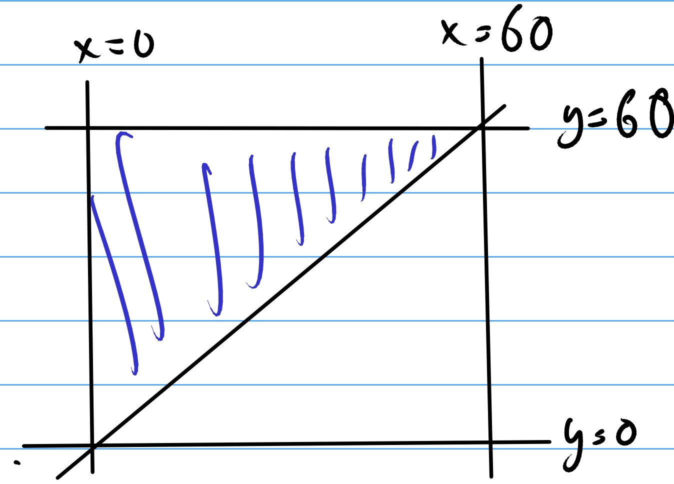
By independence

$$f(x,y) = f_X(x) f_Y(y) = \begin{cases} \frac{1}{(60)^2} & 0 < x < 60 \text{ and } 0 < y < 60 \\ 0 & \text{otherwise} \end{cases}$$



$$P\{X<Y\} = \iint_{\{x<y\}} f(x,y)\,dx\,dy$$

$$= \iint_{\left\{\begin{array}{c} 0<x<60 \\ 0<y<60 \\ x<y \end{array}\right\}} \frac{1}{(60)^2}\,dx\,dy$$



$$= \int_0^{60}\int_0^{y} \frac{1}{(60)^2}\,dx\,dy$$

$$= \frac{1}{(60)^2}\int_0^{60} y\,dy = \frac{1}{(60)^2}\,\frac{1}{2}\,(60)^2 = \frac{1}{2}$$

Next HW

Problems 6.18, 6.20, 6.29, 6.30, 6.32, 6.38, 6.42
Theoretical
Exercises: 6.11, 6.14

Sums of independent Random variables

Suppose X and Y are independent random variables
Suppose X and Y are continuous

density functions $f_X(x)$, $f_Y(y)$, $f_{X,Y}(x,y)$ joint

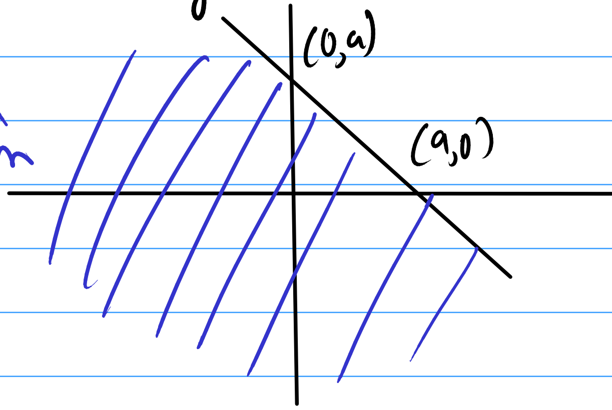
$$X \text{ and } Y \text{ independent} \iff f_{X,Y}(x,y) = f_X(x)\,f_Y(y)$$

Now consider $X+Y$: find distribution & density

$$F_{X+Y}(a) = P\{X+Y \le a\}$$

$$= \iint_{\{x+y \le a\}} f_{X,Y}(x,y)\,dx\,dy$$

Region of integration



$x+y \le a$

$x \le a-y$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_{X,Y}(x,y)\,dx\,dy$$ So far, true for any R.V.s X, Y

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) f_Y(y)\,dx\,dy$$ using the fact that X and Y are independent

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{a-y} f_X(x)\,dx \right] f_Y(y)\,dy$$

$$= \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\,dy$$

$$\therefore \quad F_{X+Y}(a) = \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\,dy$$

$$f_{X+Y}(a) = \frac{d}{da} F_{X+Y} = \int_{-\infty}^{\infty} \frac{d}{da} F_X(a-y) f_Y(y)\,dy$$

$$= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y)\,dy$$

Definition If f and g are two functions The convolution of f and g is new function

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x-y) g(y)\,dy$$

Properties

(1) $f * g = g * f$

(2) $(f * g)' = (f' * g) = (f * g')$

To summarize if X and Y are independent R.V.s

then $F_{X+Y} = F_X * f_Y = f_X * F_Y$

$f_{X+Y} = f_X * f_Y$

Discrete case: integral replaced by sum

$$F_{X+Y}(a) = \sum_y F_X(a-y)\, p_Y(y)$$

$$= \sum_x p_X(x)\, F_Y(a-x)$$

$$p_{X+Y}(a) = \sum_y p_X(a-y)\, p_Y(y)$$

<u>Example</u> · Suppose X_1 is normal with mean μ_1 and variance σ_1^2

· Suppose X_2 is normal with mean μ_2 and variance σ_2^2

· Suppose X_1 and X_2 are independent

THEN X_1+X_2 is normal with mean $\mu_1+\mu_2$ and variance $\sigma^2 = \sigma_1^2+\sigma_2^2$

ie. $f_{X_1+X_2} = f_{X_1} * f_{X_2}$ we're saying

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{(a-y-\mu_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}\, dy$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} e^{-\frac{(a-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}}$$

Straightforward but lengthy computation

<u>Example</u> Basketball team $P(\text{home win}) = .6$
$P(\text{away win}) = .5$

play 41 home and 41 away.

$X = \#$ home wins $\sim$ binomial with $n=41, p=.6$

$Y = \#$ away wins $\sim$ binomial with $n=41, p=.5$

X and Y are independent.

By DeMoivre-Laplace limit theorem

$X \sim$ normal w/ $\mu = np = 24.6$
$\sigma^2 = np(1-p) = 9.84$

$Y \sim$ normal w/ $\mu = np = 20.5$
$\sigma^2 = np(1-p) = 10.25$

Since "normal + normal = normal"

$X+Y \sim$ normal w/ $\mu = 24.6 + 20.5 = 45.1$
$\sigma^2 = 20.09$

P(50 or more wins in the entire season)

$= P(50 \le X+Y) = P(49.5 < X+Y)$

$$= P\left(\frac{49.5-45.1}{\sqrt{20.09}} < Z\right) = 1 - \Phi\left(\frac{49.5-45.1}{\sqrt{20.09}}\right)$$

X and Y independent

X and Y are both exponential with rate λ

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x>0 \\ 0 \end{cases} \qquad f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & y>0 \\ 0 \end{cases}$$

X and Y are "independent identically distributed"

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y)\,dy$$

in order for $f_X(a-y) f_Y(y)$ to be nonzero

we need $\quad a-y>0 \Rightarrow y<a$
$\qquad\qquad\quad y>0$

$$= \int_0^a \lambda e^{-\lambda(a-y)} \lambda e^{-\lambda y}\,dy$$

$$= \lambda^2 e^{-\lambda a} \int_0^a \overset{1}{\cancel{e^{+\lambda y} e^{-\lambda y}}}\,dy$$

$$= \lambda^2 e^{-\lambda a} a = \lambda e^{-\lambda a} \frac{(\lambda a)^1}{\Gamma(2)}$$

Observe that X+Y is a Gamma variable with rate λ and parameter $\alpha = 2$

$$\left(\begin{matrix} \text{Gamma} \\ \lambda, \alpha \end{matrix} \quad f(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \right)$$

Observe that exponential RV. = Gamma random variable with $\alpha = 1$

in general if X is Gamma w/ λ, α

Y is Gamma w/ λ, β

and X and Y are independent

then $X+Y$ is Gamma w/ $\lambda, \alpha+\beta$

Conceptual explanation: if α and β are integers

then X = Gamma w/ λ, α = time to wait for α events in Poisson process w/ rate λ

Y = Gamma w/ λ, β = time to wait for β events

$X+Y$ = time to wait for $\alpha+\beta$ events

= Gamma w/ $\lambda, \alpha+\beta$

Example of sum of independent discrete R.V.s

Exam 3 next friday
as before one sheet of notes is allowed.

If X and Y are independent R.V.s and discrete

$$p_{X,Y}(x,y) = p_X(x)\,p_Y(y)$$

$$p_{X+Y}(a) = \sum_y p_X(a-y)\,p_Y(y)$$

Example X poisson R.V. with parameter λ_1
Y poisson R.V. with parameter λ_2

Assume X and Y are independent: find p_{X+Y}

$$p_X(k) = e^{-\lambda_1}\frac{\lambda_1^k}{k!} \qquad p_Y(k) = e^{-\lambda_2}\frac{\lambda_2^k}{k!}$$

$$p_{X+Y}(n) = \sum_k p_X(n-k)\,p_Y(k)$$

$$= \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^{n-k}}{(n-k)!}\,e^{-\lambda_2}\frac{\lambda_2^k}{k!}$$

$p_X(n-k) > 0 \Rightarrow k \le n$

$p_Y(k) > 0 \Rightarrow 0 \le k$

$$= e^{-(\lambda_1+\lambda_2)} \sum_k \frac{\lambda_1^{n-k}}{(n-k)!} \frac{\lambda_2^{k}}{k!}$$

$$\left.\begin{aligned} (\lambda_1+\lambda_2)^n &= \sum_k \binom{n}{k} \lambda_1^{n-k} \lambda_2^{k} \\ &= \sum_k \frac{n!}{(n-k)!\,k!} \lambda_1^{n-k} \lambda_2^{k} \\ &= n! \sum_k \frac{\lambda_1^{n-k}}{(n-k)!} \frac{\lambda_2^{k}}{k!} \end{aligned}\right\} \text{Binomial Theorem}$$

$$= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!}$$

Conclusion: $X+Y$ has a Poisson distribution with parameter $\lambda_1+\lambda_2$

<u>Next</u> Conditional distribution:
(what happens when variables are not independent)

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

For discrete R.Vs X and Y

$$P\{X=x \mid Y=y\} = \frac{P\{X=x \text{ and } Y=y\}}{P\{Y=y\}}$$

We have a conditional probability mass function

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \quad \left(= \frac{\text{joint PMF}}{\text{PMF of } Y} \right)$$

Condition distribution function

$$F_{X|Y}(a|y) = P\{X \le a \mid Y = y\} = \sum_{x \le a} p_{X|Y}(a|y)$$

X and Y are independent $\iff$ $p_{X|Y}(x|y) = p_X(x)$

$$\left(\text{because} \quad p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)} = \frac{p_X(x)\, p_Y(y)}{p_Y(y)} = p_X(x) \right)$$

<u>Example</u>
- X is Poisson w/ parameter λ_1
- Y is Poisson w/ parameter λ_2

- Assume X and Y are independent
- We saw X+Y is Poisson w/ parameter $\lambda_1 + \lambda_2$

Compute conditional mass function of X given X+Y

$$P\{X = k \mid X+Y = n\} = \frac{P\{X = k,\ X+Y = n\}}{P\{X+Y = n\}}$$

$$= \frac{P\{X=k, Y=n-k\}}{P\{X+Y=n\}} = \frac{P\{X=k\}P\{Y=n-k\}}{P\{X+Y=n\}}$$

$$= \frac{\left(\cancel{e^{-\lambda_1}} \frac{\lambda_1^k}{k!}\right)\left(\cancel{e^{-\lambda_2}} \frac{\lambda_2^{n-k}}{(n-k)!}\right)}{\left[\cancel{e^{-(\lambda_1+\lambda_2)}} \frac{(\lambda_1+\lambda_2)^n}{n!}\right]}$$

$$= \frac{n!}{k!(n-k)!} \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n} = \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

$$\left(p = \frac{\lambda_1}{\lambda_1+\lambda_2} \qquad 1-p = \frac{\lambda_2}{\lambda_1+\lambda_2} \right)$$

$$= \binom{n}{k} p^k (1-p)^{n-k}$$

Binomial w/ parameters n and $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$

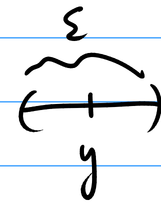
Continuous case: X and Y jointly continuous R.V.s

What if we try $P\{X \le a \mid Y=y\} = \frac{P\{X \le a \text{ and } Y=y\}}{P\{Y=y\}}$

Then have $P\{Y=y\} = \int_y^y f_Y(t)\,dt = 0$ we would divide by 0!

Can't divide by $P\{Y=y\} \Rightarrow$ need different approach

Condition on $y-\frac{\varepsilon}{2} < Y < y+\frac{\varepsilon}{2}$



$$P\{y-\tfrac{\varepsilon}{2} < Y < y+\tfrac{\varepsilon}{2}\} = \int_{y-\frac{\varepsilon}{2}}^{y+\frac{\varepsilon}{2}} f_Y(t)\,dt \approx f_Y(y)\cdot\varepsilon$$

for small ε

$$P\{a<X<b \mid y-\tfrac{\varepsilon}{2} < Y < y+\tfrac{\varepsilon}{2}\}$$

$$= \frac{P\{a<X<b \text{ and } y-\frac{\varepsilon}{2} < Y < y+\frac{\varepsilon}{2}\}}{P\{y-\frac{\varepsilon}{2} < Y < y+\frac{\varepsilon}{2}\}}$$

$$= \frac{\int_a^b \int_{y-\frac{\varepsilon}{2}}^{y+\frac{\varepsilon}{2}} f(x,t)\,dt\,dx}{\int_{y-\frac{\varepsilon}{2}}^{y+\frac{\varepsilon}{2}} f_Y(t)\,dt} \approx \frac{\int_a^b f(x,y)\,\varepsilon\,dx}{f_Y(y)\,\varepsilon}$$

$$= \int_a^b \left(\frac{f(x,y)}{f_Y(y)}\right) dx$$

The conditional density function of X given Y is $f_{X|Y}(x|y) = \dfrac{f(x,y)}{f_Y(y)}$

so $P\{a<X<b \mid Y=y\} = \int_a^b f_{X|Y}(x|y)\,dx$

Ex $f(x,y) = \begin{cases} \dfrac{e^{-x/y}e^{-y}}{y} & x>0 \text{ and } y>0 \\ 0 & \text{Otherwise} \end{cases}$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^{\infty} \frac{1}{y} e^{-x/y} e^{-y}\,dx$$

$$= e^{-y}$$

$$f_{X|Y}(x|y) = \frac{e^{-x/y}e^{-y}}{y} \Big/ e^{-y} = \frac{e^{-x/y}}{y}$$

So X given Y is exponentially distributed with parameter $\lambda = \frac{1}{y}$.

Expectation of sums of Random Variables

$Z = g(X,Y)$ What $E[Z] = E[g(X,Y)]$

Lemma: If X and Y are discrete with joint mass function $p(x,y)$

Then $E[g(X,Y)] = \sum_y \sum_x g(x,y)\, p(x,y)$

If X and Y are continuous with joint density function $f(x,y)$

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\, dx\, dy$$

Proof is similar to the single variable case

In case $g \geq 0$ is nonnegative

$$E[g(X,Y)] = \int_0^{\infty} P\{g(X,Y) > t\}\, dt$$

$$= \int_0^{\infty} \iint_{g(x,y)>0} f(x,y)\, dx\, dy\, dt = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^{g(x,y)} f(x,y)\, dt\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\, dx\, dy$$

Corollary: $E[X+Y] = E[X] + E[Y]$

Already proved in discrete case

In continuous case, we have joint density function $f(x,y)$

$$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)\,dx\,dy$$

Recall: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$

$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$

$$= \int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} f(x,y)\,dy\right]dx + \int_{-\infty}^{\infty} y\left[\int_{-\infty}^{\infty} f(x,y)\,dx\right]dy$$

$$= \int_{-\infty}^{\infty} x f_X(x)\,dx + \int_{-\infty}^{\infty} y f_Y(y)\,dy$$

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

Does not require independence.

<u>Corollary</u> if $X \geq Y$ (always)

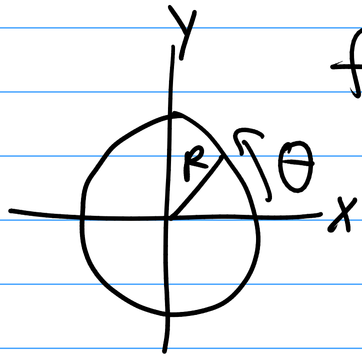
then $E[X] \geq E[Y]$

<u>Proof</u> $X - Y \geq 0$ always

$$E[X-Y] \geq 0 \Rightarrow E[X] - E[Y] \geq 0$$

e.g. $\int (x-y) f(x,y)\, dx\, dy \geq 0$ or $\sum_{x,y} (x-y) p(x,y) \geq 0$

<u>Example</u>: uniform distribution of a disk of radius R



$$f(x,y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } x^2 + y^2 \leq R^2 \\ 0 & \text{otherwise} \end{cases}$$

What is the expected distance from the point (X,Y) to the center $(0,0)$?

$$D = \sqrt{X^2 + Y^2}$$

$$E[D] = \iint \sqrt{x^2+y^2}\, f(x,y)\, dx\, dy$$

Polar coordinates r, θ $\quad r = \sqrt{x^2+y^2}$

$$dx\, dy = r\, dr\, d\theta$$

$$E[D] = \iint r\, f(x,y)\, r\, dr\, d\theta$$

$$= \int_0^{2\pi}\int_0^R r^2 \frac{1}{\pi R^2}\, dr\, d\theta = \frac{1}{\pi R^2}(2\pi)\frac{R^3}{3}$$

$$= \frac{2}{3}R$$

Sometimes its useful to undrstand a given R.V. as a sum of simpler ones

X binomial w/ parameters n, p ← prob. of success
trials ↗

let $X_i = \begin{cases} 1 & \text{if } i\text{th trial is a success} \\ 0 & \text{otherwise} \end{cases}$

Then $X = X_1 + X_2 + \cdots + X_n$

Note X_i is a "0 or 1" Random Variable

$$E[X_i] = 0\, P\{X_i = 0\} + 1\, P\{X_i = 1\} = P\{X_i = 1\}$$

For "0 or 1" RV $E[X] = P\{X=1\}$

$$E[X] = E[X_1] + \cdots + E[X_n]$$

$$= p + p + \cdots + p = np$$

Recall Hypergeometric distribution

N balls in urn m are white $N-m$ are black

Select n balls without replacement

$X = \#$ of white balls selected

$$P\{X = i\} = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

Suppose we number the m white balls $1, \dots, m$

Define $X_i = \begin{cases} 1 & \text{if the } i\text{th white ball get selected} \\ 0 & \end{cases}$

$$X = X_1 + X_2 + \cdots + X_m$$

$E[X_i] = P\{X_i = 1\} = P\{i\text{th white ball selected}\}$

$$= \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

$$E[X] = \sum_{i=1}^{m} \frac{n}{N} = \frac{mn}{N}$$

$$Y_i = \begin{cases} 1 & \text{if } i\text{th ball selected is white} \\ 0 & \end{cases}$$

$$X = Y_1 + \cdots + Y_n$$

$$E[Y_i] = \frac{m}{N}$$

because when we consider all possible outcomes, ith ball is equally likely to be any of the N balls

$$E[X] = \frac{nm}{N}$$

(same as if selection is done w/ replacement)

Next Homework

Problems 7.5, 7.9, 7.11

Theoretical exercises: 7.4, 7.5

Chapter 5 continuous random variables

distribution function $F_X(a)$

density function $f_X(x)$

$$P\{a < X < b\} = \int_a^b f_X(x)\,dx = F_X(b) - F_X(a)$$

Expectation and variance

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx$$

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

Types of R.V.s Uniform

Normal

Exponential

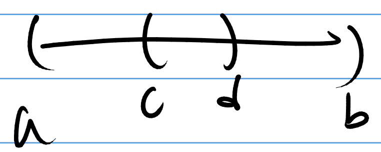
Gamma

What to use them for

Uniform – when different values are equally likely

$$f(x) = \begin{cases} \frac{1}{b-a} & a<x<b \\ 0 & \text{otherwise} \end{cases}$$

X uniform on (a,b)



$$\frac{d-c}{b-a} = P(c<X<d)$$

Normal with mean μ variance σ^2
std deviation σ

density and distribution are complicated transcendental function.

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} \qquad \Phi(x)$$

→ Always try to transform into a standard normal RV. Z normal with $\mu=0$, $\sigma=1$

X has mean μ variance σ^2 $\Leftrightarrow$ $Z = \frac{X-\mu}{\sigma}$ has mean 0 variance 1

$$\Phi(a) = P\{Z \leq a\}$$

X = IQ test $\overset{\text{normal}}{\checkmark}$ has mean $\mu = 100$, $\sigma = 15$

$P\{X > 130\}$ $\qquad$ $130 = 100 + 2(15)$

$= P\{X > \mu + 2\sigma\} = P\{Z > 2\} = 1 - P\{Z < 2\}$

$= 1 - \Phi(2) = 1 - .9772 = .0228 \approx 2\%$

Can approximate Binomial distribution n, p

$\approx$ Normal with $\mu = np$, $\sigma^2 = np(1-p)$

if n is large $(\sigma^2 > 10)$

(Binomial to Poisson is for when p is small, n large)

Exponential $\quad f_X(x) \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{other} \end{cases}$

Exponential models time to wait for something to happens (in a Poisson process)

$\rightarrow$ Memorylessness $\quad P\{X > s+t \mid X > t\} = P\{X > s\}$

Gamma dist $\Leftrightarrow$ time to wait for α events to happen.

$$f_X(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \qquad \underline{\text{If } \alpha \text{ integer}} \quad \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}}{(\alpha-1)!}$$

Function of RV

$g(x)$ function

increasing $g'(x) > 0$

thm $Y = g(X)$

$$f_Y(y) = f_X(g^{-1}(y))\,(g^{-1})'(y)$$

$$F_Y(y) = F_X(g^{-1}(y))$$

valid iff y is in the range of g

Chapter 6 multiple RV. joint density/mass functions

(X, Y) joint continuous

$$P\{a < X < b,\ c < Y < d\} = \int_c^d \int_a^b f(x,y)\,dx\,dy$$

(Discrete) joint mass function $p(x,y) = P\{X = x, Y = y\}$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

$$p_X(x) = \sum_y p(x,y) \qquad p_Y(y) = \sum_x p(x,y)$$

X = roll one die 1 - 6

Y = roll another die 1 - 6

$$p(x,y) = P\{X=x, Y=y\} = \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36} \quad \text{if } 1\le x\le 6,\ 1\le y\le 6$$

$$p_X(4) = p(4,1) + p(4,2) + p(4,3) + p(4,4) + p(4,5) + p(4,6)$$

$$= \sum_{y=1}^{6} p(4,y) = \underbrace{\frac{1}{36} + \cdots + \frac{1}{36}}_{6} = 6\cdot\frac{1}{36} = \frac{1}{6}$$

$$P\{X^2+Y^2 < 10\} = \iint_{\{x^2+y^2<10\}} f(x,y)\,dx\,dy$$

joint density function lets us express independence of R.V.s

$$f(x,y) = f_X(x)\, f_Y(y)$$

or $p(x,y) = p_X(x)\, p_Y(y)$

Conditional prob $P\{X=x \mid Y=y\} = \dfrac{p(x,y)}{p_Y(y)}$ discrete

Conditional prob. mass func. $= p_{X|Y}(x|y)$

Continuous

Conditional density func. $f(x|y) = \frac{f(x,y)}{f_Y(y)}$

X and Y indep if $\begin{cases} p_{X|Y}(x|y) = p_X(x) \\ f(x|y) = f_X(x) \end{cases}$

Take sums of independent R.V.s

discrete $P\{X+Y=a\} = \sum_y P\{X=a-y, Y=y\}$

$= \sum_y P\{X=a-y\} P\{Y=y\}$

$p_{X+Y}(a) = \sum_y p_X(a-y)\, p_Y(y)$

Continuous $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y)\, dy$

- $\text{Normal}(\mu_1, \sigma_1^2) + \text{Normal}(\mu_2, \sigma_2^2) = \text{Normal}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$
  if independent

- Poisson λ_1 + Poisson λ_2 = Poisson $\lambda_1+\lambda_2$

$\boxed{\text{exponent + exponent} = \text{Gamma}}$

Exam Stats

$X = P1 + P2 + P3 + P4$ First 4 problems

$X' = 80 - \frac{1}{2}(80 - X)$ curve for first 4

$Y = \frac{100}{80} X'$ out of 100

$Z = Y + P5$

mean	86.1
1Q	75.6
2Q	94.25
3Q	96.2

Reminder: Course Instructor survey

Due: 5/4

Last time If X and Y are R.V.s.

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)\, dx\, dy$$

$$E[X+Y] = E[X] + E[Y]$$ (does not require independence)

Lemma
If X and Y are independent then

$$E[g(X)h(Y)] = E[g(X)]\cdot E[h(Y)]$$

Pf. $E[g(X)h(Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)\,dx\,dy$

$= \iint g(x)h(y)f_X(x)f_Y(y)\,dx\,dy$ ← uses independence

$= \int g(x)f_X(x)\left[\int h(y)f_Y(y)\,dy\right]dx$

$= \left[\int g(x)f_X(x)\,dx\right]\left[\int h(y)f_Y(y)\,dy\right]$

$= E[g(X)]\cdot E[h(Y)]$

Cor $E[XY] = E[X]E[Y]$

PROVIDED that X and Y are independent

Let X and Y be random varriable, not assumed to be independent:

Notation $\mu_X = E[X]$ $\mu_Y = E[Y]$

$\sigma_X^2 = Var(X)$ $\sigma_Y^2 = Var(Y)$

<u>Def</u> Covariance:

$$\mathrm{Cov}(X,Y) := E[(X-\mu_X)(Y-\mu_Y)]$$
$$= E[(X-E[X])(Y-E[Y])]$$

<u>Prop</u>: $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$

$$E[(X-\mu_X)(Y-\mu_Y)] = E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y]$$
$$= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y$$
$$= E[XY] - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y$$
$$= E[XY] - \mu_X \mu_Y = E[XY] - E[X]E[Y]$$

<u>Properties</u> (I) X, Y independent $\Rightarrow$ $\mathrm{Cov}(X,Y) = 0$

(I') $\mathrm{Cov}(X,Y) \neq 0 \Rightarrow X$ and Y are dependent.

(II) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ symmetric

(III) $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$

(IV) $\mathrm{Cov}(aX,Y) = a\,\mathrm{Cov}(X,Y)$

(V) $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$

(IV') $\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right)$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \mathrm{Cov}(X_i, Y_j)$$

Variance of a sum of random variables

$$\mathrm{Var}(X_1+X_2) = \mathrm{Cov}(X_1+X_2, X_1+X_2)$$

$$= \mathrm{Cov}(X_1, X_1+X_2) + \mathrm{Cov}(X_2, X_1+X_2)$$

$$= \mathrm{Cov}(X_1,X_1) + \mathrm{Cov}(X_1,X_2) + \mathrm{Cov}(X_2,X_1) + \mathrm{Cov}(X_2,X_2)$$

$$= \mathrm{Cov}(X_1,X_1) + \mathrm{Cov}(X_2,X_2) + 2\,\mathrm{Cov}(X_1,X_2)$$

$$= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\,\mathrm{Cov}(X_1,X_2)$$

Summarize $\mathrm{Var}(X_1+X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\,\mathrm{Cov}(X_1,X_2)$

In general

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} \mathrm{Cov}(X_i, X_j)$$

Cor If $X_1, X_2, \ldots, X_n$ are independent

$$\mathrm{Cov}(X_i,X_j) = 0 \Rightarrow \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Caution

$\mathrm{Cov}(X,Y)=0$ does not necessarily imply that X and Y are independent

Ex X has values $-1, 0, 1$ each w/ probability $\frac{1}{3}$

$$E[X] = -\frac{1}{3} + 0\frac{1}{3} + 1\cdot\frac{1}{3} = 0$$

$$Y = \begin{cases} 0 & \text{if } X \neq 0 \\ 1 & \text{if } X = 0 \end{cases}$$

So X and Y are not independent

BUT $XY \equiv 0$ $\quad E[XY] - E[X]E[Y] = 0 - 0\,E[Y] = 0$

Just mention correllation of X and Y

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

FACT $-1 \leq \rho(X,Y) \leq 1$

$\rho(X,Y) = 1 \Rightarrow$ perfect linear relationship between X and Y

$$Y = aX + b \text{ with } a > 0$$

$\rho(X,Y) = -1 \Rightarrow$ perfect liner relationship $Y = aX + b$ with $a < 0$.

Last HW

Ch7: Problems: 7.32, 7.36, 7.38
Theoretical: 7.19

Ch8: Problems 8.1, 8.4, 8.5, 8.7
Theoretical: 8.1

Today: Start Limit Theorems

Let X be a random variable with some distribution
(which may be unknown)

X represents the out come of some experiment,

We can run several independent trials of this experiment

Generates a sequence of random variables $X_1, X_2, X_3, \ldots$

X_i = value on the ith trial

All random variables X_i are "essentially the same as X"

$X, X_1, X_2, X_3, \ldots$ all have same distribution

$X_1, X_2, X_3, \ldots$ are independent identically distributed

Particular values of $\overbrace{X_1, X_2, X_3, \dots}^{\text{samples}}$ are like a dataset we need to analyze.

How to find the mean $\mu = E[X]$?

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} \quad \text{sample mean}$$

Why good: $E[\overline{X}] = E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$

each X_i has $E[X_i] = \mu$

$$= \frac{1}{n} n\mu = \mu$$

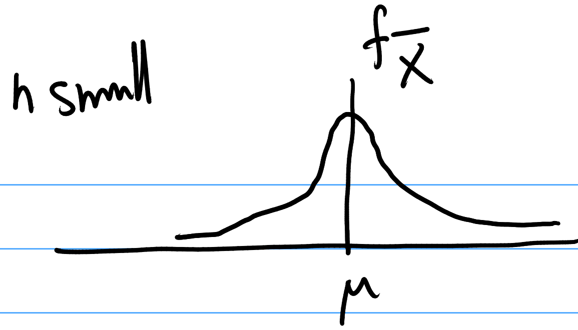
Q: How lose is $\overline{X}$ to being μ?
Law of Large numbers gives an answer.

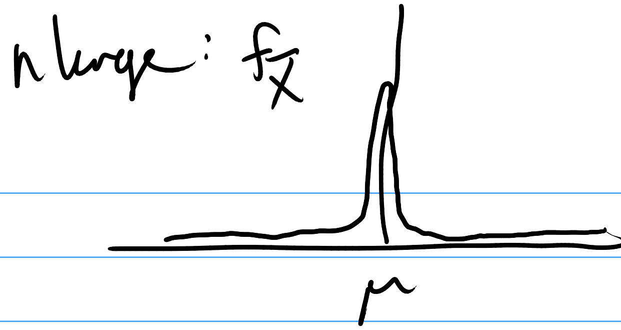
$$Var(\overline{X}) = \sum_{i=1}^{n} Var\left(\frac{X_i}{n}\right) = \sum_{i=1}^{n} \frac{1}{n^2} Var(X_i)$$

$$= \sum_{i=1}^{n} \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}$$

$\boxed{\sigma^2 = Var(X)}$ Variance of sample mean goes to 0 as $n \to \infty$

If you take $n = \#$ of samples to be very large, the it becomes very likely that $\overline{X}$ is close to μ





Weak Law of Large numbers is a precise statement of this.

<u>Theorem</u> (Weak Law of Large number)

For any value of $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{ \left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| \geq \varepsilon \right\} = 0$$

$$\lim_{n\to\infty} P\left\{ \left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| < \varepsilon \right\} = 1$$

Even more quantitatively:

$$P\left\{ \left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| \geq \varepsilon \right\} \leq \frac{\sigma^2}{n\varepsilon^2}$$

X has $\sigma^2 = 25$

Q: How large does n have to be?

$$P\left\{ \left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| \geq .1 \right\} \leq .05$$

$$P\left\{ \left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| < .1 \right\} \geq .95$$

i.e. 95% confidence that sample mean is within .1 of the true value.

Need $\frac{\sigma^2}{n\varepsilon^2} \leq .05$

$$\frac{25}{n \cdot (.1)^2} \leq .05 \iff \frac{2500}{n} \leq .05 \iff n \geq 2500 \cdot 20 = 50000$$

Proof involves two lemmas

<u>Lemma</u> (Markov's inequality) Suppose X is a NONNEGATIVE Random variable

(i.e. $P\{X<0\}=0$)

<u>Then</u>: $P\{X \geq a\} \leq \frac{E[X]}{a}$ (true for any $a>0$)

<u>Proof</u> let $I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$

$$E[I] = P\{X \geq a\}$$

Also: $\frac{X}{a} \geq I$: indeed if $X \geq a$ $\frac{X}{a} \geq 1 = I$

if $X<a$ $\frac{X}{a} \geq 0 = I$

(because $X \geq 0$)

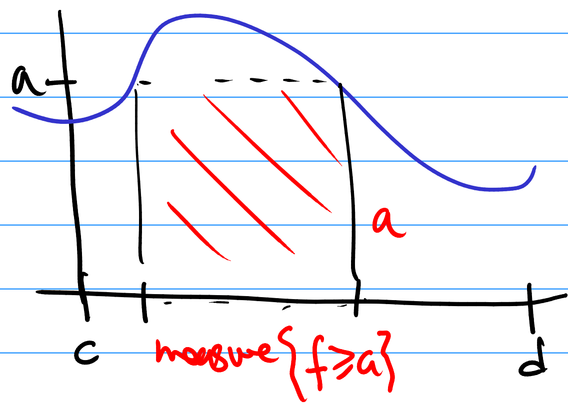
therefore $E\left[\frac{X}{a}\right] \geq E[I]$

$$\frac{E[X]}{a} \geq E[I] = P\{X \geq a\}$$

OPTIONAL

Analogy in calculus: $f(x)$ is a nonnegative function

then $\text{measure}\{f \geq a\} \leq \dfrac{\int_c^d f\,dx}{a}$



$\int_c^d f\,dx$ = Area under curve

$\geq$ Area of the rectangle

$= a \cdot \text{measure}\{f \geq a\}$

Corollary (Chebyshev's inequality) $(k>0)$

Apply Markov to $(X-\mu)^2$ $\quad E[(X-\mu)^2] = \sigma^2$

$$P\{(X-\mu)^2 \geq k^2\} \leq \frac{\sigma^2}{k^2}$$

use $(X-\mu)^2 \geq k^2 \iff |X-\mu| \geq k$

get
$$\boxed{P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}}$$

Proof of Weak Law of Large numbers:

Apply Chebyshev to $\frac{X_1+\cdots+X_n}{n}$

$$E\left[\frac{X_1+\cdots+X_n}{n}\right]=\mu \qquad \mathrm{Var}\left(\quad\right)=\frac{\sigma^2}{n}$$

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|\geq\varepsilon\right\}\leq\frac{\sigma^2}{n}\frac{1}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2}$$

<u>Example</u> of Markov's inequality

X has mean $=50$ $\quad\sigma^2=25$

$$P\{X\geq 75\}\leq\frac{50}{75}=\frac{2}{3}$$

Central limit theorem

Last time inequalities

Markov: $X \geq 0$ then $P\{X \geq a\} \leq \frac{E[X]}{a}$

Chebyshev: $P\{|X-\mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$

Weak law of large numbers

$X_1, X_2, \ldots,$ independent identically distributed

then $$\lim_{n\to\infty} P\left\{\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| \geq \varepsilon\right\} = 0$$

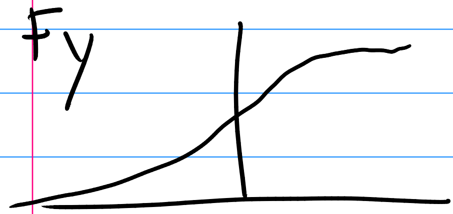
The weak law of large numbers (and the Central limit theorem) are statements about convergence in distribution:

Let Y_n be a sequence of Random variables with CDF's $F_{Y_n}(a)$

If Y is a random variable with $F_Y(a)$

We say the sequence Y_n converges to Y in distribution if

$\lim_{n\to\infty} F_{Y_n}(a) = F_Y(a)$ For every a at which F_Y is continuous



Apply this to weak law of large numbers

$$Y_n = \frac{X_1 + \cdots + X_n}{n} \qquad E[Y_n] = \mu$$

$$\mathrm{Var}(Y_n) = \frac{\sigma^2}{n}$$

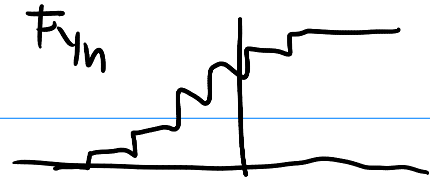
$$\lim_{n\to\infty} P\left\{ \frac{X_1+\cdots+X_n}{n} \le \mu - \varepsilon \right\} = 0$$

$$\lim_{n\to\infty} P\{Y_n \le \mu-\varepsilon\} = 0 \qquad \text{for any } \varepsilon > 0$$

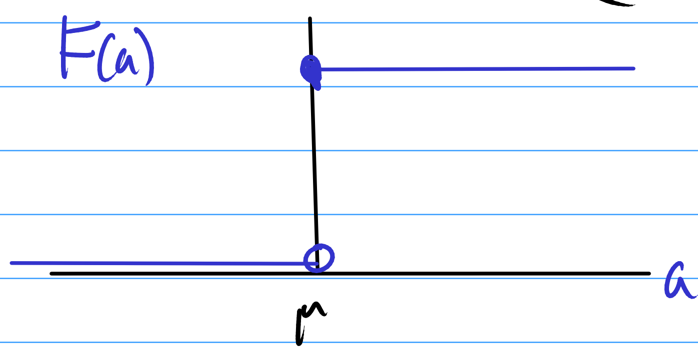
similarly $\displaystyle \lim_{n\to\infty} P\left\{ \frac{X_1+\cdots+X_n}{n} \le \mu+\varepsilon \right\} = 1$

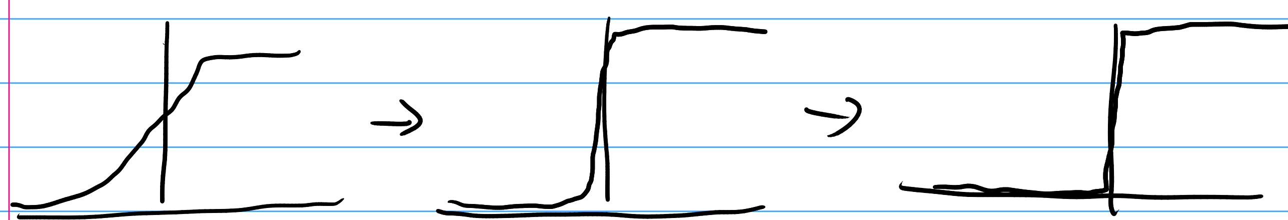
$$\lim_{n\to\infty} P\{Y_n \le \mu+\varepsilon\} = 1$$

$$\lim_{n\to\infty} F_{Y_n}(\mu-\varepsilon) = 0 \qquad \lim_{n\to\infty} F_{Y_n}(\mu+\varepsilon) = 1$$



limiting distribution $F(a) = \begin{cases} 0 & \text{if } a<\mu \\ 1 & \text{if } a\geq\mu \end{cases}$





The step function is the CDF of a constant random variable.

$Y=\mu$ with probability 1

then $F_Y(a) = \begin{cases} 0 & a<\mu \\ 1 & a\geq\mu \end{cases}$

$$Y_n = \frac{X_1+\cdots+X_n}{n} \xrightarrow{\text{distribution}} \mu \quad (\text{constant})$$

The central limit theorem says that Y_n is asymptotic to a sequence of normal random variables.

$$Y_n = \frac{X_1 + \cdots + X_n}{n} \qquad E[Y_n] = \mu \qquad Var(Y_n) = \frac{\sigma^2}{n}$$

$$Z_n = \frac{\left(\frac{X_1 + \cdots + X_n}{n} - \mu\right)}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{(X_1 + \cdots + X_n - n\mu)}{\sigma\sqrt{n}}$$

Z_n is like the average, but it has mean 0 and variance instead.

Let Z be standard normal Random variable

$$F_Z(a) = \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2} dt$$

<u>Theorem</u> (CLT): $Z_n \xrightarrow{\text{distribution}} Z$.

i.e., $$\lim_{n \to \infty} P\left\{ a \leq \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq b \right\} = \Phi(b) - \Phi(a)$$

* Remember, this is all under the assumption
* that $X_1, X_2, \ldots$ are independent and
* have a common distribution w/ finite μ, σ^2

Other ways to phrase it

For n large

$$X_1 + \cdots + X_n \overset{\text{dist.}}{\sim} \text{normal with mean } n\mu \text{ and variance } n\sigma^2$$

$$\frac{X_1 + \cdots + X_n}{n} \overset{\text{dist}}{\sim} \text{normal with mean } \mu \text{ and variance } \sigma^2/n$$

Extension to when variables are not identically distributed, but independent.

$X_1 + \cdots + X_n$ mean $\sum_{i=1}^{n} \mu_i$

variance $\sum_{i=1}^{n} \sigma_i^2$

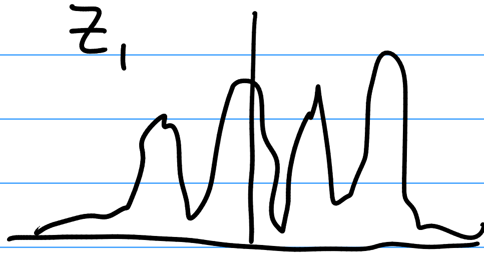
$$\frac{\sum_{i=1}^{n} (X_i - \mu_i)}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \overset{\text{dist}}{\sim} \text{normal with mean 0 and variance 1}$$

(Under some technical hypotheses)

This theorem explains the ubiquity of the normal distribution: Appears whenever we look at a Random variable which is a sum of many small independent contributions.

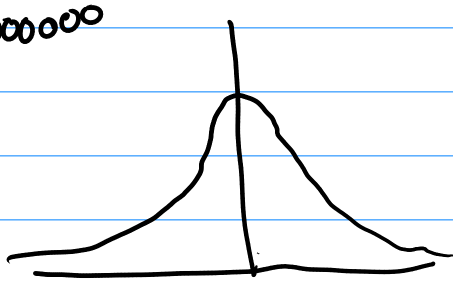
In pictures of PDF

Z_1



Z_2

$Z_{1000000}$



Application of CLT: compute approximate probabilities.

X is binomial with (n, p)

$X = \sum_{i=1}^{n} X_i$ $\quad$ X_i is binomial with $(n=1, p)$

$$E[X_i] = p$$
$$Var(X_i) = p(1-p)$$

Large $n \Rightarrow$

$$\frac{X - np}{\sqrt{np(1-p)}}$$ is approximately standard normal

This is just DeMoivre-Laplace limit theorem

(for this approximation would use the continuity correction!)

X = Poisson with parameter $\lambda = 100$

$$X = \sum_{i=1}^{100} X_i \qquad X_i \text{ is Poisson with parameter } \lambda = 1$$

Mean $(X) = \lambda = 100$ Mean $(X_i) = \lambda = 1$

Var $(X) = \lambda = 100$ Var $(X_i) = \lambda = 1$

So we can approximate X by a normal distribution with $\mu = 100$ $\sigma^2 = 100$

Q: $$P\{X \geq 120\} = e^{-100} \sum_{i=120}^{\infty} \frac{(100)^i}{i!}$$

Because we're approximating a discrete distribution by a continuous one, we use the continuity correction

$$P\{X \geq 120\} = P\{X > 119.5\}$$

$$\overset{CLT}{\approx} P\left\{Z > \frac{119.5 - 100}{\sqrt{100}}\right\} = P\{Z > 1.95\}$$

$$= 1 - \Phi(1.95) = .0256$$

If X is an integer (continuity correction)

$$\begin{cases} X \leq n \iff X < n + \frac{1}{2} \\ X < n \iff X < n - \frac{1}{2} \\ \\ X \geq n \iff X > n - \frac{1}{2} \\ X > n \iff X > n + \frac{1}{2} \end{cases}$$

Final exam Friday May 11, 2-5pm, usual room

→ Two sheets of notes are permitted

→ Office hours

Monday May 7 1-4pm

Wednesday May 9 9:30-12

$X_1, X_2, \dots, X_n, \dots$ a sequence of independent random variables, all having the same distribution

Let $\mu = E[X_i]$ common value of mean.

Weak law of large numbers

$$\lim_{n\to\infty} P\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \varepsilon\right\} = 0$$

One enhancement is the central limit theorem

→ gives the limiting/asymptotic shape of the distribution:

$\boxed{\text{Assume finite variance}}$

$Var(X_i) = \sigma^2$

$Var(\Sigma X_i) = n\sigma^2$

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

is approximately normal with mean 0 and variance 1

Another enhancement of the weak law of large numbers is the strong law of large numbers, which involves a stronger notion of convergence.

What if we look at the sequence

$$A_n = \frac{X_1 + \cdots + X_n}{n}$$

Does this sequence have a limit $\lim_{n\to\infty} A_n$?

Strong law of large numbers says the limit exists and is equal to μ, with probability 1

("with probability 1" =: "almost surely")

X = uniform random variable on $(0,1)$

$$f_X(x) = \begin{cases} 1 & \text{if } 0<x<1 \\ 0 & \text{otherwise} \end{cases}$$

$$P\{X=\tfrac{1}{2}\} = \int_{1/2}^{1/2} 1\,dx = 0$$

Intuitively, "with probability 1" means "certainly"!

Strong law of large numbers.

$$P\left\{\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = \mu\right\} = 1$$

Eg.

n	1	2	3	4	5	
A_n	0	1	0	1	0	...

limit doesn't exist because sequence oscillates. But Strong law of large numbers says that this situation has probability zero.

$$P\left\{\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} \text{ does not exist or exists but does not equal } \mu\right\} = 0$$

Application Let E be some event has probability $P(E)$

$$X_i = \begin{cases} 1 & \text{if } E \text{ occurs on the } i\text{th trial} \\ 0 & \text{if } E \text{ does not occur} \end{cases}$$

$$E[X_i] = P(X_i=1) = P(E)$$

With probability 1: $\displaystyle\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = P(E)$

Weak law $\lim_{n\to\infty} P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| \geq \varepsilon\right\} = 0$

Means if we take n sufficiently large

then the prob that $\left|\frac{X_1+\cdots X_n}{n}-\mu\right| \geq \varepsilon$ is small

But this leaves open the possibility that

$\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| \geq \varepsilon$ for infinitely many n.

The strong law of large numbers rules this out:

with probability 1, there are only

finitely many values of n for which

$\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| \geq \varepsilon$.

ie., there is some n_0 such that for all

$n \geq n_0$, $\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| < \varepsilon$

Proof of Strong law of large numbers

We will assume that $E[X_i^4] = K < \infty$

finite forth moment. [The theorem is true without this assumption.]

Case 1: assume $\mu = E[X_i] = 0$.

$S_n = X_1 + \cdots + X_n$ sequence of sums

Want to show $\lim_{n\to\infty} \frac{S_n}{n} = 0$

$$E[S_n^4] = E[(X_1 + \cdots + X_n)^4]$$

= sum of terms like

$E[X_i^4]$, $E[X_i^3 X_j]$, $E[X_i^2 X_j^2]$

$E[X_i^2 X_j X_k]$, $E[X_i X_j X_k X_l]$ $\quad i \neq j \neq k \neq l$

Most terms go away

$$E[X_i^3 X_j] = E[X_i^3] E[X_j] \text{ by independence}$$

$$= E[X_i^3] \cdot 0 \quad b/c \ \mu = 0.$$

$$E[X_i^2 X_j X_k] = E[X_i^2] E[X_j] E[X_k] = 0$$

$E[X_i X_j X_k X_\ell] = 0$ similarly

Left with $E[X_i^4]$ and $E[X_i^2 X_j^2]$ $i \neq j$

$$\overset{\shortparallel}{E[X_i^2]E[X_j^2]}$$

$E[X_i^4]$ occurs once for each i

$E[X_i^2 X_j^2]$ occurs $\binom{4}{2} = 6$ times for each pair i,j

Assumed $E[X_i^4] = K < \infty$

so $0 \leq Var(X_i^2) = E[X_i^4] - (E[X_i^2])^2 = K - (E[X_i^2])^2$

$E[X_i^2 X_j^2] = E[X_i^2]E[X_j^2] = (E[X_i^2])^2 \leq K$

Thus

$$E[S_n^4] = \sum_{i=1}^{n} E[X_i^4] + 6 \sum_{i \neq j} E[X_i^2 X_j^2]$$

$$\leq nK + 6\binom{n}{2} K$$

$$= nK + 3n(n-1)K$$

$$\leq nK + 3n^2 K$$

$$E\left[\frac{S_n^4}{n^4}\right] \leq \frac{K}{n^3} + \frac{3K}{n^2}$$

$$E\left[\sum_{n=1}^{\infty} \frac{S_n^4}{n^4}\right] = \sum_{n=1}^{\infty} E\left[\frac{S_n^4}{n^4}\right] \le \sum_{n=1}^{\infty} \frac{K}{n^3} + \frac{3K}{n^2} < \infty$$

converges by p-test

$E\left[\sum_{n=1}^{\infty} \frac{S_n^4}{n^4}\right] < \infty$ implies

that $P\left\{ \sum_{n=1}^{\infty} \frac{S_n^4}{n^4} < \infty \right\} = 1$

Finite expectation $\Rightarrow$ almost surely finite

$$\sum_{n=1}^{\infty} \frac{S_n^4}{n^4} < \infty \Rightarrow \lim_{n\to\infty} \frac{S_n^4}{n^4} = 0$$

$\Rightarrow \lim_{n\to\infty} \frac{S_n}{n} = 0$ by nth term test

$$P\left\{ \lim_{n\to\infty} \frac{S_n}{n} = 0 \right\} = 1$$

Case $\mu \neq 0$, just replace X_i by $X_i - \mu$

Moment Generating functions

and proof of the Central Limit Theorem

Recall CLT: if $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of independent identically distributed random variables, with mean μ and variance σ^2

$$Y_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{distribution}} \text{standard normal random variable } Z$$

$$F_{Y_n}(a) \longrightarrow \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2}\, dt$$

Recall: X and Y are independent random variables with density functions f_X and f_Y

then $X+Y$ has density function $f_X * f_Y$

$$f_X * f_Y(a) = \int_{-\infty}^{\infty} f_X(x) f_Y(a-x)\, dx$$

Nice fact about convolution: Fourier transform of convolution is the product of the Fourier transforms.

$f(x) \to$ Fourier transform $\hat{f}(\xi) = \int_{-\infty}^{\infty} e^{-2\pi i \xi x} f(x)\,dx$

$\hat{f}(\xi) \to$ inverse transform $f(x) = \int_{-\infty}^{\infty} e^{2\pi i \xi x} \hat{f}(\xi)\,d\xi$

One version of Fourier transform in probability theory is the moment generating function

Let X be a random variable

The moment generating function of X is

$$M_X(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} P\{X=x\} & \text{discrete} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx & \text{continuous} \end{cases}$$

I.e., for continuous X w/ density $f_X(x)$

$$M_X(-2\pi i \xi) = \hat{f}_X(\xi)$$

Why the name? Moments of distribution

$$n\text{th moment} = E[X^n] = \begin{cases} \sum x^n p(x) \\ \int x^n f(x)\,dx \end{cases}$$

("higher order variance")

$$M_X(t) = E[e^{tX}] = E\left[\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right] = \sum_{n=0}^{\infty} E[X^n] \frac{t^n}{n!}$$

So

$M_X(t)$ is a function whose Taylor coefficients at $(t=0)$ are exactly the moments of X.

If a_n is any sequence of numbers

$$G(t) = \sum_{n=0}^{\infty} a_n \frac{t^n}{n!} = a_0 + a_1 t + a_2 \frac{t^2}{2} + a_3 \frac{t^3}{6} + \cdots$$

is called exponential generating function of the sequence $\{a_n\}_{n=0}^{\infty}$

$a_0 = G(0) \qquad a_1 = G'(0) \qquad a_2 = G''(0)$

in general $\quad a_n = G^{(n)}(0)$

Point is many combinatorial properties of the sequence a_n translate into properties of generating function

e.g. linear recursion relation on a_n's $\Leftrightarrow$ differential equation for the generating function.

Fact: if X and Y are independent R.V.s,

Then $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

<u>Proof</u> $E[e^{t(X+Y)}] = E[e^{tX} e^{tY}] = E[e^{tX}]\, E[e^{tY}]$

because e^{tX} and e^{tY} are independent

Ex Binomial dist (n,p)

$$\sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^{n}\binom{n}{k}(e^tp)^k(1-p)^{n-k}$$

$$\overset{\text{binomial theorem}}{=} \left(e^tp+(1-p)\right)^n = M(t)$$

Exponential dist

$$M(t) = \int_0^\infty e^{tx}\,\lambda e^{-\lambda x}\,dx = \lambda\int_0^\infty e^{-(\lambda-t)x}\,dx$$

$$= \lambda\left[-\frac{1}{\lambda-t}e^{-(\lambda-t)x}\right]_0^\infty = \frac{\lambda}{\lambda-t}$$

$$E[X] = M'(0) = \left.\frac{\lambda}{(\lambda-t)^2}\right|_{t=0} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

$$E[X^2] = M''(0) = \left.\frac{2\lambda}{(\lambda-t)^3}\right|_{t=0} = \frac{2}{\lambda^2}$$

$$\mathrm{Var}(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Z Standard normal variable $f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$M_Z(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

$$e^{tx} e^{-x^2/2} = e^{tx - x^2/2} = e^{t^2/2} e^{-\frac{t^2}{2} + tx - \frac{x^2}{2}} = e^{t^2/2} e^{-\frac{(x-t)^2}{2}}$$

$$e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}}\,dx = e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2}\,du$$

$$\begin{pmatrix} u = x - t \\ du = dx \end{pmatrix}$$

$$= e^{t^2/2}$$

Levy Continuity theorem

If X_n is a sequence of random variables, X som RV.
Then
if $M_{X_n}(t) \to M_X(t)$ for all t

then $X_n \to X$ in distribution
ie. $F_{X_n}(a) \to F_X(a)$ for all a at which $F_X(a)$ is continuous

Proof CLT if X_i has mean 0 and variance 1

$M(t) = M_{X_i}(t)$ the common MGF of the X_i's (since identially distribution)

Now consider

$$Y_n = \frac{X_1 + \cdots + X_n - n0}{1\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}}$$

$$M_{Y_n}(t) = M_{\frac{X_1+\cdots+X_n}{\sqrt{n}}}(t) = M_{\frac{X_1}{\sqrt{n}}}(t) \cdots M_{\frac{X_n}{\sqrt{n}}}(t)$$

$$M_{\frac{X_i}{\sqrt{n}}}(t) = E\left[\exp\left(t\frac{X_i}{\sqrt{n}}\right)\right] = M_{X_i}\left(\frac{t}{\sqrt{n}}\right) = M\left(\frac{t}{\sqrt{n}}\right)$$

$$M_{Y_n}(t) = \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

Need to show $\lim_{n\to\infty} \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = e^{t^2/2}$

$$L(t) := \log M(t)$$

$$\operatorname{Log}\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = n\,L\left(\frac{t}{\sqrt{n}}\right)$$

$$\log e^{t^2/2} = t^2/2$$

Need to show $\lim_{n\to\infty} n L\left(\frac{t}{\sqrt{n}}\right) = \frac{t^2}{2}$

What is the taylor series for $L(t)$

$$L(0) = \log M(0) = \log 1 = 0$$

$$L'(0) = \frac{M'(0)}{M(0)} = \frac{\mu}{1} = \mu = 0$$

$$L''(0) = \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} = \frac{1(\mu^2+\sigma^2) - \mu^2}{1^2}$$

$$= \sigma^2 = 1$$

$$L(t) = 0 + 0t + 1\frac{t^2}{2} + a_3\frac{t^3}{3!} + a_4\frac{t^4}{4!} + \cdots$$

Now consider what happens when $L(t) \to nL\left(\frac{t}{\sqrt{n}}\right)$

$$t^k \longrightarrow n\left(\frac{t}{\sqrt{n}}\right)^k = n^{1-\frac{k}{2}}\, t^k$$

ie. k th term is rescaled by $n^{1-\frac{k}{2}}$

$k=2$ $\quad n^0$

$k=3$ $\quad n^{-1/2}$

$k=4$ $\quad n^{-1}$

as $n \to \infty$ all terms but quadratic go away

$$\lim_{n\to\infty} nL\left(\frac{t}{\sqrt{n}}\right) = \frac{t^2}{2}$$

Final Review

Course Evaluations end Today

Exam 3 hours Friday May 11 2-5pm
in 6.122

2 sheets of notes permitted

~10 problems

Final is cumulative: test all parts of course

30% about chapters 7 8

Summary

<u>Prelude</u> combinatorics ch 1

permutations (some indistinguishable objects)

combinations & binomials coeffs $\binom{n}{k}$

binomial theorem and <u>combinatorials proofs</u>

prove by counting same thing two different ways.

Discrete Probability Ch 2, 3, 4

2) Sample spaces and events. Axioms of probability

S' sample space E event = subset of S'

$P(E)$ $0 \le P(E) \le 1$

$P(S') = 1$

$P(E \cup F) = P(E) + P(F)$

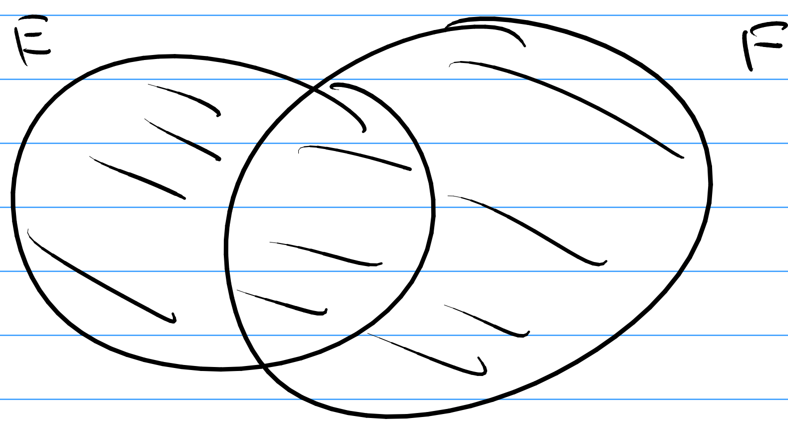
provided that $E \cap F = \emptyset$

Set operations $\cap$, $\cup$ de Morgan's laws, complement

Venn diagrams

Inclusion-exclusion formula

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$



- Examples of discrete probability coming from situations where all outcomes are equally likely.
  → reduces to combinatorics

3) Conditional probability

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

independent events: $P(EF) = P(E)P(F)$

$$P(E|F) = P(E)$$

Two key lemmas:

Conditioning on a complete set of mutually exclusive events

$H_1, H_2, H_3, \ldots, H_n$

(i) $H_i \cap H_j = \emptyset$ unless $i=j$.

(ii) $H_1 \cup H_2 \cup \cdots \cup H_n = S$

$$P(E) = P(E|H_1)P(H_1) + \cdots + P(E|H_n)P(H_n)$$

Trick: "reversing the order of conditioning"

$$P(F|E) = \frac{P(FE)}{P(E)} = \frac{P(E|F)P(F)}{P(E)}$$

Bayes law:

$$P(H_1|E) = \frac{P(E|H_1)P(H_1)}{\sum_{i=1}^{n} P(E|H_i)P(H_i)}$$

H_1 = guilty H_2 = not guilty

prior probability

$$P(\text{guilty}|E) = \frac{P(E|\text{guilty})P(\text{guilty})}{P(E|\text{guilty})P(\text{guilty}) + P(E|\text{not guilty})P(\text{not guilt}}$$

conditional independence

$$P(E_1 E_2|F) = P(E_1|F)P(E_2|F)$$

we say E_1 and E_2 are conditionally independent given F.

4] random variables, mass function, expectation & variance

Gambling examples: Expected winnings determine whether its a good idea to play.

Binomial (n,p), Poisson λ, geometric, negative binomial, hypergeometric

Poisson approximation to binomial p small n large, $\lambda = np$

Continuous RVs Ch 5

density function, Cumulative distribution function.

$P(a<X<b) = \int_a^b f_X(x)\,dx$

$F_X(a) = \int_{-\infty}^{a} f_X(x)\,dx$

expectation and variance

memorylessness property

Uniform, normal, exponential random variables

X normal with mean μ, variance σ^2

$\frac{X-\mu}{\sigma}$ normal with mean 0 and variance 1

Theory of RVs 6, 7, 8

6] joint mass / density functions

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

independent RV's.

X, Y independent if $f(x,y) = f_X(x) f_Y(y)$

$$p(x,y) = p_X(x)\,p_Y(y)$$

Sum of independent random variables

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(a-x)\, dx$$

7] Expectations of functions of random variables

$$E[g(X,Y)] = \iint g(x,y) f(x,y)\, dx\, dy$$

→ sums products $E[X+Y] = E[X] + E[Y]$

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

X, Y independent $\Rightarrow \mathrm{Cov}(X,Y) = 0.$

$\mathrm{Var}(X) = \mathrm{Cov}(X,X)$ ⇒ variances of sums

$$\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{cov}(X_i, Y_j)$$

if $X_1, X_2, \ldots, X_n$ are independent

then $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$

8]: Markov's, Chebyshev's inequality and Weak law of large numbers.

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Central limit theorem

$X_1, \dots, X_n$ independent identically distributed

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{n\to\infty}$$ standard normal distribution with $\mu = 0$, $\sigma = 1$

Weak law

$$P\left\{ \left| \frac{X_1 + \cdots + X_n - n\mu}{n} \right| \geq \varepsilon \right\} \xrightarrow[n\to\infty]{} 0$$

Strong law of large numbers:

$$P\left\{ \lim_{n\to\infty} \frac{X_n + \cdots + X_n}{n} = \mu \right\} = 1$$

Office hours Mon 1-4
Wed 9:30-12

otherwise email.