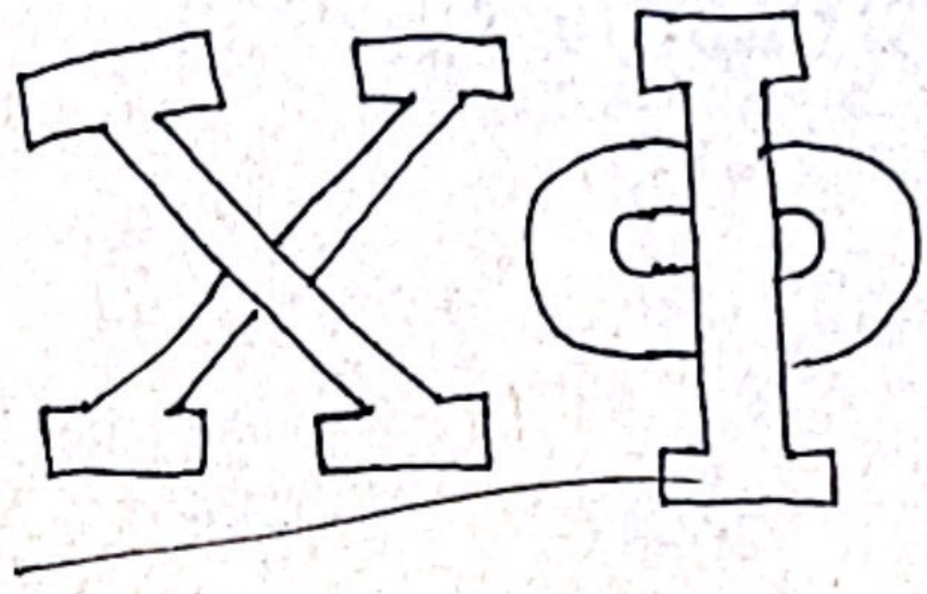


spiralnote

6.5mm RULED

40 SHEETS | NO.

MATH 201: Intro to Probability Notes - University of Rochester

SPRING 2024

N236 252×179mm | 6.5mm RULED | 40 SHEETS

maruman®
www.e-maruman.co.jp

MPS-N 70g/m²
MADE IN JAPAN

JAN : 4979093236404

Scanned with CamScanner

NO Class #1 1

DATE 01 17 24

Dr. Peter Oberly → Hylan 811

Office hours W/F 2-3pm

What is this class about? → Precision understanding randomness

How to count:

Suppose one coin is flipped 3 times. What is the probability of flipping exactly one heads.

↳ All possibilities:

HTT ★

HHH → 3/8 ✓

HHT

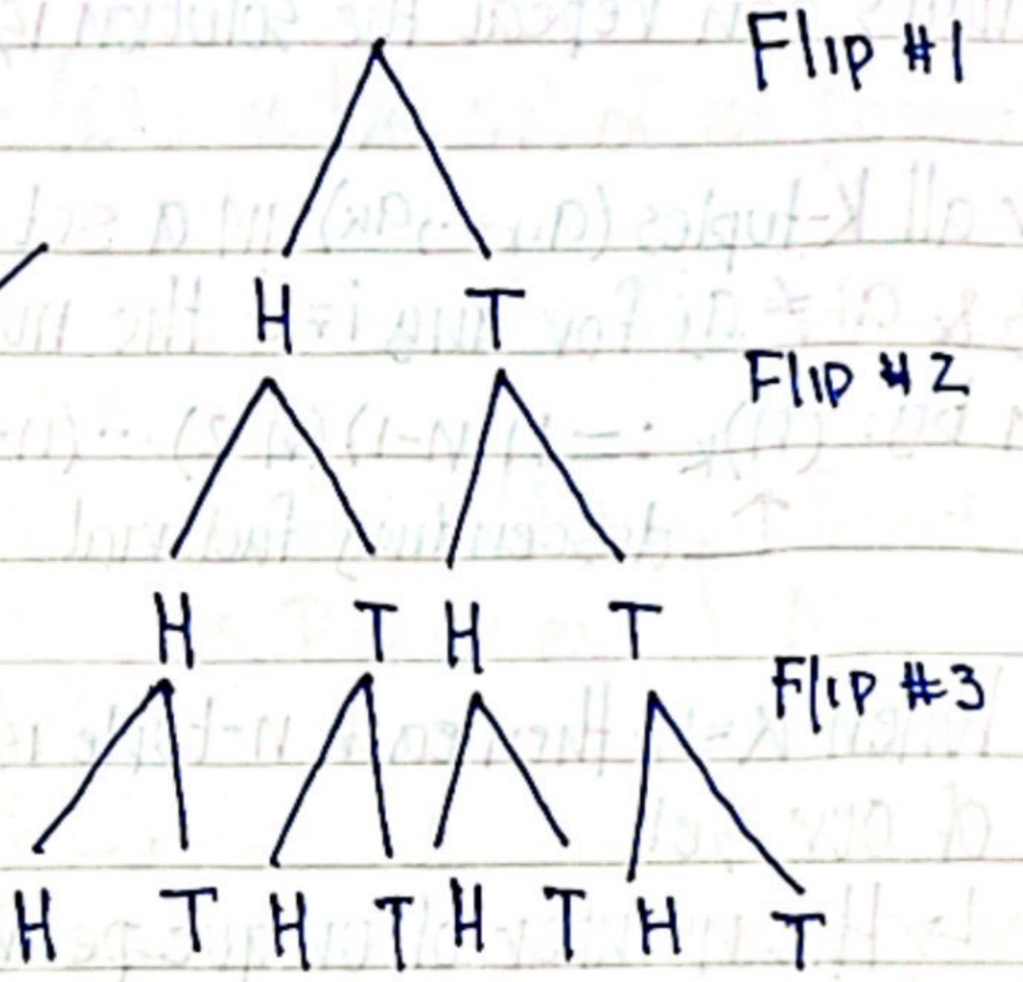
THH ★

TTT

THH

THT ★

HTH



This is tedious... There must be another way!

Multiplication Principle: Suppose that a set of n-tuples satisfies

1. the number of choices for a_1 is r_1
2. once the first k entries $a_1, \dots, a_k$ have been chosen there are r_{k+1} choices remaining for a_{k+1} & r_{k+1} does not depend on the previous choices

Then, the total # of n-tuples is $r_1 \cdot r_2 \cdot r_3 \dots \cdot r_n$

↳ This concept defines the factorial(!)

Scanned with CamScanner

Class #12
01/17/24

Multiplication Principle Example:
Suppose license plates have 3 letters followed by 3 digits
no repeated letters. How many license plates are possible
↳ Think of this situation as sequential choices, When chosing the first letter there are 26 possibilities.
↳ Since letters cannot repeat there only 25 choices for the 2nd letter & 24 for the 3rd
↳ since digits can repeat the solution is $(26 \cdot 25 \cdot 24)(10 \cdot 10 \cdot 10)$

Fact: Consider all k-tuples $(a_1, \ldots, a_k)$ in a set A where A has n elements & $a_i \neq a_j$ for any $i \neq j$ the number of such k-tuples is given by: $(n)_k := n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$
↑ descending factorial

Special Case: When $k=n$ then each n-tuple is an odering (or permutation) of our set
↳ The number of unique permutations a given set has is given by n! (n factorial)

What about cases where order is irrelevant? → Binomial Coefficients
Let k & n be intergers with $0 \leq k \leq n$ The # of distinct subsets of size k in the set n is given by:
$\binom{n}{k} := \frac{n!}{k!(n-k)!}$ → called binomial coefficients
↳ "n chose k" → Example how many unique 5 card hands can be made with a standard deck?.
$\binom{52}{5} = \frac{52!}{5!(47!)}$

Scanned with CamScanner

Class #2 3
01/22/24

Multinomial Coefficients:

Let n & r be nonnegative Intergers such that $K_1 + \cdots + K_r = n$ When chosing subgroups of varying sizes they are multinomial coeffients

$$\binom{n}{K_1, K_2, \cdots, K_r} = \frac{n!}{K_1! K_2! \cdots K_r!}$$

1.1. Sample Spaces & Probabilies:

Kolmogorov Axioms:

1. The sample space (Ω) is the set of all possible outcomes of the experiments
2. Subsets of Ω are called events, the collection of events in Ω is denoted by F
3. The Probability Distribution (P) is a function from F into the real numbers, P of an event A,
   - ↳ $0 \leq P(A) \leq 1$
   - ↳ $P(\Omega) = 1$ & $P(\varnothing) = 0$
   - ↳ If $A_1, A_2, A_3, \cdots$ is a sequence of pairwise disjoint events, then: $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

The triple of (Ω, F, P) is called a probability space

1.2. Random Sampling: It does exist lol

If the sample space Ω has finitely many elements & each outcome is equally likely then for any event $A \subset \Omega$ (↑ within)

Then: $P(A) = \frac{\#A}{\#\Omega}$

Scanned with CamScanner

Class #2 4

DATE 01/22/24

1.3 Infinitely Many Outcomes:

Geometric Probability Distribution:

$\sum_{k=1}^{\infty} P\{k\} = \sum_{k=1}^{\infty} 2^{-k} = 1 \rightarrow$ Implying $P\{\infty\} = 0$

Countably Infinite Sample Spaces:

$P(A) = \sum_{\omega : \omega \in A} P\{\omega\}$ for any event $A \subset \Omega$

↳ meaning its elements are sequential

~~Counting & Permutation Examples~~

~~Binomial & Multinomial Examples~~

1.4. Consequent of the Rules of Probability:

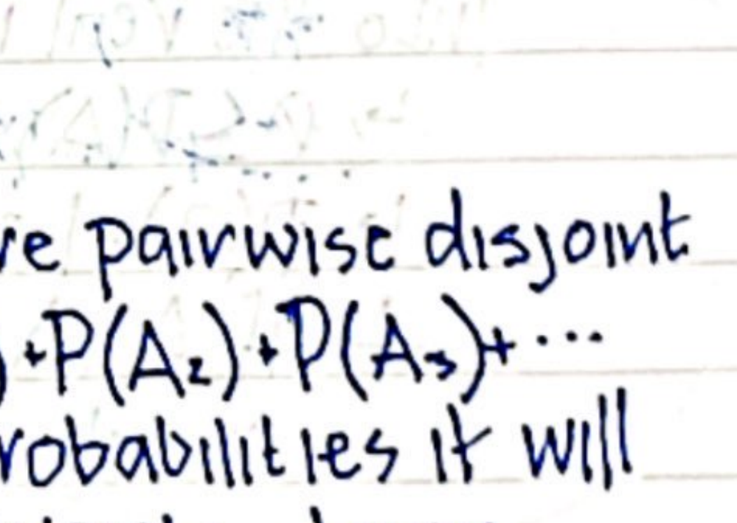
Event decomposition: If $A_1, A_2, A_3 \ldots$ are pairwise disjoint an A is their union then: $P(A) = P(A_1) + P(A_2) + P(A_3) + \cdots$

↳ When calculating complicated probabilities it will very often be neccessary to break events down into several simpler ones

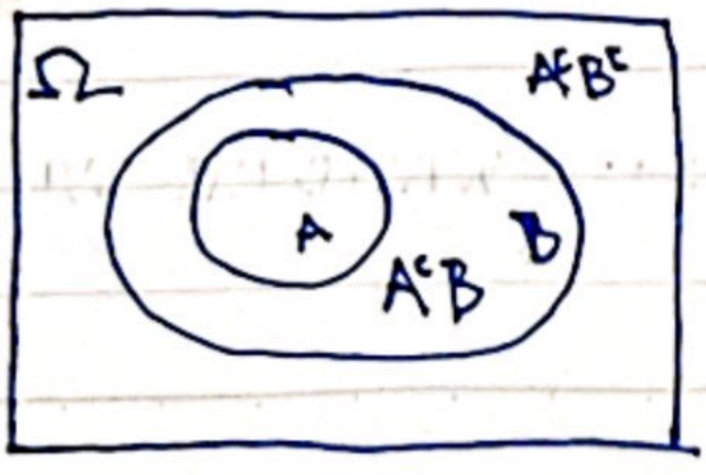
Complementary Event: If two events A & A^c are ME and together encompass the entire sample space then $P(A) + P(A^c) = 1$

↳ A useful extrapolation of this fact:

$P(B) = P(A \cap B) + P(A^c \cap B) \rightarrow$



Monotonicity: Larger events have larger probabilities: if $A \subseteq B$, then $P(A) \leq P(B)$



Scanned with CamScanner

Class #35
01/24/24

Inclusion & Exclusion:

↳ Recall: when two events are disjoint their probabilities can be added, but what about events that are not mutually exclusive?

Inclusion-Exclusion Formula (2 events):

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

Inclusion-Exclusion Formula (3 events):

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

↳ Example: An urn has 30 red, 20 green, & 10 yellow marbles. When drawing 2 without replacement. What is the probability of drawing 1 red or 1 yellow

$$P(R \cup Y) = P(R) + P(Y) - P(R \cap Y)$$

$$= \frac{30 \cdot 30}{60/2} + \frac{10 \cdot 50}{60/2} - \frac{30 \cdot 10}{60/2} = \frac{110}{177}$$

1.5: Introduction to Random Variables

Random Variable: A function that inputs a sample space & outputs real numbers

↳ Random variables are denoted by capital letters

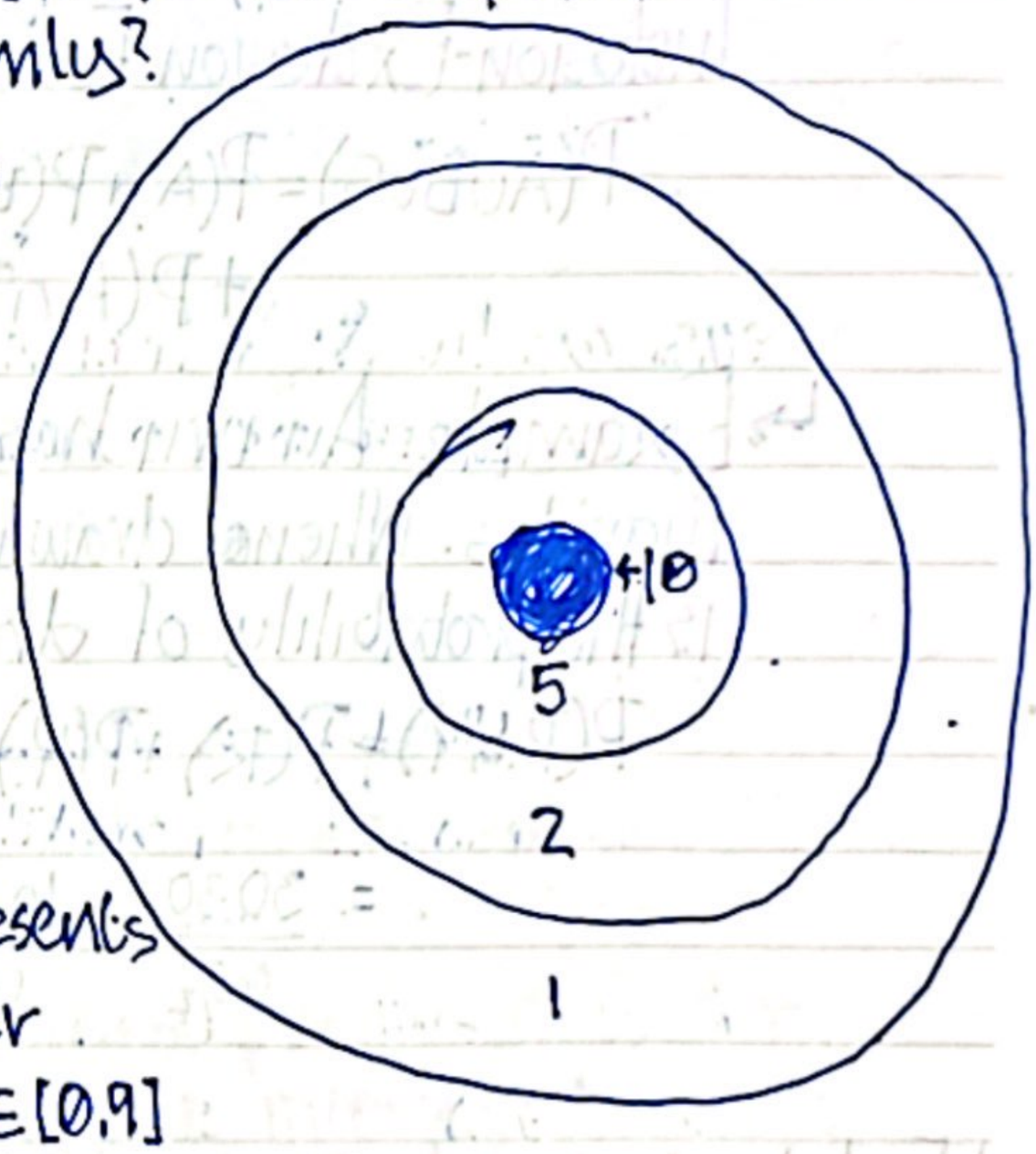
Probability Distribution: for a random variable X, is the collection of probabilities $P\{X \in B\}$ for sets B of real numbers

↳ Think of them as assignments of probabilities to subsets of $\mathbb{R}$ that satisfies the axioms of probability

Discrete Random Variable: Any random variable X such that there exists a finite or countable infinite set $\{k_1, k_2, k_3, \ldots\}$ of real numbers such that $\sum_i P(X = k_i) = 1$ where the sum ranges over the entire set of points $\{k_1, k_2, k_3, \ldots\}$

Scanned with CamScanner

Probability mass function of a DRV is given by p
$p(k) = P(X=k)$ for possible value k of X
↳ Example: A dart board with a radius of 9in is divided into four regions by 3 concentric circles with raidii 1, 3, & 6 respectively with the bullseye being worth 10, with the next ring being worth 5, then 2 & finally 1. What is the distribution of points when throwing darts randomly?
↳ The possible values of X are: 1, 2, 5, 10
↳ The probability of a random dart hitting each region is given by the probability mass function



Lets introduce another random variable that represents the distance from the center
↳ Let $R \in [0,9]$ for any $a \in [0,9]$

$$P\{R \le a\} = \frac{\pi \cdot a^2}{\pi \cdot 9^2} = \frac{a^2}{81}$$

We can use this to calculate the probability mass function

$P\{X=10\} = P\{R \le 1\} = 1/81$
$P(X=5) = P\{1 < R \le 3\} = P\{R \le 3\} - P(R \le 1) = 8/81$
$P\{X=2\} = P\{3 \le R \le 6\} = P\{R \le 6\} - P\{R \le 3\} = 27/81$
$P\{X=1\} = P\{6 \le R \le 9\} = P\{R \le 9\} - P\{R \le 6\} = 45/81$

Scanned with CamScanner

1.6: Continuity & Measurability:

With our shiny new probability functions we can establish a definition of continuity as we can with Cartesian

↳ If we have increasing events in a sequence then the probability of the union of all the events is equal to the limit of the individual probabilities

$$\lim_{n\to\infty} P(A_n) = P(A_\infty)$$

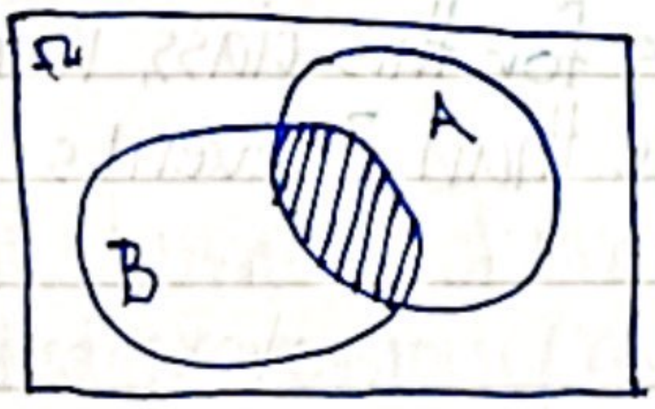
a probability function, then, can be said to be continious if $f(x_n)\to f(x)$ as $n\to\infty$

Measurability & Notation of Sample Events

↳ TLDR it can get so much worse

2.1 Conditional Probability:

Let B be an event in the sample space Ω such that $P(B)>0$. Then, for all events A the conditional probability of A given B is given by: $P(A|B)=\frac{P(AB)}{P(B)}=\frac{P(A\cap B)}{P(B)}$



Example: A bag has 4 red & 6 green balls, with a sample of 3 without replacement find the conditional probability of having exactly 2 red balls in the sample given that there is one red ball

A = there are 2 red balls

B = there is at least 1 red ball

$P(B)=1-\frac{\binom{6}{3}}{\binom{10}{3}}=5/6$ → Thus,

$$P(AB)=P(A)=\frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}=3/10$$

Finally, $P(A|B)=\frac{3/10}{5/6}=\boxed{9/25}$ ✓

Multiplycation Rule: $P(AB)=P(A)P(B|A)$

Scanned with CamScanner

2.2: **Bayes' Formula**: If $P(A), P(B), P(B^c) > 0$, then.

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

Example:

A medical device detects 96% of real cases but gives false positives 2% of the time. Assuming 0.5% of the population carries the disease. If a random person tests positive what is the probability they actually do have the disease?

$D = \{\text{Person has the disease}\}$

$A = \{\text{Person tests positive}\}$

↳ $P(A|D) = \frac{96}{100}$ $P(A|D^c) = \frac{2}{100}$ $P(D) = \frac{5}{1000}$

Applying Bayes Formula gives:

$$P(D|A) = \frac{(96/100)(5/1000)}{(96/100)(\frac{5}{1000} + \frac{2}{100})(995/1000)} = 96/494$$

↳ 0.194

Note: As with most of the formulas for this class, it can be generalized to include more than 2 events, but the formula is ardious

2.3: Independence:

Bayes Formula helps us calculate the probability that an event will occur based on the occurance of another event → this shows the dependence of the events on one another

A&B are **Independent** if: $P(AB) = P(A)P(B)$

Note: Independence only holds when A & B are in the same sample space

Scanned with CamScanner

Revisiting The Different Types of Random Sampling: (1.2)

Sampling with replacement, Order matters:

↳ The number of unique outcomes is determined by:

n^k → Where k is the number of trials & n is the number of outcomes each trial can have

Since there is replacement the probability remains constant

↳ thus, the probability of each set of outcomes is given by n^{-k}

Sampling without replacement. Order matters

↳ The number of unique sets of outcomes is given by the descending factorial → $(n)_k$ Which is defined as

↳ $(n)_k = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)}$

Similar to the case above the probability of each set of outco is given by $\frac{1}{(n)_k}$ → P.S. $(n)_k = \frac{n!}{(n-k)!}$

Sampling without replacement. Order does not matter

↳ The number of unique sets of out comes is given by the expression: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ → "n chose k

As with the other sampling methods, taking the reciprocal of this yields the probability of each set of outcomes

↳ $\frac{k!(n-k)!}{n!}$

Sampling without replacement, Order does not matter:

↳ The number of unique sets of outcomes is given by

$\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$

Scanned with CamScanner

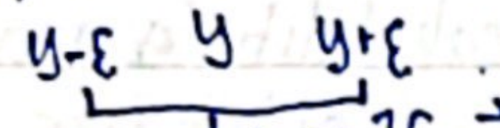
Uniform Infinite Sample Space Example:

Let $\Omega = [0,1]$ & X is a randomly chosen number in Ω

We can use monotonicity to prove X is not discrete

↳ let y be an arbitrary point in $(0,1)$

for any $\varepsilon > 0$, $\{X=y\} \subseteq \{y-\varepsilon \leq X \leq y+\varepsilon\}$

↳ 

↳ $= 2\varepsilon$ → Thus, $0 \leq P\{X=y\} \leq P\{y-\varepsilon \leq X \leq y+\varepsilon\} = 2\varepsilon$

↳ Then it follows that $\lim_{\varepsilon \to 0} \left(P(X=y)\right) = 0$!

1.4: Event Decomposition Example:

An urn contains 30 red, 20 green, & 10 yellow balls. 2 balls are drawn at random. What is the probability that the sample contains 1 red & 1 yellow ball?

↳ $\{R,Y\} \cup \{R,R\} \cup \{R,G\}^{or} \cup \{Y,G\}$

$P(R \cup Y) = P(R,Y) + P(R,G) + P(Y,G) = \frac{110}{177}$

More Random sample examples:
↳ Unordered Sample. Without Replacement:
[redacted] When drawing 3 cards from a deck

Discrete Random Variables (Continued)
↳ Recall: Discrete Random variables are any finite or countably Infinite RVs
Example $\Omega = [0,1]$, Pick a # Uniformly at random
let $X(\omega) = \begin{cases} 1, & 1/2 \le \omega \le 1 \\ 0, & 0 \le \omega \le 1/2 \end{cases}$ → This uncountably infinite Ω can be made discrete

Probability Mass Function: Let X be a discrete R.V.

$$P_X(k) = P\{X = k\} = P\{\omega \in \Omega : X(\omega) = k\}$$

↳ This probability mass function determines the Probability Distribution & If $B \subseteq R$, then

$$P\{X \in B\} = \sum_{k \in B} P(X = k) = \sum_{k \in B} P_X(k)$$

↳ See dart board example on page 6.

Scanned with CamScanner

NO.

DATE

Sample Space Partitioning & Bayes Theorem

↳ Recall Baye's Theorem:

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

$P(A_1) = 0.2$ $P(E|A_1) = 0.5$

$P(A_2) = 0.1$ $P(E|A_2) = 0.4$

$P(A_3) = 0.7$ $P(E|A_3) = 0.6$

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

$$P(E|A_1) = \frac{P(EA_1)}{P(A_1)} = \frac{P(E)\,0.2}{0.2} \neq 0.5$$

$$\frac{P(EA_1)}{0.2} = 0.5 \qquad P(EA_1) = 0.1$$

$$P(B|A) = \frac{P(A|B)\cdot P(B)}{P(A)} \rightarrow P(A_1|E) = \frac{P(E|A_1)\cdot P(A_1)}{P(E)}$$

$$P(A_1|E) = \frac{0.5\cdot 0.2}{P(E)} = \frac{0.1}{P(E)}$$

$$P(A_2|E) = \frac{P(E|A_2)\cdot P(A_2)}{P(E)} = \frac{0.04}{P(E)}$$

$$P(A_3|E) = \frac{P(E|A_3)\cdot P(A_3)}{P(E)} = \frac{0.42}{P(E)}$$

$\cdot\, P(E) = 0.56$

Scanned with CamScanner

NO

DATE

2.4: Independent Trials & Distributions:

~~Bernoulli Distribution:~~

~~Let $0 \leq p \leq 1$ X is a Bernoulli Distribution:~~

2.3. Independence:

Two events A and B are Independent if

$P(AB) = P(A)P(B)$

Note: A & B must be in the same sample space

↳ Additionally, if two events are independent so are their compliments

Pairwise Independence: If two events are independent under specific indicies

Scanned with CamScanner

NO

DATE

MATH 201: Midterm #1 Topic List 02/20/24

1.1: Sample Spaces & Probabilities
1.2: Random Sampling
1.3: Infinitely many outcomes
1.4: Consequenses of the Rules of Probability
1.5: Random Variables
2.1: Conditional Probability
2.2: Bayes Theorem
2.3: Independence
2.4: Independent Trials

Key Concepts:

1.1: Multiplication Principals
1. With Replacement, Order matters
   # of unique outcomes = n^k
2. Without Replacement, Order does not matter
   # of unique outcomes = $\binom{k}{n} = n!/k!(n-k)!$
3.

Scanned with CamScanner

NO.

DATE

Midterm #1 (2021) – Practice

Q1. (5 points) Suppose an Urn contains 5 red balls, 7 green balls and 9 orange balls. 5 balls are drawn one at a time without replacement:

a. What is the probability that the sample contains 3 red balls

P(exactly 3 red balls) = 3/5 are red

↳ how many sets of 5 balls can be drawn?

$$\binom{5}{25} \rightarrow \frac{n!}{k!(n-k)!} \rightarrow \frac{25!}{5!\,20!}$$

↳ $(5/25)(4/24)(3/23)$ = P(choosing 3 red balls)

how many different ways could these draws appear in our sample of 5? → $\binom{5}{3} = 10$

XXXOO

OOXXX

OXOXX

OXXOX

OXXXO

XOXXO

$10(5/25)(4/24)(3/23)(16/22)$

$10(5/21)(4/20)(3/19)(16/18)(15/17)$

b. What is the probability that at least 1 red & 1 green?

P(≥1 red & ≥1 green) = P(≥1 red) · P(≥1 green)

Scanned with CamScanner

NO.

DATE

Q2: Peter & Mary take take turns rolling a fair die. If Peter rolls 1 he wins & the games ends. If Mary rolls a 3 or 6 she wins & the game ends. Peter rolls first:

a. What is the probability Mary wins on her 5th roll?

Who?	roll #	Prob of continuing:
1. P	1	5/6
2. M	1	4/6
3. P	2	5/6
4. M	2	4/6
5. P	3	5/6
6. M	3	4/6
7. P	4	5/6
8. M	4	4/6
9. P	5	5/6
10. M	5	2/6

$$(5/6)^5 (4/6)^4 (2/6)$$

2b. What is the probability Mary Wins?

$\lim_{n\to\infty}$ ~~$(2/6)^n$~~ ~~$\lim_{n\to\infty}(2/6)^n$~~

$(2/6)^{n-1}(5/6)^n$ ← P(the game continues)

↑ P(mary wins on a given roll)

$\lim_{n\to\infty}((2/6)^{n-1}(5/6)^n)$ Where n is the dice being rolled

↳ $\frac{2^{n-1}}{6^{n-1}} \cdot \frac{5^n}{6^n}$ → ~~decreases/grows faster~~

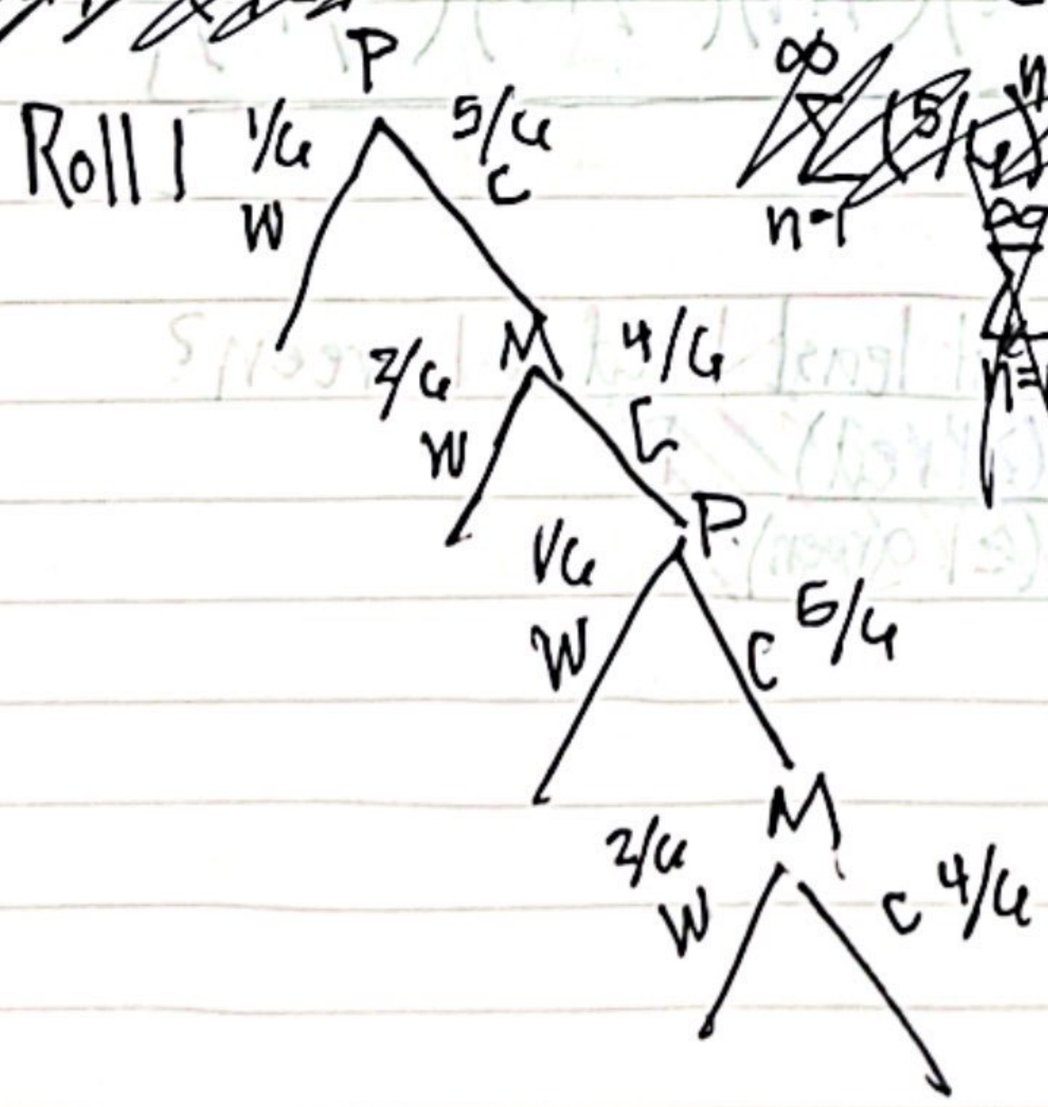
~~∞ $n=1$~~

∞ n

$(5/6)^n (4/6)^{n-1} \cdot (2/6)$

$n=1$ $n-1$

↳ as $n \to \infty$ the p that mary wins goes to 2/4



Scanned with CamScanner

Consider events A, B, & C which are mutually independent (meaning $P(A \cap B \cap C) = P(A)P(B)P(C)$) $P(A) = 1/2$, $P(B) = 1/4$ $P(C) = 1/2$

a. Compute $P(A \cup B)$

↳ A & B are independent meaning $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B) = \frac{1}{2} + \frac{1}{4} - 1/8 = \boxed{5/8}$

b. Are $A \cup B$ and C independent:

↳ yes, since A & C are independent and B & C are independent we calculate $P(A \cup B)$ independent of $P(C)$

$P(A \cup B | C) = P(C | A \cup B)$

A fair Coin is flipped 4 times. Let A be the event that tails comes up at least 3 times, let B be the event that the first 3 flips are tails.

$P(A)$ = Recall Binomial Formula: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ k

↳ $P(X_t \geq 3) = \binom{n}{3} 0.5^3 0.5^{n-3}$ $k = 0, \dots, n$

$P(X_t \geq 3) = \binom{4}{3} (0.5^3)(0.5)^1$ $= 4(0.5)^4 + (0.5)^4 = 5(0.5)^4$

$= \boxed{5/16}$

↳ $\frac{4!}{3!(4-3)!} = \frac{4 \cdot 3 \cdot 2}{3 \cdot 2} = 4$

Find $P(B|A)$

Recall Baye's Thm: $P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$

$P(A|B) = 1$

$P(B) = 0.5^3$

$P(A|B^c) = 3/16$

$0.5^3 / 0.5^3 + \frac{3}{16}(1 - 0.5^3) = \frac{1/8}{1/8 + 1/8(7/8)} = \frac{1}{8} \cdot \frac{64}{}$

$1/8 \cdot 7/8 = 7/64$

$\frac{8}{64} + \frac{7}{64} = \frac{15}{64}$

$\frac{1/8}{1/8 + (\frac{3}{16} \cdot \frac{7}{8})} = \frac{1/8}{37/128}$

$\frac{1}{8} \cdot \frac{128}{37} = \boxed{\frac{128}{296}} = 0.432$

↓

$\frac{21}{128}$

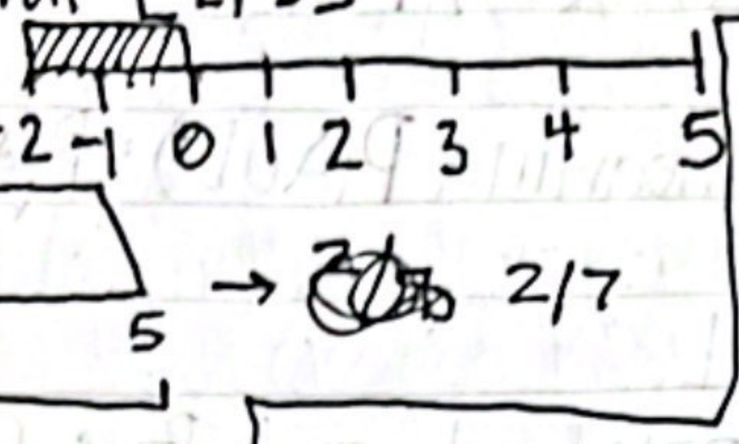
Scanned with CamScanner

NO

DATE

The Continious Random Variable X is uniformly distributed on the Interval [-2, 5]

find $P(X<0)$



find $P(X^2>1)$

X	X^2	$X^2>1$?
-2	4	Y.
-1	1	N.
0	0	N.
1	1	N
2	4	Y.
3	9	Y.
4	16	Y.
5	25	Y.

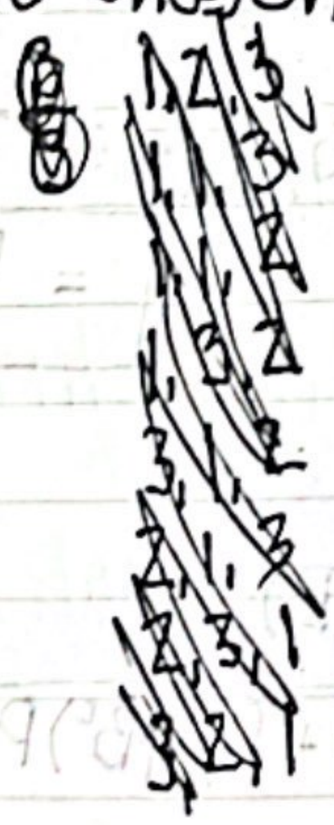
5/7

6. 3 numbers are chosen w/ replacement from the set {1, 2, 3}

What is the chance no number is chosen twice?

$\frac{3}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} = \frac{6}{27} = 2/9$

What is the chance at least two different numbers are chosen.



$\frac{2}{3} \cdot \frac{2}{3} \rightarrow 4/9$

Scanned with CamScanner

7. There are three types of coin in circulation. Fair ($P(H)=1/2$), biased ($P(H)=1/3$) & rigged ($P(H)=1/5$). Suppose 1/2 of the coins are fair, 1/4 are biased, & 1/4, A coin is flipped twice and the outcome is heads followed by tails What is the probability the coin is fair?.

$P(\text{the coin is fair} \mid n=2, H, T)$

↳ Recall Bayes Thm:

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

$P(n=2, H,T \mid \text{the coin is fair}) = 1/4$

$P(\text{the coin is fair}) = 1/2$

$P(n=2, H,T, \mid \text{the coin is NOT fair}) = \frac{1}{2}\left(\frac{1}{4}\cdot\frac{3}{4}\right) + \frac{1}{2}\left(\frac{1}{5}\cdot\frac{4}{5}\right) =$

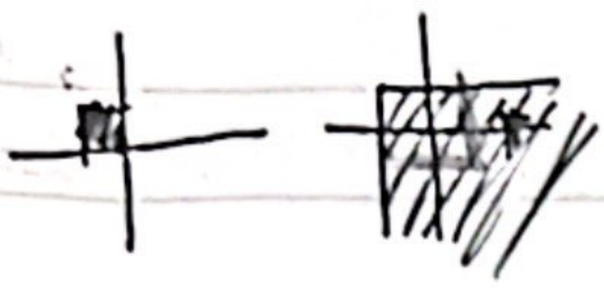
↳ $\frac{139}{800}$ $P(\text{The coin is NOT fair}) = \frac{1}{2}$ $\frac{1}{2}\frac{3}{16} + \frac{1}{2}\frac{4}{25}$

$\frac{3}{32} + \frac{2}{25} = \frac{75+64}{800} = \frac{139}{800}$

$$\frac{\frac{1}{4}\cdot\frac{1}{2}}{\frac{1}{4}\cdot\frac{1}{2} + \frac{139}{600}\cdot\frac{1}{2}} = \frac{1}{4}\Big/\frac{289}{600}$$

$\frac{1}{4}\cdot\frac{600}{289} = \frac{600}{1156}$

Scanned with CamScanner



MATH 201 Formulas to know (Midterm #1):

$(n)_k = \frac{n!}{(n-k)!}$ → (Sampling W/O replacement, order matters)
↳ (a.k.a. descending factorial)

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ → (Sampling W/O replacement, No order)
↳ "n choose k"

Inclusion-Exclusion Rules:

↳ 2 Events: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

↳ 3 Events: $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A \cap B)}{P(B)}$ → Conditional Probability

$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$ → Baye's Theorem

For Independent events: $P(A \cap B) = P(A)P(B)$

Binomial Distribution: $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

Scanned with CamScanner

2.4 Indepent Trials:

The idea of Independence can easily be extended to the idea of several repeated trials conducted independently

Bernouli Distributions:

↳ Records the result of a single trial with 2 possible outcomes

Binomial Distributions:

↳ When counting the number of successes in a series of Independent binary trials

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k} \quad \text{for } K=0,1,\cdots,n$$

Where n = the number of trials
k = the number of successes
p = the chance of success

Geometric Distributions:

↳ When considering an indefinate number of repeated trials it is neccessary to use the geometric distribution

$$P(X=k)=(1-p)^{k-1}p$$

Where p = the chance of success
k = the number of successes necessary

Together these encompass binary distribution

Example: What is the probability that it takes more than 7 rolls of a fair dice to roll a 6?

$$P(N>7)=\sum_{K=8}^{\infty}P(N=K)=(1-p)^{k-1}p=\sum_{K=8}^{\infty}(\tfrac{5}{6})^{k-1}\tfrac{1}{6}=\frac{\frac{1}{6}(\frac{5}{6})^7}{1-5/6}=(5/6)^7$$

Scanned with CamScanner

NO.

DATE

3.1: Probability Distributions of Random Variables

Probability Mass Functions:

Recall that a random variable is discrete if there is a finite or countably infinite number of elements in the sample space that sum to 1

↳ The probability distribution of a discrete random variable is entirely determined by its probability mass function

$P(k) = P(X=k)$ → The values of a pmf must sum to 1

$$\sum_k P_X(k) = \sum_k P(X=k) = 1$$

Probability Density Functions:

↳ Let X be a random variable. If a function f satisfies

$$P(X \leq b) = \int_{-\infty}^{b} f(x)dx$$ for all real values of b, then f is the probability density function of X

In the case that X has a pdf then point values have a probability of zero

Uniform distributions

↳ Let [a,b] be a bounded interval on the real line. A random variable has the uniform distribution on the interval [a,b] if X has the density function $f(x) = \begin{cases} \frac{1}{b-a} & \text{if } X \in [b,a] \\ 0 & \text{if } X \notin [b,a] \end{cases}$

If $X \sim \text{Unif}[a,b]$ and $[c,d] \subset [a,b]$, then

$$P(c \leq X \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}$$

Scanned with CamScanner

NO

DATE

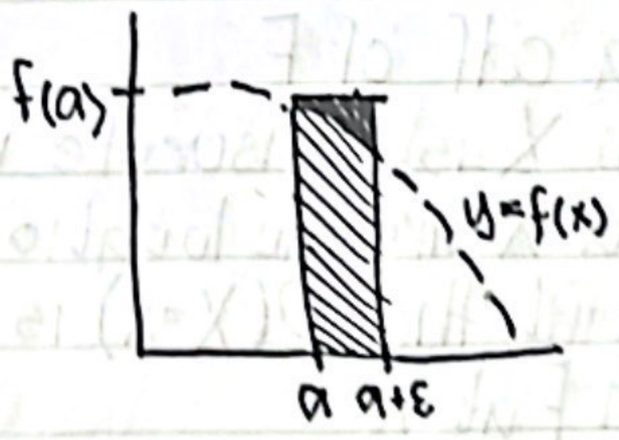
Infantesimal Method:

Note: the value $f(x)$ of a density function is NOT a probability
However, multiplied by the length of a tiny interval it approximates the probability of the interval
Suppose that X has a density function f that is continious at the point a. Then for small $\varepsilon > 0$

$$P(a < X < a+\varepsilon) \approx f(a) \cdot \varepsilon$$

↳ This is derrived from:

$$\lim_{\varepsilon \to 0}\left(\frac{P(a \le X \le a+\varepsilon)}{\varepsilon}\right) = \lim_{\varepsilon \to 0}\left(\frac{1}{\varepsilon}\int_a^{a+\varepsilon} f(x)dx\right) = f(a)$$



3.2: Cumulative Distribution Function

Just as probability mass functions are only defined for discrete random variables, and density functions are only defined for continious ones
Probability Distribution Functions describe any random variable
The **cumulative distribution function** (C.D.F) of a random variable (c.d.f.) is given in it

$$F(s) = P(X \le s) \quad \text{for all } s \in \mathbb{R}$$

↑ note that the cdf gives probabilities of left-open right-closed intervals of the form $(a,b]$: $P(a < X < b) = P(X \le b) - P(X \le a)$
$= F(b) - F(a)$

Scanned with CamScanner

NO

DATE

CDF of Discrete Random Variables:

↳ Recall the ~~cumulative~~ pmf. It is related to the cdf

$$F(s) = P(X \leq s) = \sum_{k:k \leq s} P(X = k)$$

CDF of Continious Random Variables:

↳ recall the pdf, it is related to the cdf

$$F(s) = P(X \leq s) = \int_{-\infty}^{s} f(x)dx$$

Finding pmf and pdf from the cdf

Let the random variable X have a cdf of F

a. If F is piecewise constant, then X is a discrete random variable. The possible values of X are the locations where F jumps, and if x is such a point, then $P(X=x)$ is equal to the magnitude of the jump in F at x

b. If F is continious and the derivative $F'(x)$ exists everywhere on the real line, except possibly at finitely many points. Then X is a continious random variable and $f(x) = F'(x)$ is the density function of X. If F is not differentiable at x then the value $f(x)$ can be set arbitrarily

Let X be a random variable with a c.d.f. of F. Then for any $a \in \mathbb{R}$

$$P(X < a) = \lim_{s \to a^-} (F(s))$$

Point Probabilites then are given by: $P(X=a) = P(X \leq a) - P(X < a)$

$= F(a) - F(a^-)$

= size of jump in F at point a

Scanned with CamScanner

How can we find the p.d.f. of X? → Infantesimal Method Example

If $f(x)$ is continious at a, then

$$\frac{1}{\varepsilon}\int_a^{a+\varepsilon} f(x)dx \to f(a) \text{ as } \varepsilon \to 0^+ \longrightarrow P(a \le X \le a+\varepsilon) \approx f(a)\cdot\varepsilon$$

Suppose $f(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

find $P(0.5 \le X \le 0.51) \to \varepsilon = 0.01$

$$\int_{0.5}^{0.51} 3x^2 dx = (0.5)^3 - (0.51)^3$$

$$\approx 0.007651$$

$$(0.007651)(0.01) = \boxed{0.0075}$$

↳ $f(x) = \lim_{\varepsilon \to 0^+} \frac{P(x \le X \le x+\varepsilon)}{\varepsilon}$

Discrete Random Variable Example:

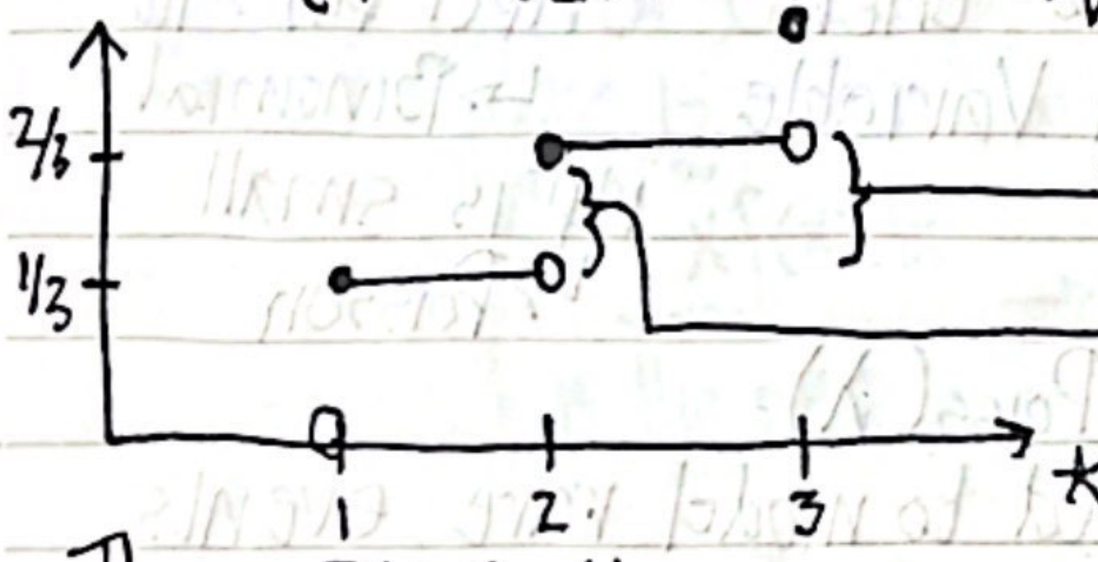
Suppose the c.d.f. of a random variable X is given by

$$F(x) \begin{cases} 0 & x<1 \\ 1/3 & 1 \le x < 2 \\ 2/3 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

→ Since $F(x)$ is piecewise constant X is a discrete random variable

↳ find the p.m.f.

↳ how? graph $F(x)$!



$F(2) - \lim_{s \to 2^-}(f(s) \cdot P(X=2))$

$F(1) - \lim_{s \to 1^-}(f(s)) = P(X=1)$

Thus, $P(X=1) = 1/3$

$P(X=2) = 1/3$

$P(X=3) = 1/3$

* note that since X is discrete rather than continious we find the p.m.f. for Continious RV we would find the p.d.f.

Scanned with CamScanner

NO

DATE

Binomial Random Variables:

↳ Counting the number of successes in a <u>fixed number of trials</u> $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ Where:

n = number of trials

k = number of successes

p = Probability of Success

$X \sim Bin(n, p)$

$\mu(X) = n \cdot p = E(X)$

$\sigma^2(X) = np(1-p) = Var(X)$

Geometric Random Variables $X \sim Geom(p)$

↳ $P(X=k) = (1-p)^{k-1} p$

$\mu(X) = 1/p = E(X)$, $\sigma^2(X) = \frac{1-p}{p^2} = Var(X)$

Standard Normal (Gaussian) Random Variables

↳ If $X \sim N(\mu, \sigma^2)$ with σ & $\mu > 0$ then $\frac{X-\mu}{\sigma}$ = a <u>standard</u> Normal Variable

$np(1-p) > 10$

↳ Binomial

np^2 is small

↳ Poisson

Poisson Random Variables: $X \sim Poiss(\lambda)$

$P(X=k) = \frac{e^{-\lambda} \lambda^n}{n!}$ $\mu = \lambda$ used to model rare events

$\sigma^2 = \lambda$

Exponential Random Variables

Scanned with CamScanner

NO

DATE

3.3: Expectation:

Probability distributions are useful for describing RVs but they are big and can be difficult to calculate

↳ Expectation, variance, and moments provide useful information about RVs in a single quantity

Expectation of Discrete Random Variables

↳ Also called the <u>mean</u> is given by

$$E(X) = \sum_k k P(X=k) = \mu$$

Where the sum ranges over all the possible values k of X

The expectation is also called the first moment.

Mean of a binomial RV

$$\mu = E(X) = \sum_{k=0}^{n} kP(X=k) = \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \xrightarrow{\text{simplifies to}} (n \cdot p)$$

Expectation of Continious Random Variables

↳ the mean is given by:

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Where f is the p.d.f.

→ Mean of a uniform RV

$$E(X) = \int_{-\infty}^{\infty} x \cdot \frac{1}{b-a} dx \rightarrow \frac{a+b}{2}$$

Note: it is possible for RVs to have a non-finite or- even undefined means

The median of a random variable X is any real value m that satisfies $P(X \geq m) \geq \frac{1}{2}$ $P(x \leq m) = 1/2$

For $0 < p < 1$, the pth quartile of X is any real value x that satisfies $P(X \geq x) \geq 1-p$ & $P(X \leq x) \geq p$

Scanned with CamScanner

Expectation Example: Let X be uniformly distributed on the set $\{-100, 1, 2, 3, \ldots, 9\}$, so X has a pmf given by

$P_X(-100) = 1/10 \qquad P_X(k) = 1/10$ for $k = 1, 2, 3, \ldots, 9$

Then $E(X) = -5.5$ while the median is any number $m \in [4, 5]$

The median closely reflects the fact that 90% of the probability is in the interval $[1, 9]$

3.4: Variance: A measure of how much a random variable fluctuates around the mean

Let X be a random variable with mean μ. The variance of X is defined by: $\text{Var}(X) = E((X-\mu)^2)$

The variance is given by $\sigma^2 = \text{var}(X)$

Square rooting the variance gives the standard deviation

Continious Variables:

$\text{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$ → Where f is the density function

Discrete Random Variables:

$\text{Var}(X) = \sum_{k} (k-\mu)^2 P(X=k)$

Larger Deviations from the mean create larger variances

Alternative Forms of the Variance Formula:

$\text{Var}(X) = E(X^2) - (E(X))^2$

$E(aX+b) = aE(X) + b$

$\text{Var}(aX+b) = a^2 \text{Var}(X)$

Variance of a Binomial RV → $np(1-p)$

Scanned with CamScanner

Summarization of the Properties of Random Variables

Discrete	Continious
Probability Mass function	Probability density function
$P_X(k) = P(X=k)$	$f_X(x)$
$P(X \in B) = \sum_{k \in B} P_X(k)$	$P(X \in B) = \int_B f(x)dx$

Cumulative Distribution: $F_X(a) = P(X \leq a)$

$F_X(a) = \sum_{k \leq a} P_X(k)$	$F_X(a) = \int_{-\infty}^{a} f(x)dx$
Step wise	continious

$$P(X<a) = \lim_{t \to a^-} F(t) = F(a^-)$$

$$P(X=a) = F(a) - \lim_{t \to a^-} F(t) = F(a) - F(a^-)$$

Expectation:	
$E(X) = \sum_k k\, p_X(k)$	$E(X) = \int_{-\infty}^{\infty} x f(x)dx$

$$\hookrightarrow E(aX+b) = aE(X)+b \hookleftarrow$$

Variance:	
$Var(X) = E(X-E(X))^2$ $= E(X^2) - (E(X))^2$	$Var(X) = E((X-E(X))^2)$ $= E(X^2) - (E(X))^2$

$$\hookrightarrow Var(aX+b) = a^2 Var(X) \hookleftarrow$$

Scanned with CamScanner

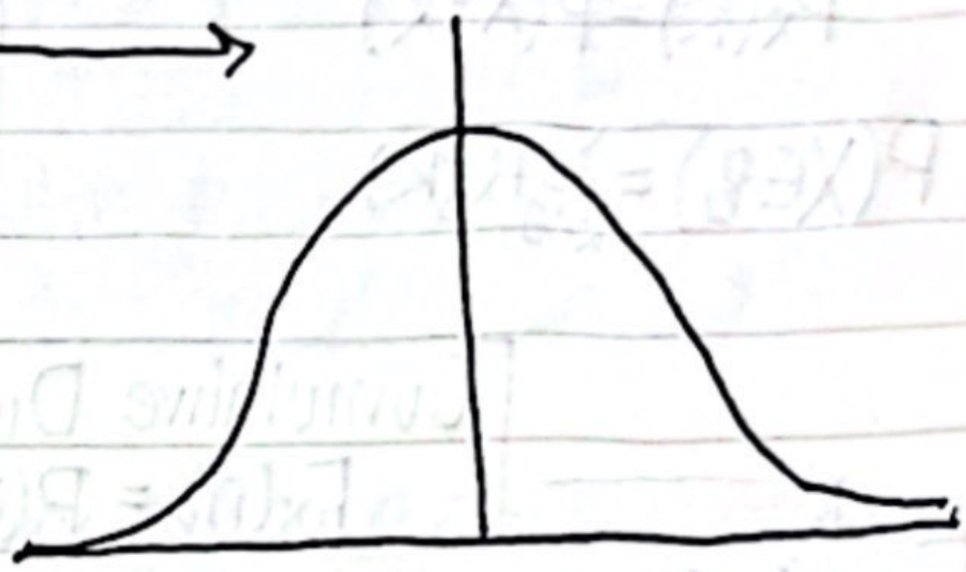
NO.

DATE

3.5: Gaussian (Normal) Distribution:

A random variable Z has the standard normal distribution if its density function is given by:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \rightarrow Z \sim N(0,1)$$



The c.d.f. is given by:

$$\int_{-\infty}^{\infty} e^{-s^2/2} ds = \sqrt{2\pi} = \Phi(x)$$

↳ Since this is so foul to integrate values of $\Phi(x) \rightarrow$ Z table!

When $Z \sim N(0,1)$

$E(Z) = 0$ &

$Var(Z) = 1$

With this standard normal distribution defined we can define non-standard normal distributions in terms of transformations applied to the standard normal distribution, of the form

$$X = \sigma Z + \mu \rightarrow X = \sqrt{Var(Z)}Z + E(Z)$$

Let μ be real and $\sigma > 0$. A random variable X has the normal distribution if its density function is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

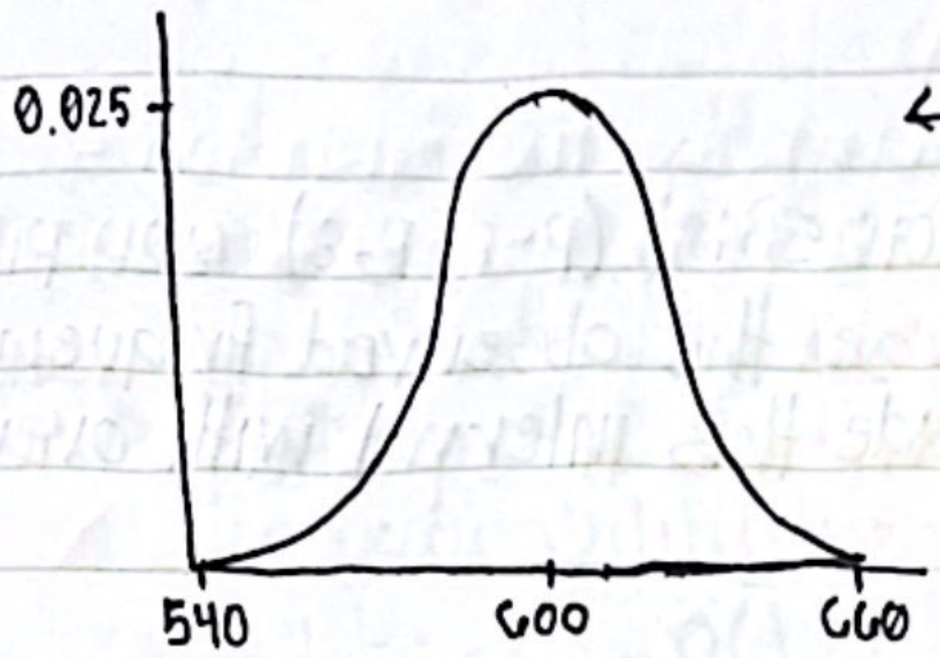
↳ $X \sim N(\mu, \sigma^2)$

If $X \sim N(\mu, \sigma^2)$ & $Y = aX + b$ then, $Y \sim N(a\mu + b, a^2\sigma^2)$

or $Z = \frac{X-\mu}{\sigma}$ is a standard normal variable

Scanned with CamScanner

4.1: Normal Approximation:
Central Limit Theorem (CLT) for the Binomial:



← this curve is the p.m.f. of Bin(1000, 0.6) which is equal to N(600, 240) in the interval [540, 660]

The process of Normal Approximation for Binomial distributions is the same as Z-standardization

CLT for Binomial Random Variables:

Let $0<p<1$ be fixed and suppose that $S_n \sim Bin(n,p)$ then for any fixed interval $-\infty<a<b<\infty$ we have the:

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$$

Also called the de Moivre-Laplace theorem

$P\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right)$ is close to $\Phi(b) - \Phi(a)$

if $np(1-p) > 10$

Example: A fair coin is flipped 10,000 times. Estimate the probability that the number of heads is between 4850 & 5100? S = the # of heads $S \sim Bin(10,000, 0.5)$

$E(S) = 10,000 \cdot 0.5 = 5,000$ & $Var(S) = 10,000 \cdot 0.5 \cdot 0.5 = 2,500$

$$P(4850 \le S \le 5100) = P\left(\frac{4850-5000}{\sqrt{2500}} \le \frac{S-5000}{\sqrt{2500}} \le \frac{5100-5000}{\sqrt{2500}}\right)$$

$$= P\left(-3 \le \frac{S-5000}{\sqrt{2500}} \le 2\right) \approx \Phi(2) - \Phi(-3)$$

$$= 0.9772 - (1 - 0.9987) = 0.9759 \checkmark$$

Scanned with CamScanner

NO.

DATE

4.2 Law of Large Numbers:

For any binomial Random Variable & fixed ε

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = 1$$

This means that no matter how small $(p-\varepsilon, p+\varepsilon)$ you put around p, as n becomes large the observed frequency of successes (S_n/n) will lie inside this interval with overwhelming probability

The CLT relies on the law of large numbers

↳ The CLT gives the error of the LOLN:

$$\frac{S_n}{n} = p + \frac{\sigma}{\sqrt{n}} \cdot \frac{S_n - np}{\sigma\sqrt{n}} \approx p + \frac{\sigma}{\sqrt{n}} Z$$

Where $\sigma = \sqrt{p(1-p)}$

Example: Let S_n denote the # of 6s in n rolls of a die

$\frac{S_n}{n}$ = the observed freq of 6s in n rolls $\quad \lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \frac{1}{6}\right| < 0.0001\right) = 1$

↳ this limit says that as the number of rolls grows deviations of S_n/n from $1/6$ by more than 0.0001 becomes extremely unlikely

Scanned with CamScanner

4.3 Applications of Normal Approximation

Confidence Intervals:

Suppose a biased coin with an unknown p of landing on heads is flipped. How can we estimate p?

↳ Normal Approximation can help!

Let's estimate that the error $|\hat{p}-p|$ is bound by a tight margin of error ε. First, rearrange the inequality inside the probability inequality & standardize

↳ $P(|\hat{p}-p|<\varepsilon) = P\left(\left|\frac{S_n}{n}-p\right|<\varepsilon\right) = P(-n\varepsilon < S_n - np < n\varepsilon)$

$= P\left(-\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} < \frac{S_n-np}{\sqrt{np(1-p)}} < \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)$

$P\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} < \frac{S_n-np}{\sqrt{np(1-p)}} < \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \approx \Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)$

↳ Solving for the lower bound $= 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1$ involves some pretty ugly calculus gives us ⟶

$\geq P(|\hat{p}-p|<\varepsilon) \geq 2\Phi(2\varepsilon\sqrt{n}) - 1$

We can use this inequality to find how many flips we would need to be 99% confident as to the probability of flipping heads within 5%?

↳ $P(|\hat{p}-p|<\varepsilon) \geq \underline{2\Phi(2\varepsilon\sqrt{n}) - 1 \geq 0.99}$

↳ $\Phi(2\varepsilon\sqrt{n}) \geq 0.995$

$2\varepsilon\sqrt{n} \geq 2.58$

$\rightarrow n \geq \frac{2.58^2}{4\varepsilon^2} = \frac{2.58^2}{4(0.05)^2} \approx 665.64$

Scanned with CamScanner

Another common application of confidence intervals is to find the interval around $\hat{p}$ that captures the true probability with a given confidence.

↳ "The 100r% confidence interval for the unknown success probability p given by $(\hat{p}-\varepsilon, \hat{p}+\varepsilon)$ where ε is chosen to satisfy $P(|\hat{p}-p|<\varepsilon) \geq r$. In other words, the random interval $(\hat{p}-\varepsilon, \hat{p}+\varepsilon)$ contains the true p with probability at least r%."

Example: We repeat a trial 1000 times and observe 450 successes. Find the 95% confidence interval for the true success probability p?

$P(|\hat{p}-p|<\varepsilon) \geq r \rightarrow 2\Phi(2\varepsilon\sqrt{n})-1 \geq 0.95$

$r=0.95$

↳ We need to solve this inequality for ε

① Simplify

$\Phi(2\varepsilon\sqrt{n}) \geq 0.975$

② Φ table $\rightarrow 2\varepsilon\sqrt{n} \geq 1.96$

③ Solve! $\rightarrow \varepsilon \geq \frac{1.96}{2\sqrt{1000}} \approx 0.031$

Thus, if $n=1000$, then with probability at least 95% the true success probability p is $(0.45-0.031, 0.45+0.031)$

$= (0.419, 0.481)$

↑ 95% confidence interval for the true probability

Likelyhood function: $L(p) = P(S_n=k) = \binom{n}{k} p^k (1-p)^{n-k}$

(a function of p with k & n as constants)

Scanned with CamScanner

Polling: Confidence intervals are helpful when estimating public opinion, but! Recall that the binomial distribution requires Independent trials (Sampling <u>With</u> replacement) where as random polling <u>samples</u> <u>without</u> replacement making its distribution not exactly binomial → (by definition its actually hypergeometric)

However! Since the population of interest is so much larger than the sample that the change in probability is negligable

↳ through a limit theorem that is so evil im actually gonna write it out lol

$$\left(\frac{N_A}{N}\right)^k \cdot \prod_{i=1}^{k} \frac{\left(1-\frac{i-1}{N_A}\right)}{\left(1-\frac{i-1}{N}\right)} \left(1-\frac{N_A}{N}\right)^{n-k} \cdot \prod_{i=k+1}^{n} \frac{\left(1-\frac{i-k-1}{N-N_A}\right)}{\left(1-\frac{i-k}{N}\right)}$$

Some the fuck how it solves to → $p^k(1-p)^{n-k}$ → Binomial

4.4. Poisson Approximation

Let $\lambda > 0$, a random variable X has the the Poission Distribution with parameter λ if X is non-negative interger-valued and has a pmf given by:

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{for } k \in \{0,1,2,\dots\}$$

denoted as $X \sim Pois(\lambda)$

↳ This formula is a result of Taylor Series form of e^λ

$$e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \xrightarrow{\text{which gives}} 1 = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}$$

When $X \sim Pois(\lambda) \rightarrow E(X) = \lambda$ & $Var(X) = \lambda$

Poission Limit of the Binomial (Law of Rare Events)

If $\lambda > 0$ & positive intergers n for which $\lambda/n < 1$ & $S_n \sim Bin(n, \lambda/n)$ then, $\xrightarrow{\lim_{n\to\infty}} P(S_n = k) = e^{-\lambda}\frac{\lambda^n}{k!}$ for all $k \in \{0,1,2,\dots\}$

↳ i.e. if S_n counts the number of successes in n independent trials & the mean $E(S_n) = \lambda$ does not change with n, then, as $n \to \infty$ the distribution of $S_n \rightarrow Poisson(\lambda)$

Scanned with CamScanner

If a random variable X counts the occurance of rare events that are not stronglu dependent upon each other. X can be approximated with a Poisson(λ) distribution where

$P(X=k)$ is close to $e^{-\lambda}\frac{\lambda^k}{k!}$ for $k \in \{0,1,2,\dots\}$ & $\lambda = E(X)$

Example: Suppose a factory averages 3 accidents a month What is the probability of 2 accidents next month?

↳ by Assuming independence between accidents X can be aproximated to $X \sim Pois(3)$

$P(X=2) = e^{-3}3^2/2! \approx 0.224 \rightarrow 24\%$ ✓

Use Poissoin Approximation when np^2 is small

↳ (use Norm. Approximation if $np(1-p) > 10$)

4.5: Exponential Distribution:

↳ The exponential distribution is the geometric distributions Continious cousin

A random variable X has the exponential distribution with the parameter λ if X has the density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$ denoted by $X \sim Exp(\lambda)$

Integrating to find the cdf yields: $\int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}$

To find the mean & variance evaluate the first two moments by Integration by parts:

$E(X) = \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$ $E(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$

Thus if $X \sim Exp(\lambda)$ then $\mu(X) = 1/\lambda$ & $Var(X) = 1/\lambda^2$

Scanned with CamScanner

Memoryless property of the exponential distribution:
If $X \sim Exp(\lambda)$. Then for any $s,t>0$ we have

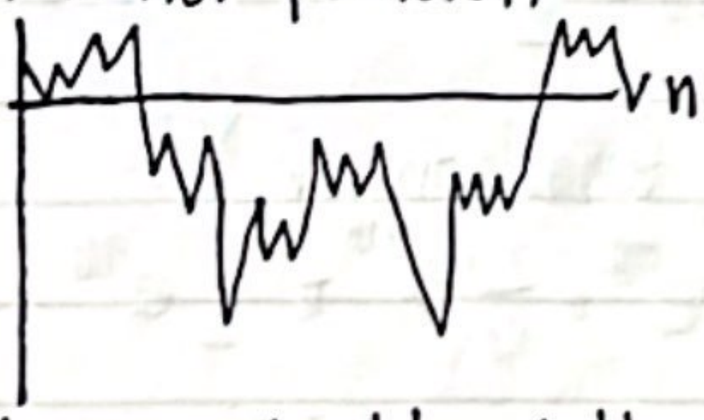
$$P(X>t+s \mid X>t) = P(X>s)$$

↳ This basically means that the probability that s has the same likelyhood following t as it does at the start of the distribution

Random Walk(Advanced Applications):

↳ Imagine someone who repeatedly flips a coin that gives heads with probability $0<p<1$. Every time she flips heads, she takes a step to the right. Everytime she flips tails she takes a step to the left

↳ "Let $X_1, X_2, X_3, \dots$ be indepent random variables, such that each X_j satisfies $P(X_j=1)=p$ and $P(X_j=-1)=1-p$ & $S_n = X_1 + \dots + X_n$. The random variable X_j is her j^{th} step & S_n is her position" → Simple random walk



Simulation of a simple random walk ↩

Using the central limit theorem we can analyze what the distribution of S_n looks like after a long time. Let T_n count the number of times the coin came up heads in the first n flips. Then $S_n = T_n - (n-T_n) = 2T_n - n$. Since $T_n \sim Bin(n,p)$ we get the mean $E(S_n) = 2E(T_n) - n = n(2p-1)$ & the variance $Var(S_n) = 4Var(T_n) = 4np(1-p)$

↳ By the CLT after a long time a simple random walk S_n approaches normality

Scanned with CamScanner

NO

DATE

Continuity Correction of the Binomial Normal Approximation:

If $S_n \sim Bin(n,p)$ then it can take only interger values, thus. If k_1 & k_2 are intergers then,

$$P(k_1 \leq S_n \leq k_2) = P(k_1 - 1/2 \leq S_n \leq k_2 + 1/2)$$

Since the interval $[k_1, k_2]$ contains the same intergers as $[k_1 - 1/2, k_2 + 1/2]$, expanding the range of the interval gives a better approximation of the binomial:

$$\Phi\left(\frac{k_2 + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - 1/2 - np}{\sqrt{np(1-p)}}\right)$$

5.1. Moment Generating Functions:

The moment generating function of an RV X is given by $M(t) = E(e^{tx})$ that describes the distribution of X by producing its moments sequentially

Example: Let X be a discrete RV with pmf given by

$P(X=1) = \frac{1}{3}$ $P(X=4) = \frac{1}{6}$ & $P(X=9) = \frac{1}{2}$

What is the moment generating function of X?

$$M_X(t) = E(e^{tX}) = \sum_k e^{tk} P(X=k) = \frac{1}{3}e^{-t} + \frac{1}{6}e^{4t} + \frac{1}{2}e^{9t}$$

For a Continious RV? X is a continious RV w/ pdf

$$f(x) = \begin{cases} \frac{e^x}{e-1} \rightarrow 0 < x < 1 \\ 0 \rightarrow \text{otherwise} \end{cases}$$

What is the moment generating function of X?

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \int_0^1 e^{tx} \frac{e^x}{e-1}\,dx$$

↳ If $t = -1$ then $M_X(t) = \frac{1}{e-1}$

↳ If $t \neq -1$ then $M_X(t) = \frac{e^{t+1} - 1}{(e-1)(t+1)}$

→ these are the $M_X(t)$ for X

Scanned with CamScanner

MATH 201 Midterm #2:
- 3.1: Probability distributions of RV
- 3.2: Cumulative distribution function
- 3.3: Expectation
- 3.4: Variance
- 3.5: Gaussian Distribution
- 4.1: Normal Approximation
- 4.2: Law of large numbers
- 4.3: Application of the normal approximation
- 4.4: Poisson Distributions

→ 201 Midterm #2: Useful Formulas

Distribution:	p.d.f/p.m.f.	cdf	μ	σ^2
General disc.	$P(X=k)$	$\sum_{k} P(X=k)$	$\sum_{k} P(X=k)k$	
General cont.				
Uniform				
Binomial				
Geometric				
Poisson				
Exponential				
Norm. (Gaussian)				

Scanned with CamScanner

Example: Moment Generating Function of the Poisson
Let $X \sim \text{Poiss}(\lambda)$ then

$$E(e^{tX}) = \sum_{k=0} e^{tk} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0} \frac{(e^t\lambda)^k}{k!} = e^{-\lambda} \cdot e^{e^t\lambda} = e^{\lambda(e^t-1)}$$

Moment Generating Function for the Normal:
Let $X \sim \text{Norm}(0,1)$ then.

$$E(e^{tX}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2}\, dx = e^{t^2/2}$$ → To generalize this formula. We can write $X = \sigma z + \mu$

$$\rightarrow E(e^{tx}) = e^{\mu t + \sigma^2 t^2/2}$$

Moment Generating function of the Exponential:
Let $X \sim \text{Exp}(\lambda)$ then.

$$E(e^{tx}) = \int_0^{\infty} e^{tx} \cdot \lambda e^{-\lambda x}\, dx = \lambda \lim_{b\to\infty} \int_0^b 1\, dx = M(t) = \begin{cases} \infty & \text{if } t \geq \lambda \\ \frac{\lambda}{\lambda - t} & \text{if } t < \lambda \end{cases}$$

When the moment generating function $M(t)$ of a random variable X is finite in an open interval around the origin the moments of X are given by $E(X^n) = M^{(n)}(0)$

Two random Variables X & Y are equal in distribution if $P(X \in B) = P(Y \in B)$ for all subsets B of $\mathbb{R} \rightarrow X \overset{d}{=} Y$

Scanned with CamScanner

6.1: Joint Distribution of Discrete Random Variables:

Let $X_1, X_2, \dots, X_n$ be discrete rvs, all defined on the same sample space. Their Joint Probability Mass Function is defined by: $P(k_1, k_2, \dots k_n) = P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n)$ for all possible values, $k_1, k_2, \dots k_n$ of $X_1, X_2, \dots, X_n$

Let $g: \mathbb{R}^n \to \mathbb{R}$ be a real-valued function of an n-vector
If $X_1, \dots, X_n$ are discrete rv with joint probability mass function of P then: $E(g(X_1, \dots, X_n)) = \sum_K g(k_1, \dots k_n)\, P(k_1, \dots, k_n)$

Marginal Probability mass fnc. ↳ provided this sum is defined

Let $P(k_1, \dots, k_n)$ be the joint probability mass function of $(X_1, \dots, X_n)$ let $1 \le j \le n$. Then the pmf of X_j is given by

$$P_{X_j}(k) = \sum P(\ell_1, \dots, \ell_{j-1}, k, \ell_{j+1}, \dots, \ell_n)$$

↳ This is called the marginal probability mass func.

The Multimoial Distribution:

Consider the trial with r possible outcomes. Outcome j appears with probability p_j. There are n independent repitions of this trial. For $j = 1, \dots, r$ let X_j denote the number of times we see outcome j in n trials. Find the joint dist.

↳ Enter the multinomial coeffients

$$\binom{n}{k_1, k_2, \dots, k_r} = \left(\frac{n!}{k_1! k_2! \cdots k_r!}\right) p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$$

the special case where $r = 2$ is the binomial

$(X_1, \dots, X_r) \sim \text{Mult}(n, r, p_1, \dots, p_r)$

Note: If X, Y are jointly continious then X & Y are continious random variables

↳ The converse is FALSE!

Scanned with CamScanner

6.2. Joint Distribution of Continious Random variables

↳ As with most discrete things in this class there is a continious counterpart

Random Variables $X_1, \dots, X_n$ are jointly continious if there exists a joint density function f on $\mathbb{R}^n$ such that for subsets $B \subseteq \mathbb{R}^n$

$$P((X_1, \dots, X_n) \in B) = \int \cdots \int_B f(x_1, \dots, x_n)\, dx_1 \cdots dx_n$$

Example of Discrete Random Variables:

Roll two fair 4 sided dice. Let X_1, X_2 be the outcomes

let $Y_1 = \min(X_1, X_2)$ & $Y_2 = |X_2 - X_1|$

a. find the joint pmf of (X_1, X_2)

↳ Possible values of X_1 & X_2: 1, 2, 3, 4

$$P(X_1 = k_1, X_2 = k_2) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}, \quad 1 \le k_1, k_2 \le 4$$

b. find the joint pmf of Y_1, Y_2

↳ possible values of $Y_1 = \text{Min}(X_1, X_2)$: 1, 2, 3, 4

↳ Possible values of $Y_2 = |X_2 - X_1| = 0, 1, 2, 3$

$P(Y_1 = 1, Y_2 = 0) = P(\min(X_1, X_2) = 1, X_1 = X_2) = 1/16$

$P(Y_1 = 2, Y_2 = 3) = P(\min(X_1, X_2) = 2, |X_2 - X_1| = 3) = 0$

P	$y_2=0$	$y_2=1$	$y_2=2$	$y_2=3$
$y_1=1$	1/16	2/16	2/16	2/16
$y_1=2$	1/16	2/16	2/16	0
$y_1=3$	1/16	2/16	0	0
$y_1=4$	1/16	0	0	0

c. find $E[Y_1, Y_2]$

$= E[\min(X_1, X_2)\ |X_2 - X_1|]$

$$\sum_{\ell=0}^{3} \sum_{K=1}^{4} K \cdot \ell\, P(K, \ell)$$

$$\frac{2}{16} + 2\frac{2}{16} + 3\frac{2}{16} + 4 \cdot 0 + 2\frac{2}{16} + 2 \cdot 2\frac{2}{16} + 2 \cdot 3 \cdot 0 + 2 \cdot 4 \cdot 0 + 3\frac{2}{16}$$

$= 15/8$

Scanned with CamScanner

NO

DATE

Example of Jointly Continuous Random Variables:

Let the density function of X & Y be given by

$$f(x,y) = \begin{cases} \frac{3}{2}(xy^2+y), & (x,y) \in [0,1]\times[0,1] \\ 0, & \text{otherwise} \end{cases}$$

a. verify that $f(x,y)$ is a density function

→ $f(x,y) \geq 0$ ✓

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_0^1\int_0^1 \frac{3}{2}(xy^2+y)\,dx\,dy$$

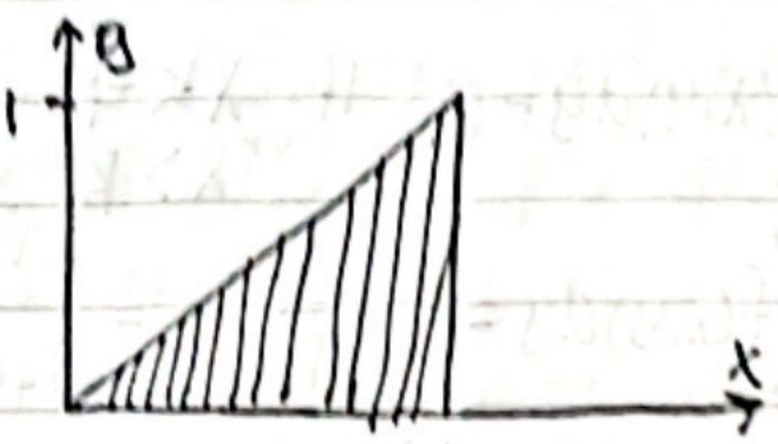
↓ treat y as a constant, then integrate w/ respect to y

$$\int_0^1 \left[\frac{3}{2}\left(\frac{x^2}{2}y^2+xy\right)\right]_0^1 dy$$

$$= \int_0^1 \frac{3}{2}\left(\frac{1}{2}y^2+y\right)dy = \frac{3}{2}\left[\frac{y^3}{6}+\frac{y^2}{2}\right]_0^1 = \frac{3}{2}\left(\frac{1}{6}+\frac{1}{2}\right) = 1 \checkmark$$

b. find $P(X<Y)$

$$P(X<Y) = \iint_{x<y} f(x,y)\,dx\,dy$$



$$= \int_0^1\int_0^y \frac{3}{2}(xy^2+y)\,dx\,dy$$

$$= \int_0^1 \left[\frac{x^2}{2}y^2+xy\right]_0^y dy = \int_0^1 \frac{3}{2}\left(\frac{y^4}{2}+y^2\right)dy = \frac{3}{2}\left(\frac{y^5}{10}+\frac{y^3}{3}\right)\Big|_0^1$$

$$P(X<Y) = \frac{3}{2}\left(\frac{1}{10}+\frac{1}{3}\right) = \boxed{13/20}$$

c. find $E[X^2Y] \rightarrow \iint x^2y\,f(x,y)\,dx\,dy \rightarrow \int_0^1\int_0^1 x^2y\left(\frac{3}{2}(xy^2+y)\right)dx\,dy$

$$\frac{3}{2}\int_0^1\int_0^1 x^3y^3+x^2y^2\,dx\,dy = \frac{3}{2}\int_0^1 \left[\frac{x^4}{4}y^3+\frac{x^3}{3}y^2\right]_0^1 dy$$

$$= \frac{3}{2}\int_0^1 \frac{1}{4}y^3+\frac{1}{3}y^2\,dy = \frac{3}{2}\left[\frac{y^4}{16}+\frac{y^3}{9}\right]_0^1 = \frac{3}{2}\left(\frac{1}{16}+\frac{1}{9}\right) = \boxed{25/96}$$

Scanned with CamScanner

NO.

DATE

Joint Uniform Distributions:

Recall: $X \sim \text{Unif}[a,b]$ if $f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$

Let $D \subseteq \mathbb{R}^2$ have finite area. Then (X,Y) is uniformly distributed on D if their joint density is

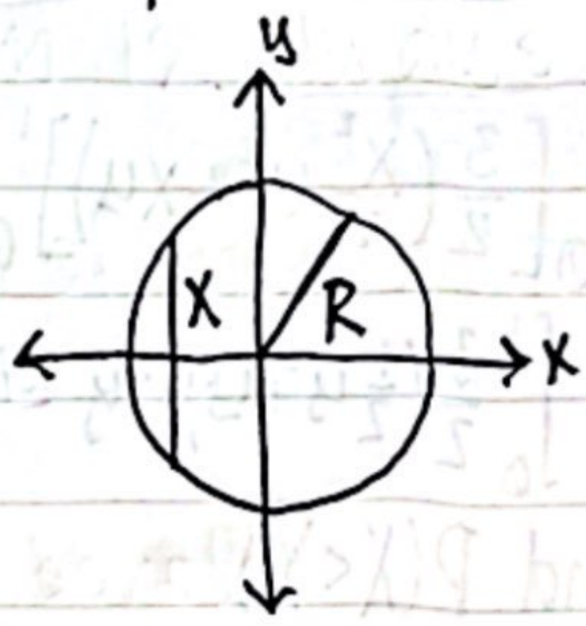
$$f(x,y) = \begin{cases} \frac{1}{\text{area}(D)} & \text{if } (x,y) \in D \\ 0 & \text{otherwise} \end{cases}$$

Example: Let (X,Y) be a uniform random point on the disk D with radius R centered at $(0,0)$

a. find the marginals $f(x)$ & $f(y)$

→ Joint density

$$f(x,y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } (x,y) \in D \\ 0 & \text{otherwise} \end{cases}$$



$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = 0 \quad \text{if } \begin{matrix} x < -R \\ \text{or } x > R \end{matrix}$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int \frac{1}{\pi R^2}\,dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2}\,dy = \frac{2\sqrt{R^2-x^2}}{\pi R^2}$$

↳ by symetry $f_Y(y) = \frac{2\sqrt{R^2-y^2}}{\pi R^2}$ $\quad -R \le y \le R$

Scanned with CamScanner

6.3: Independence & Joint Distributions

Discrete Case: Let $P(k_1,\dots,k_n)$ be joint p.m.f. of discrete RVs $X_1,\dots,X_n$. Let $p_{X_i}(k)=P(X=k)$. Then: $X_1,\dots,X_n$ are independent if & only if and only if

$$P(k_1,\dots,k_n)=p_{X_1}(k_1)\cdots p_{X_n}(k_n)$$

Example: Roll two fair dice let X_1, X_2 be the outcomes & $S=X_1+X_2$ are X_1 & S independent

↳ NOPE! $\rightarrow P(k,\ell)=P_{X_1,S}(k,\ell)$

$$P_{X_1,S}(1,12)=P(X_1=1,S=12)=0$$
$$\neq \frac{1}{6}\cdot\frac{1}{36}=P(X_1=1)P(S=12)=P_{X_1}(1)P_S(12)$$

Continious Analouge

Thm: Let $X_1,\dots,X_n$ be random variables on the same sample Space Ω

1. If joint density of $X_1,\dots,X_n$ is $f(x_1,\dots,x_n)=f_{X_1}(x_1)\cdots f_{X_n}(x_n)$ then $X_1,\dots,X_n$ are independent
2. Conversely: $X_1,\dots,X_n$ independent, then $f(x_1,\dots,x_n)=f_{X_1}(x_1)\cdots f_{X_n}(x_n)$

Example: Suppose X, y have joint density

$$f(x,y)=\begin{cases}\frac{7}{\sqrt{2\pi}}e^{\frac{-x^2}{2}-7y}, & \text{if } -\infty<x<\infty \quad y>0\\ 0 & \text{otherwise}\end{cases}$$

a. Are X and Y indepent? $\rightarrow$ Probably

$$\frac{7}{\sqrt{2\pi}}e^{\frac{-x^2}{2}-7y}=\frac{7}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}e^{-7y}=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}7e^{-7y}$$

↳ yes! $f(x,y)$ product of standard normal & Exp(7)

Scanned with CamScanner

NO.

DATE

8.1: Linearity of Expectation:

Let $g_1, g_2, \cdots, g_n$ be single variable functions & $X_1, X_2, \cdots, X_n$ random variables defined on the same sample space. Then,

$$E[g_1(X_1) + g_2(X_2) + \cdots + g_n(X_n)] = E[g_1(X_1)] + E[g_2(X_2)] + \cdots + E[g_n$$

& provided all the expectations are finite

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

This means the expectation of a sum is <u>always</u> the sum of expectations

↳ Note: This fact holds regardless of the assumption that the variables are Independent

Linearity of Expectation Example (Discrete):

Adam must pass both a written test & a road test to get his license. Each time he takes the road test he passes $7/10$ of the time & $4/10$ of the time he passes the written one. What is the total expected number of times he will have to take the tests before he passes?

written: $X_1 \sim Geom(\frac{4}{10})$ Then, $X = X_1 + X_2$

road: $X_2 \sim Geom(\frac{7}{10})$

$$E(X) = E(X_1) + E(X_2] = \frac{5}{2} + \frac{10}{7} = \frac{55}{14} \approx 3.93 \text{ times!}$$

Scanned with CamScanner

The Indicator Method (Linearity of Expectation)

Let $X \sim Bin(n, p)$ → Recall: a binomial RV counts the number of successes among n independent trials with p chance of a success

we can think of each trial as its own sample $(X_1, X_2, \cdots, X_n)$

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n)$$
$$= p + p + \cdots + p = np \checkmark$$

If a non-negative discrete RV X can be written as a sum of indicator variables, then E(X) is the sum of the expectations of the indicators

Example: There are n guests at a party. Assume that each pair of guests know each other with probability $\frac{1}{2}$ Let X denote the number of groups size 3 where everybody knows each other. Find E(X)

↳ For 3 distinct numbers $1 \leq a < b < c \leq n$ let $I_{a,b,c}$ denote the indicator of the event that guests a, b, & c all know each other → $X = \sum_{a,b,c} I_{a,b,c}$ • $E(X) = \sum_{a,b,c} E[I]$

$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$ ← P(the first 3 guests know each other)

$E(X) = \frac{1}{8} \cdot \frac{n}{3}$ ✓

Scanned with CamScanner

NO.

DATE

8.2: Expectation & Independence

Variance of a sum of Independent RVs:

↳ If $X_1, \dots, X_n$ are independent random variables with finite variances, then:

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)$$

Example: Variance of the Binomial Distribution

Recall: $X \sim \text{Bin}(n, p)$ counts the number of successes

$X = X_1 + X_2 + \dots + X_n$ ↓ Where we consider X as the sum of independent trials

$$\text{Var}(X) = \text{var}(X_1) + \dots + \text{Var}(X_n) = np(1-p)$$

Sample Mean & Sample Variance:

↳ these are examples of functions of random variables & better known as statistics

Expectation & Variance of the Sample Mean

↳ In a sequence of Independent RVs

$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ denotes the average of the first n random variables

This is the definition of the sample mean (of the first n observations). As a result of this:

$$E[\bar{X}_n] = \mu \quad \& \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

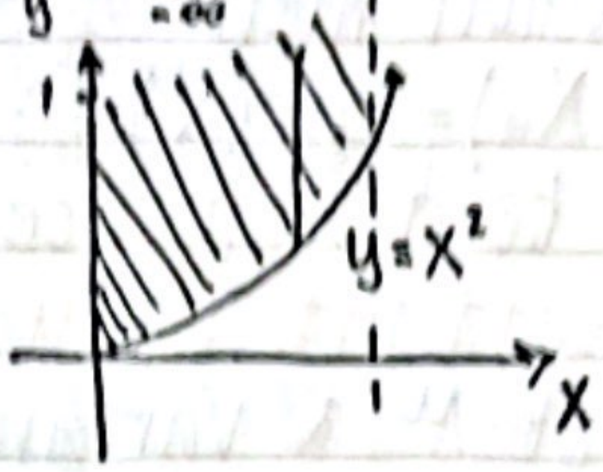
↑ This fact makes the sample mean a "unbiased estimator"

Scanned with CamScanner

Example of Jointly Continous:
Suppose X, Y have joint density. Find the marginal distributions.

$$f(x,y) = \begin{cases} 2xe^{x^2-y} & 0 < x < 1 \\ 0 & \text{else} \end{cases}$$

$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = 0$ if $x \notin (0,1)$



When $0 < x < 1$

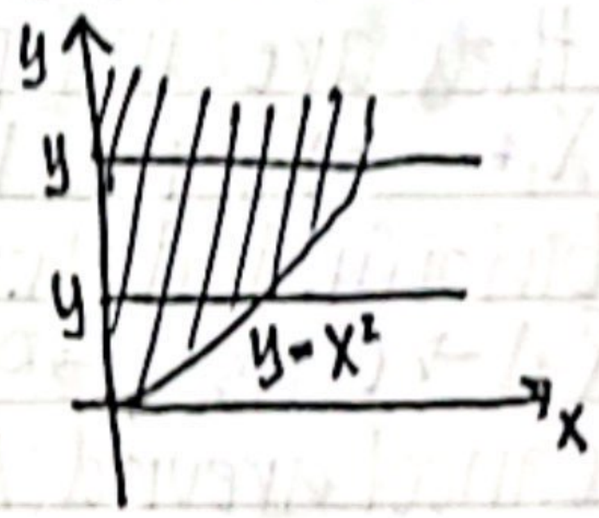
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_{x^2}^{\infty} 2xe^{x^2-y}\,dy$$

$$= 2xe^{x^2}\int_{x^2}^{\infty} e^{-y}\,dy = 2xe^{x^2}\left[\frac{e^{-y}}{-1}\right]_{x^2}^{\infty} = 2xe^{x^2}e^{-x^2} = 2x$$

$$f_X(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{else} \end{cases}$$

Now for $f_Y(y)$ $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$ Break into Cases:



When $y \geq 1$:

$$f_Y(y) = \int_0^1 f(x,y)\,dx = \int_0^1 2xe^{x^2-y}\,dx$$

$$= e^{-y}\int_0^1 2xe^{x^2}\,dx = e^{-y}\int_0^1 e^u\,du$$

$u = x^2$

$$= e^{-y}(e-1)$$

When $0 < y < 1$:

$$f_Y(y) = \int_0^{\sqrt{y}} 2xe^{x^2-y}\,dx = e^{-y}\int_0^{\sqrt{y}} 2xe^{x^2}\,dx = 1 - e^{-y}$$

W

Scanned with CamScanner

8.4. Covariance & correlation:
↳ measures simple dependence
↳ Joint variance

If $X_1, \dots, X_n$ are independent random variables with finite expectation then

$$Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n)$$

Let $X_1, X_2, X_3, \cdots$ be a sequence of random variables which are indepent and ~~[illegible]~~ Identically distributed with a mean of μ & variance σ^2, set

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

Then. $E[\bar{X}_n] = \mu$ & $Var(\bar{X}_n) = \sigma^2/n$

↳ Note: $\bar{X}_n$ is the "sample mean" & since $E[\bar{X}_n] = \mu$
We can say $\bar{X}_n$ is an unbiased estimator

↳ The elements of $\bar{X}_n$ are not (necesarilly) bernoulli but they are independent & Identically distributed
↳ $X_1 + \cdots + X_n$ is a kind of a generalization of the binomial distribution

↳ $Var(\bar{X}_n) \to 0$ as $n \to \infty$, so the more fluctuations will cancel around μ

Scanned with CamScanner

8.2: Covariance

Let X, Y be random variables...
With mean μ_X & μ_Y respectively...
The Covariance of X & Y is given by:

$$\text{cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \text{ or}$$
$$\text{cov}(X,Y) = E[(XY)] - \mu_X\mu_Y$$

If $\text{cov}(X,Y) > 0$ they are positively correlated.
Note: If X & Y are independent their covariance is zero
BUT! Just because the covariance of X & Y is zero does not necessarily mean they are Independent

Example: Let $A, B \subset \Omega$ be events and let I_A & I_B be their indicator random variables such that

$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{else} \end{cases}, \text{ & } I_B(\omega) = \begin{cases} 1 & \omega \in B \\ 0 & \text{else} \end{cases}$$

Find the covariance of I_A & I_B
↳ Note: $I_A(\omega)I_B(\omega) = 1 \longleftrightarrow \omega \in A \text{ & } \omega \in B \longleftrightarrow \omega \in A \cap B$

So... $\text{cov}(I_A, I_B) = E[I_A I_B] - E[I_A]E[I_B]$
$= P(A \cap B) - P(A)P(B)$
$= P(B)\left(\frac{P(AB)}{P(B)} - P(A)\right) = P(B)\left(P(A|B) - P(A)\right)$

Note: Covariance is bilinear

Scanned with CamScanner

NO.

DATE

Let X & Y be random variables with finite positive variance
The correlation of X & Y is

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$$

$$\frac{a}{|a|} corr(x,y) = corr(ax+b, Y)$$

Cauchy-Schwartz Inequality: $-1 \leq Corr(X,Y) \leq 1$
↳ $|corr(X,Y)| = 1$ if & only if $Y = aX+b$ for some $a \neq 0$
$b \in \mathbb{R}$

Properties of the covariate:

i. $Cov(X,Y) = Cov(Y,X)$
ii. $Cov(aX+b, Y) = a\,Cov(X,Y)$ for any real numbers
iii. For any RV X_i Y_j & real numbers a, b

$$Cov\left(\sum_{i=1}^{m} a_i X_i, \right.$$

Scanned with CamScanner

NO.

DATE

Moment Generating Functions:

↳ Let X be a random variable. the moment generating function of X is $M_X(t) = E[e^{tx}]$, $t \in \mathbb{R}$

M_X is not random. M_X is a function of a real variable

↳ M_X might only be finite on an interval about, or even only at 0 & $M_X(0) = E[1] = 1$

Example: Let X be discrete w/ p.m.f. $P(X=-10) = 1/3$, $P(X=2) = 1/2$, $P(X=0) = 1/6$

$$M_X(t) = E(e^{tx}) = \sum_k e^{kt} P(X=k) = e^{-10t} \cdot \frac{1}{3} + e^{0t} \cdot \frac{1}{6} + e^{2t} \cdot \frac{1}{2}$$

Example #2: Let $X \sim \text{Unif}[a,b]$ w/ p.d.f.

$$f(x) = \begin{cases} \frac{1}{b-a} & , x \in [a,b] \\ 0 & , \text{else} \end{cases} \qquad M_X(t) = E[e^{tx}] = \int_a^b e^{tx} \cdot \frac{1}{b-a} dx$$

$$= \frac{e^{bt} - e^{at}}{t(b-a)}, \quad t \neq 0$$

$$M_X(0) = 1 \quad \&$$

Example #3: Let $Z \sim N(0,1)$ find M_{Zt}

$$M_Z(t) = E(e^{tz}) = \int_{-\infty}^{\infty} e^{tz} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \left(e^{t^2/2} e^{-z^2/2 + tz - t^2/2} \right) dz = e^{t^2/2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}} dz}_{\text{density of } N(t,1)}$$

$$= e^{t^2/2} \cdot 1$$

In General: $X \sim N(\mu, \sigma^2)$ has $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

Example #4: Let $X \sim \text{Poisson}(\lambda)$, find $M_X(t)$

$$M_X(t) = E(e^{tx}) = \sum_{k=0}^{\infty} e^{kt} P(X=k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \lambda)^k}{k!} = e^{-\lambda} e^{e^t \lambda} = e^{\lambda(e^t - 1)}$$

Scanned with CamScanner

Properties of Moment Generating Functions:

↳ Let X be a r.v., then $E[X]$ & $M_X(t)$ are related by

$$\frac{d}{dt}(M_X(t)) = \frac{d}{dt}(E[e^{tx}]) = E\left[\frac{d}{dt}(e^{tx})\right] = E[Xe^{tx}] \text{ so...}$$

$$M'_X(0) = E[Xe^0] = E[X]$$

↳ If $M_X(t)$ is finite on an interval centered at 0. Then $E[X^n] = M_X^{(n)}(0)$ → Example: suppose X is a random variable w/ $M_X(t) = e^t + e^{2t} + e^{3t}$

find $E[X]$ & $E[X^2]$

$E[X] = M'_X(0)$, & $M'_X(t) = e^t + 2e^t + 3e^t$

$= e^0 + 2e^0 + 3e^0 = 1 + 2 + 3 = 6$

$E[X^2] = M''_X(0) = 1 + 4 + 9 = 14$

↳ Two random variables X & Y are equal in distribution if $P(X \in B) = P(Y \in B)$ for all $B \subseteq \mathbb{R}$ (denoted $X \overset{d}{=} Y$

↳ If X, Y are discrete w/ same p.m.f. then $X \overset{d}{=} Y$

If X, Y are continious w/ the same p.d.fs then $X \overset{d}{=} Y$

Theorem: Let X, Y be random variables suppose that M_X, M_Y are finite near 0, & $M_X(t) = M_Y(t)$ for all $t \in (-\delta, \delta)$, some $\delta > 0$ then $X \overset{d}{=} Y$

Example: Suppose $M_X(t) = \frac{1}{10}e^{-4t} + \frac{1}{5} + \frac{7}{10}e^{3t}$
find the distribution of X

Note: If X has finitely many possible values, then $M_X(t) = \sum_k e^{kt} p(x=k)$ is a finite sum. So,

$M_X(t)$ is a finite sum of exponentials
$\Rightarrow X$ has finitely many values &

$P(X=-4) =$ coeff of $e^{-4t} = 1/10$
$P(X=0) =$ coeff of $e^{0t} = 1/5$
$P(X=3) =$ coeff of $e^{3t} = 7/10$ ✓

Scanned with CamScanner

8.3: Moment Generating Functions & Sums

Fact: suppose X, Y are independent, Then $M_{X+Y}(t) = M_X(t) M_Y(t)$

Example: Let $X_1, \cdots, X_{15}$ be independent random variables w/ p.m.f. $P(X_i = -1) = 1/3$, $P(X_i = 0) = 1/2$, $P(X_i = 1) = 1/6$. Find $M_S(t)$ where $S = X_1 + \cdots + X_{15}$

$$\rightarrow M_{X_i}(t) = \sum_k e^{kt} P(X_i = k) = e^{-t}\tfrac{1}{3} + e^{0t}\tfrac{1}{2} + e^{t}\tfrac{1}{6}$$

$$= \tfrac{1}{3}e^{-t} + \tfrac{1}{2} + \tfrac{1}{6}e^{t}, \quad \text{so} \cdots$$

$$M_S(t) = M_{X_1}(t) \cdots M_{X_{15}}(t) = \left(\tfrac{1}{3}e^{-t} + \tfrac{1}{2} + \tfrac{1}{6}e^{t}\right)^{15}$$

Example#2: Let $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ be independent
Find the distribution of $X+Y$

$$M_X(t) = e^{\mu_X t + \frac{1}{2}\sigma_X^2 t^2}, \quad M_Y(t) = e^{\mu_Y t + \frac{1}{2}\sigma_Y^2 t^2} \quad \text{so...}$$

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

$$= \exp\left(\mu_X t + \tfrac{1}{2}\sigma_X^2 t^2\right) \exp\left(\mu_Y t + \tfrac{1}{2}\sigma_Y^2 t^2\right)$$

$$= \exp\left((\mu_X + \mu_Y)t + \tfrac{1}{2}(\sigma_X^2 + \sigma_Y^2)t\right)$$

$= $ Mgf of $N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ so $X+Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

If $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, $\cdots$, $X_n \sim N(\mu_n, \sigma_n^2)$ are Independent, & $a_1, \cdots, a_n, b \in \mathbb{R}$, then... $a_1X_1 + \cdots + a_nX_n + b \sim N(\mu, \sigma^2)$

Where $\mu = a_1\mu_1 + \cdots + a_n\mu_n + b$, $\sigma^2 = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2$

Scanned with CamScanner

Estimating Tail Probabilities:

Goal: Estimate $P(|X| \geq a)$ for general random variables X and $a > 0$

Fact: Suppose X, Y are random variables (on the same sample space Ω). If $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$ then $E[X] \leq E[Y]$

Markov's Inequality:

Let X be a non-negative random variable

$$P(X \geq c) \leq \frac{E[X]}{c} \quad \text{for all } c > 0$$

↳ Markov's Inequality is simple, but often not very good

Chebyshev's Inequality:

Let X be a random variable with finite mean μ & a variance of σ^2. Then, for any $c > 0$

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Markov's/Chebyshev's Inequalities are great when we know very little about our random variable. If we know the distribution, we can usually do better

Example: Suppose $X \sim Ber(p)$, so $P(X=1) = p$, $P(X=0) = 1-p$

Then... $p(X \geq 1) \leq E[X] = p$ ↑ markov

$p = P(X \geq 1/10) \leq \frac{E[X]}{1/10} = 10p$ ← markov

Scanned with CamScanner

NO

DATE

More Markov & Chebyshev's Inequality Examples:

#1. Suppose a coffee shop averages $2.000 a day in Sales. What can we say about the probability that they sell more than $3.000 tomorrow?

Let X = Sales tomorrow $E[X] = 2000$ so...

$$P(X \geq 3000) \leq \frac{E[X]}{3000} = \frac{2000}{3000} = 2/3$$

(Markov)

#2. Suppose that the variance in sales is $1.000 What can be said about the probability that the cafe has sales between $1.800 & $2,200 tommorow? What about than $3.000

Let X = Sales. $E[X] = 2000$, & $Var(X) = 1.000$

$$P(1800 \leq X \leq 2200) = P(-200 \leq X - 2000 \leq 200)$$
$$= P(|X - E[X]| \leq 200)$$
$$= 1 - P(|X - E[X]| > 200)$$
$$\geq 1 - 1/40 = 39/40$$

$$P(X \geq 3000) = P(X - 2000 \geq 1000)$$
$$\leq P(|X - 2000| \geq 1000) \leq 1/1000$$

Scanned with CamScanner

NO

DATE

9.2: Law of Large Numbers.

Let $X_1, X_2, X_3, \dots$ be a sequence of independent random variables with finite $E[X_i] = \mu$ & finite $\text{var}(X_i) = \sigma$?

Let $S_n = X_1 + \dots + X_n$ Then $\lim_{n\to\infty} P(|\frac{S_n}{n} - \mu| < \varepsilon) = 1$

Example: Suppose you have a weighted die, & you want to estimate the expected ~~value~~ value of the roll by rolling the die many times & averaging the outcomes. How many times do you need to roll to be 99% sure that your average is within 0.1 of the true expected value of a roll?

Let X_i = outcome of the i-th roll, $S_u = X_1 + \dots + X_n$

μ = expected roll ← What we want to estimate

$\mu = E[X_i]$, $\sigma^2 = \text{var}(X_i)$

Idea: $P(|\frac{S_n}{n} - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2} \geq 0.99$

↑ Find ↑ 1/10 ↑ with an n such that

$0.01 \geq \sigma^2/n\varepsilon^2 \Rightarrow n \geq \sigma^2/0.01 \cdot \varepsilon^2 = \frac{\sigma^2}{(0.01)(0.01)}$

↑ We dont know σ^2, but can bound from above: $|X| \leq b \Rightarrow \text{var}(X) \leq b^2$ so, in order for this equation to hold its enough for

$n = \frac{b^2}{(0.01)^2} = 360{,}000 \geq \frac{\sigma^2}{(0.01)^2}$

Scanned with CamScanner

Central Limit Theorem:

Let $X_1, X_2, X_3, \ldots$ be a sequence of indepent random variables with finite $E[X_i] = \mu$ & finite variance $Var(X_i) = \sigma^2$

Let $S_n = X_1 + \cdots + X_n$ Then for $-\infty \leq a \leq b \leq \infty$

$$\lim_{n\to\infty}\left(P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right)\right) = \Phi(b) - \Phi(a) = \int_a^b \frac{e^{-s^2/2}}{\sqrt{2\pi}}\, ds$$

note: $n\mu = E[S_n]$, $\sqrt{Var(S_n)} = \sigma\sqrt{n}$

Example: Roll 1,000 dice & sum the values

Estimate $P(\text{sum} \geq 3,600)$

$X_1, X_2, \ldots, X_{1000}$ = outcomes of dice rolls

$\mu = E[X_i] = 7/2$ $\quad \sigma^2 = Var(X_i) = E[(X_i - 3.5)^2] =$

$S = \sum_{j=1}^{1000} X_j$ $\quad \sum_{k=1}^{6}(k_i - 3.5) - 1/4 = 35/12$

Thus: $P(S \geq 3600) = P\left(\frac{S - n\cdot\mu}{\sigma\cdot\sqrt{n}} \geq \frac{3600 - n\cdot\mu}{\sigma\sqrt{n}}\right)$

$= P\left(\frac{S - 3500}{\sqrt{35\cdot 1000/12}} \geq \frac{100}{\sqrt{35\cdot 1000/12}}\right)$

$\approx P(Z \geq 1.8516) \approx 1 - \Phi(1.8516) \approx 0.03$

Scanned with CamScanner

7.2: Exchangeable Random Variables:

Two random variables X_1 & X_2 are exchangeable if for any $B \subseteq \mathbb{R}^n$ & any permutation $(k_1, \dots, k_n)$ of $(1, \dots, n)$,

$P((X_1, \dots, X_n) \in B) = P((X_{k_1}, \dots, X_{k_n}) \in B)$

↳ In other words: $(X_1, \dots, X_n)$ & $(X_{k1}, \dots, X_{kn})$ have the same distribution

If X & Y are jointly continious w/ density $f(x,y)$ then,

↳ X & Y are exchangeable

Independent
Identically
Distributed
} abbreviated IID →

"Let $1 \le k \le m$ & let $X_1, \dots, X_k$ be outcomes of successive draws without replacement from a set with n elements Then $X_1, \dots, X_k$ are exchangeable

Note: This does not produce independent r.v.s in ordered samples!

Example: Draw 23 cards one-by-one from a standard deck & Let

$$X_i = \begin{cases} 1 & \text{if the ith draw is an ace} \\ 0 & \text{otherwise} \end{cases}$$

Then $X_1, \dots X_{23}$ are exchangeable

Properties of Exchangeable RVs: Let $X_1, \dots, X_n$ be exchangeable

1. For any $g: \mathbb{R} \to \mathbb{R}$, $g(X_1), \dots, g(X_n)$ are exchangable
2. The RVs X_i are identically distributed
3. For any $f: \mathbb{R}^k \to \mathbb{R}$, $E[f(X_1, \dots, X_k)] = E[f(X_{i1}, \dots, X_{ik})]$ for any permutation $(i, \dots, ik)$ of $(1, \dots, k)$

Scanned with CamScanner

Exchangeable RV Example:

Suppose we deal 23 cards from a standard deck,

let $X_i = \begin{cases} 1 & i\text{-th card is an ace} \\ 0 & \text{else} \end{cases}$

Note: $X_1, X_2, \ldots, X_{23}$ are exchangable

1. Verify that X_1 & X_2 have the same distribution
2. Find the expected number of aces
3. Calculate $E[X_{12}, X_{17}]$

1. It's enough to check $P(X_1=1) = P(X_2=1)$, as this then implies $P(X_2=0) = P(X_1=0)$

↳ $P(X_2=1) = P(X_1=1 \mid X_1=0)P(X_1=0) + P(X_2=1 \mid X_1=1)P(X_1=1)$

$= \frac{4}{51}\cdot\frac{48}{52} + \frac{3}{51}\cdot\frac{4}{52} = \frac{4(48+3)}{52\cdot 51} = \frac{4}{52} = P(X_1=1)$ ✓

2. X = # of aces drawn

$X = X_1 + \cdots + X_{23}$

$E[X] = E[X_1] + \cdots + E[X_{23}] = 23E[X_1]$ ← by property #2

↑ by the linearity of expectation

$= 23P(X_1=1) = 23\cdot\frac{4}{52} \approx 1.77$ ✓

3. $E[X_{11}, X_{22}]$

$f: \mathbb{R}^{23} \to \mathbb{R}$, $f(X_1, X_2, \ldots, X_{23}) = X_1X_2$ Then...

$E[X_1X_2] = P(X_1=1, X_2=1) = \frac{4}{52}\cdot\frac{3}{51}$ ✓

Scanned with CamScanner

More moment generating functions:

#1: Suppose X is a random variable w/ $M_X(t) = e^t + e^{2t} + e^{3t}$

Find $E[X]$ & $E[X^2]$

$E[X] = M'_X(0)$ & $M'_X(t) = e^t + 2e^t + 3e^{3t}$

$= e^0 + 2e^0 + 3e^0 = 1+2+3 = 6$

$M''_X(t) = e^t + 4e^{2t} + 9e^{3t}$

$= 1 + 4 + 9 = 14$

#2: $X \sim Exp(\lambda)$ Find $E[X^n]$ for $n \geq 1$, find $M_X(t)$

$$M_X(t) = E[e^{tx}] = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx$$

$$= \lambda \int_0^\infty e^{x(t-\lambda)} dx = \lambda \left[\frac{e^{x(t-\lambda)}}{t-\lambda} \right]_{x=0}^{\infty} = \lambda \left[0 - \frac{1}{t-\lambda} \right]$$

$$= \frac{\lambda}{\lambda - t} \quad \text{valid for } t < 2$$

$$\frac{d}{dt} M_X(t) = \frac{d}{dt}\left(\frac{\lambda}{\lambda - t}\right) \leftarrow$$

$$= \frac{\lambda}{(\lambda - t)^2}$$

Sums of Independent Normal Random Variables:

Let X_1 be a normal independent RV with mean = 2 & var = 3

Scanned with CamScanner