

Tuesday
24-9-19

Statistics is a science that helps us make decisions and draw conclusions in the presence of variability

variability: successive observations at a system are not the same.

- my goal is to lessen the variables, eliminating it is impossible.

since my factors affect my response $\Rightarrow$ variable $\Rightarrow$ probability

we can't eliminate variability so we are going to:

- Describe + Quantify + Reduction

Types of models:

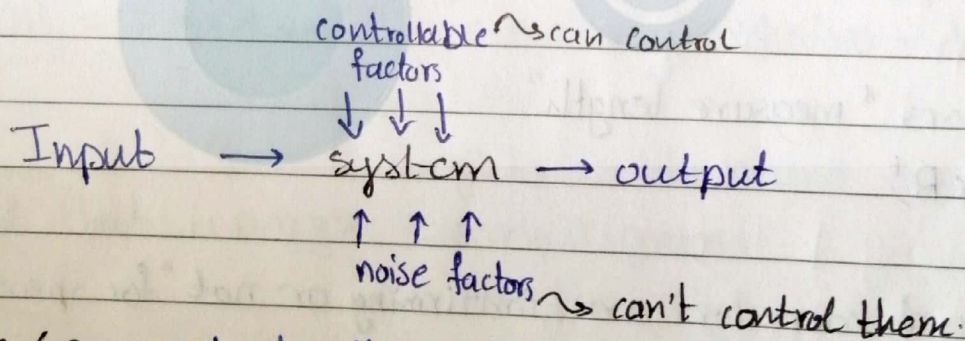
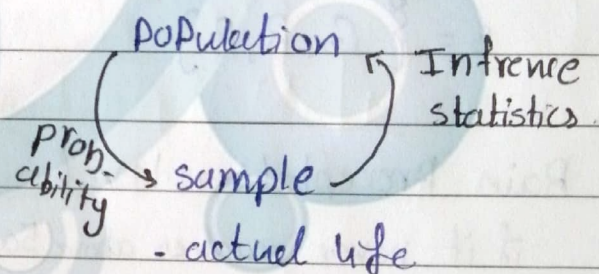
mechanistic model

physical laws
model design

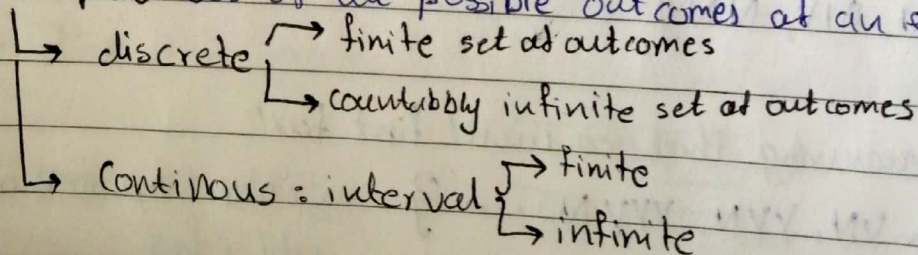
ex: $F=ma$

- quick and direct

Empirical model



Sample Space: set of all possible outcomes at an experiment (S)



- an event is a subset of sample space
- points of a sample space are called elementary events or outcome or "realization"

example: flipping a coin: $S = \{H, T\}$

Thursday

26-9-2019

- 3 machine parts, each is classified as either above or below the target specifications.

take above to be a, below, b

$S = \{aaa, aab, aba, baa, bba, abb, bab, bbb\}$ discrete finite

Number of news recorded at a high volume website in a day

$S = \{0, 1, 2, 3, \dots, n\}$ very large number countably infinite
 ↗ a vase with a cover.

- Selecting a ball from urn which has 3 balls: G, R, B

$S = \{G, R, B\}$

- Rain Precipitation level $S = R^+ = \{x | x \geq 0\}$

if it never reaches above 50 inches $S = \{x | 0 \leq x \leq 50\}$

- Two connectors "measure length"

$S = R^+ \times R^+$

in showing if connectors are confirming or not "for specifications"

$S = \{YY, YN, NY, NN\}$

Y, N

~~Two connectors~~ measuring thickness until first fail

$S = \{N, \del{YN}, YN, YYN, YYYN, \dots\} \Rightarrow$ i will keep measuring until i fail

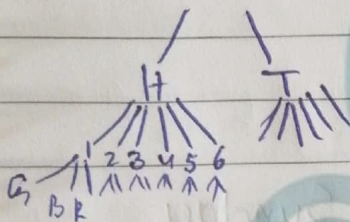
Graphical representation of sample space which is called Tree diagram

experiment is done in several steps

n_1 ways at first steps $\Rightarrow n_1$ branches

n_2 ways at second steps $\Rightarrow n_2$ branches for each branch in n_1

ex: tossing a coin, rolling a dice and choosing a ball from urn



$$S = \{H1G, H1B, H1R, H2G, \dots\}$$

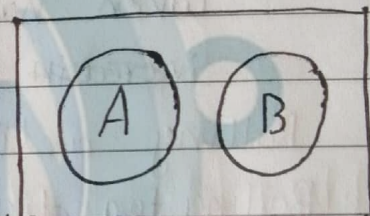
$$S = 36$$

Basics of Set Theory

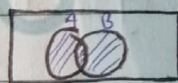
A
event

B
event

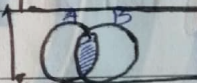
Venn diagram



union $A \cup B$



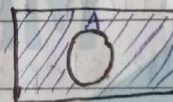
intersection $A \cap B$



$\emptyset$: empty set

complement

A^c



$$\Rightarrow S - A$$

If $A \cap B = \emptyset \Rightarrow A$ and B are mutually exclusive = disjoint

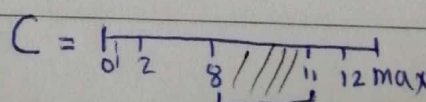
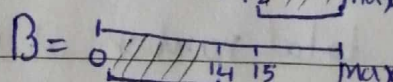
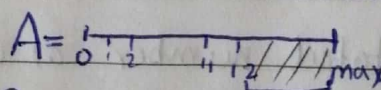
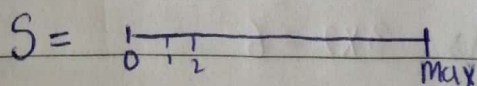
Digital scale to provide weights to the nearest gram

A: event that a weight exceeds 11 grams

B: is less than 15g

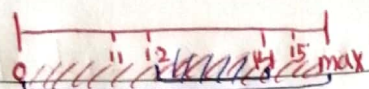
C: is greater than or equal to 8g and less than 12 grams.

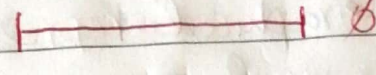
as i will not take 11

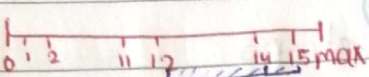


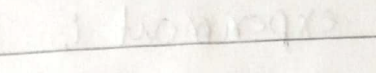
-Continued:

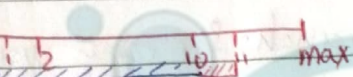
$B \Rightarrow$ or equal

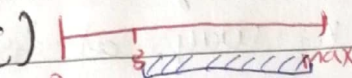
b) $A \cup B$ 

g) $A \cap B \cap C$ 

c) $A \cap B$ 

h) $B' \cap C$ 

d) A^c 

i) $A \cup (B \cap C)$ 

e) $A \cup B \cup C$ 

f) $(A \cup C)'$ 

"Counting techniques"

Sunday
29-9-2019

when number of outcomes from experiment
~~increases~~ increases.

Rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$

Rolling two dice: $S = \{(1, 1), (1, 2), \dots\}$ $36 \Rightarrow 6 \times 6$

technique 1 | Multiplication Rule

ex: operation with k steps, each step is done in a number of ways

- step 1 with n_1 ways
- step 2 with n_2 ways, and so on...

Q) How many total number of ways to do this operation?
 $= n_1 \times n_2 \times n_3 \dots n_n$

Q) Draw a marble from urn which has 3 of them: G, B, R
twice

- with replacement. total number of ways $3 \times 3 = 9$

without replacement. total number of ways: $3 \times 2 = 6$

~~6) AUB~~ ~~known~~

Now in how many ways can we select k distinct objects from n distinct objects

Technique 2] Permutations (order of items arrangement is important) $GAB \neq GBA$.

- How many possible ways to get k from n

$$n \times (n-1) \times (n-2) \times (n-3) \dots \times (n-k+1) \Rightarrow$$

$$= \frac{n(n-1)(n-2) \dots (n-k+1) \times (n-k)(n-k-1) \times \dots \times 2 \times 1}{(n-k)(n-k-1) \times \dots \times 2 \times 1} = \frac{n!}{(n-k)!} = {}^n P_k$$

example] 5 students, need to select: president, vice-president, secretary. order is important $\Rightarrow$ Permutations

$${}^5 P_3 = \frac{5!}{(5-3)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = 60$$

$$\frac{5}{5} \frac{4}{4} \frac{3}{3} = 5 \times 4 \times 3 = 6$$

- 5 students in 5 positions

$$\frac{5}{5} \frac{4}{4} \frac{3}{3} \frac{2}{2} \frac{1}{1} = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

$${}^5 P_5 = \frac{5!}{(5-5)!} = \frac{5!}{0!} = \frac{5!}{1} = 120$$

Now for a set of objects which contains k groups of identical items

$$\text{number of ways of ordering } n \text{ objects} = \frac{n!}{n_1! n_2! n_3! \dots n_k!}$$

$\sim 1-M \quad \sim 4-2 \quad \sim 4-5 \quad \sim 2-P$

ex] MISSISSIPPI

G_i will exclude the repeated identical

how many words can I write with 11 letters? (note I have repeated identical units)

$$= \frac{11!}{1! 4! 4! 2!}$$

3) Combinations $\leadsto$ order isn't important.

Tuesday

number of ways of selecting r objects from n without regard to order

1-10-19

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

ex) 4 cars, Honda, Toyota, Ford, BMW, number of ways of selecting two of them. $C_2^4 = \frac{4!}{2!2!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1 \times 2 \times 1} = 6$

ex) a group of seven engineers and five statisticians.

We need a committee of 5 people. In how many ways can we select the committee where we have 3 engineers and 2 statisticians.

$$C_3^7 * C_2^5$$

Probability: likelihood or chance of an outcome to occur for a event $A \Rightarrow P(A)$

* If A is an event that a student will achieve a B is a class.

$P(A) = 0.6 \Rightarrow$ likelihood a student will get B is 60%.

$\hookrightarrow$ 60% from students will get B

mathematical definition of probability is a function that assigns to each event A which is included (subset) in $S \Rightarrow A \subset S$ a real number $\Rightarrow P(A) \in [0, 1]$ such that (i) $P(A) \geq 0$ (ii) $P(S) = 1$ (iii) if $A_1, A_2, \dots$ are disjoint (mutually exclusive) $P(\cup A_i) = \sum P(A_i)$
 $\sum P(A_i) = P(A_1) + P(A_2) + P(A_3) + \dots$

1) Degree of belief: common sense. (weather)

2) frequency: repetition.

(flipping a coin repeatedly)

3) equally likely: N outcomes $\Rightarrow p(\text{each outcome}) = \frac{1}{N}$

ex) Urn contains 30 Red, 20 Green, 50 Blue.

What is the probability of choosing every marble with replacement

$$P(\text{choosing any marble}) = \frac{1}{100}$$

$$P(\text{Red}) = \frac{30}{100}$$

For a discrete sample space $P(E) = \sum_{i=1}^n P(\text{all outcomes in } E)$

Simple properties of probability:

1) $P(\emptyset) = 0$

2) If $A \subset B \Rightarrow P(A) \leq P(B)$

3) $0 \leq P(A) \leq 1$

4) $P(A') = 1 - P(A)$

5) If $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

6) If $A \cap B \neq \emptyset \Rightarrow P(A) + P(B) - P(A \cap B) \rightarrow$ total prob law.

ex) Batch contains six parts $\{a, b, c, d, e, f\}$ where f is defective. Two parts are selected randomly without replacement. What is the prob. that f is in the sample.

1) define the event: $E = f$ is in the sample.

$$P(E) = \frac{\text{number of samples that contain } f}{\text{number of possible samples}} = \frac{5}{15}$$

$$\rightarrow C_2^6 = \frac{6 \times 5 \times 4!}{2! \times 4!} = 15$$

$$P(E) = \frac{5}{15} = \frac{1}{3}$$

ex) problem 2-61 Nickel battery, charge state and probabilities for each state

Charge

Prob

$$P(E) = +2 + 3 + 4$$

0

0.17

$$= 0.25 + 0.33 + 0.55 = 0.83$$

+ 2

0.35

+ 3

0.33

+ 4

0.55

continued..

Thursday ♥

3-10-19

B) what is the prob that a cell has not a nickel charge greater than +3?

$$E = \{\text{charge not } > +3\} \text{ or } \{\text{charge } \nless +3\} = E\{\text{charge } \leq +3\} \\ = \{0, +3, +2\} \quad P(E) = 0.85$$

Addition Rule: A & B are two events

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

example) Prob of raining today = 20%

Prob of raining tomorrow = 30%

Prob of raining today + tom = 6%

Prob of raining today or tom =

$$A: \text{today} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$B: \text{tomorrow} \quad 20 + 30 - 6 = 44\%$$

Prob 2-78 | Stands at copper wire, 100 wires were analyzed for strength and conductivity.

conductivity

	Strength high	low
high	74	8
low	15	3

a) what is the prob that a randomly selected strand is high cond and high strength

$$P(H_c \cap H_s) = 74/100$$

b) strength is low OR cond is low $P(L_c \cup L_s) = 15 + 8 + 3/100 = 26/100$

c) Two events: E_1 = Strand has low conductivity.

E_2 = Strand has low strength

are these two events (E_1 & E_2) mutually exclusive?

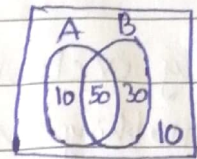
No, they intersect $P(E_1 \cup E_2) \stackrel{?}{=} P(E_1) + P(E_2)$

$$= \frac{8+3}{100} + \frac{15+3}{100} \\ = \frac{29}{100} \neq \frac{26}{100}$$

example) In a suburb, 60% of all households subscribe to the news paper published in a nearby city, 80% subscribe into the local news paper, and 50% subscribe to both newspapers.

$A = \{ \text{house hold subscribed to } \overset{\text{nearby}}{\text{city paper}} \}$

$B = \{ \text{house hold subscribed to local paper} \}$



$$P(A) = 0.6 \quad P(B) = 0.8 \quad P(A \cap B) = 0.5$$

a) the house hold subs to at least one of the two newspapers.

$$P(A \cup B) = 0.6 + 0.8 - 0.5 = 0.9$$

b) subs to exactly one of the two papers.

$\{ \text{at least 1} \} = \{ \text{exactly 1} \} \cup \{ \text{both} \}$

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

$$0.9 = \{ \text{exactly 1} \} + 0.5$$

$$P(\text{exactly 1}) = P(A \cup B) - P(A \cap B)$$

$$= P(A) + P(B) - 2P(A \cap B)$$

TIPS

at least $x \Rightarrow x$ or more

$$\{ x \geq x \}^c = \{ x < x \}$$

$\{ \text{at least } x \}^c = \text{less than } x$

$\{ \text{at most } x \} = x \text{ or less}$

$$\{ x \leq x \} = \{ x \geq x \}$$

$\{ \text{at most } x \}^c = \{ \text{more than } x \}$

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

Sunday

6-10-2019

mutually exclusive events: $E_1, E_2 \Rightarrow$ meaning they have nothing in common and might not include all the sample space

mutually exclusive and collectively exhaustive $\Rightarrow$ meaning they have nothing in common but they cover the whole sample space

*** Conditional Probability** when probabilities change as additional info becomes available

A & B Knowledge at the occurrence of an event A will change the probability at B.

$$P(B|A) \text{ Prob at B given A has happened} = \frac{P(A \cap B)}{P(A)}$$

example) Flipping two coins $S = \{HH, HT, TH, TT\}$

E_1 : at least one head $S = \{HH, HT, TH\}$

E_2 : two heads. $S = \{HH\}$

what is the probability of getting two heads if given at least one head is tossed?

$$P(E_2|E_1) = \frac{P(E_2 \cap E_1)}{P(E_1)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Problem 2-87 | 100 Disks are analyzed for shock and scratch

	Shock	
	high	low
Scratch	high 70	low 9
	low 16	5

A: event that disk has high shock

B: event that disk has high scratch.

a) $P(A) = \frac{70+16}{100}$

b) $P(B) = \frac{70+9}{100}$

c) $P(A|B) = \frac{70}{79} = \frac{70/100}{79/100}$

d) $P(B|A) = \frac{70}{86}$

* Multiplication and Total Probability Law

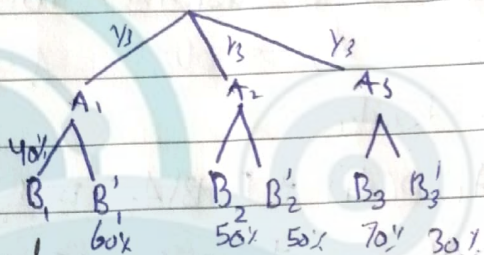
$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B) \times P(B)$$

$$\hookrightarrow = P(B|A) \times P(A)$$

Three Airlines A_1, A_2, A_3 , late take off $A_1 = 40\%$, $A_2 = 50\%$, $A_3 = 70\%$

A : choosing an airline

B : it being late.



a) Prob of selecting airline and

late take off $P(A_1 \cap B_1) = 40/100 \times 1/3 = 4/30$

b) Prob of late taking off (nevermind the airline)

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$$

$$= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)$$

$$= 40/100 \times 1/3 + 50/100 \times 1/3 + 70/100 \times 1/3 = 16/30$$

c) given that I was late taking off, what is the Prob that I selected Airline one?

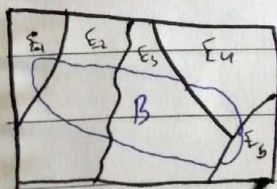
$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{4/30}{16/30} = 1/4$$

Tuesday
8-10-19

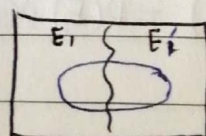
Law of Total Probability

practically used to calculate probability of individual events given it's probability and several conditions.

Let $E_1, E_2, \dots, E_n$ be a collection of mutually exclusive collectively exhaustive events, for some event $B \Rightarrow P(B) = \sum_{i=1}^n P(B|E_i)P(E_i)$



$$P(B) = P(B|E_1) + P(B|E_2) + \dots$$



$$P(B) = P(B|E_1)P(E_1) + P(B|E_2)P(E_2)$$

example) at a gas station 40% of the customers use regular gas (A₁) 35% use plus gas, 25% are premium. of those customers using regular gas only 30% fill their tanks, at those customers using plus gas 60% fill their tanks. and 50% of those using premium gas fill their tanks.

a) Prob that the next customer will fill their tanks.

A₁ ⇒ regular gas A₂ ⇒ plus A₃ ⇒ premium

B ⇒ filling their tanks.

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)$$

$$= 0.3 \times 0.4 + 0.6 \times 0.35 + 0.5 \times 0.25$$

$$= 0.455$$

prob / no text	2004 election	<u>Bush</u>	<u>Kerry</u>
	62% no college degree	50%	50%
	38% college degree	53%	47%

- what is the prob that a randomly selected respondent voted for Bush.

A: denoted the event that a respondent is a college degree.

B: denoted the event that an individual voted for bush.

$$P(B) = P(B|A)P(A) + P(B|A')P(A')$$

$$= \frac{38}{100} \times \frac{53}{100} + 0.5 \times 0.62 = 0.5114$$

Independence when occurrence of an event does not affect the probability of another event, then they are said to be independent.

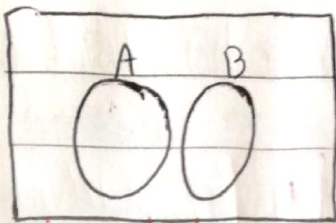
$$P(B|A) = P(B)$$

$$P(A \cap B) = P(B|A)P(A) = P(B)P(A)$$

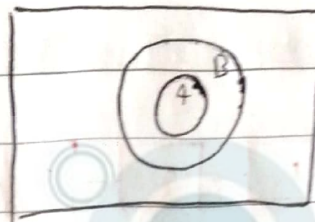
$$P(A|B) = P(A \cap B)P(B) = \frac{P(A)P(B)}{P(B)} = P(A)$$

- For a collection of independent events E₁, E₂, ..., E_n

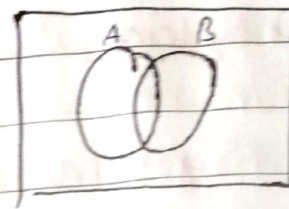
$$P\left(\bigcap_{i=1}^n E_i\right) = \prod_{i=1}^n P(E_i) = P(E_1)P(E_2)P(E_3), \dots$$



~~dependent~~
independent because
the prob didn't change.
 $P(A \cap B) = P(\emptyset) = 0$
 $P(A)P(B) > 0$



$A \subset B$
 $P(A \cap B) = P(A)$
 $\neq P(A)P(B)$
Dependent



They may or
may not
be independent

2 Cases for independence

1) As an assumption

Coin Tossing $P(H) = P(T) = 1/2$

2) Independence as a property (depends on data)

Roll a fair die $S = \{1, 2, 3, 4, 5, 6\}$

$A = \{2, 4, 6\} = P(A) = 3/6 = 1/2$

$B = \{1, 2, 3, 4\} = P(B) = 2/3$

$P(A \cap B) = 2/6 = 1/3 \Rightarrow P(A) \cdot P(B) = 1/3 = P(A \cap B) \Rightarrow$ independent

problem
2-126

100 disk

	high	low
scratch high	70	9
low	16	5

A: event that a disk has a high ~~scratch~~ shock Res

B: event that a disk has a high scratch Res

is ~~and~~ A and B dependent

$$P(A \cap B) = 70/100$$

$$P(A) = 86/100$$

$$P(B) = 79/100$$

$$P(A \cap B) \neq P(A) \cdot P(B) \text{ not independent}$$

Application of Probability Concepts (Reliability)

Reliability: the probability that a system or product will perform its intended function properly.

2 Types of systems

1) Series system —

each component works/fails independently than others

the system will work iff there is a path $\Rightarrow$ all of the components are working

A_i : event that component i works

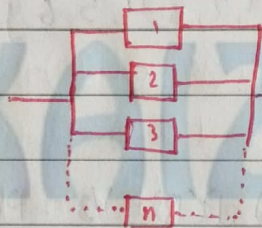
P_i : prob that component i works

$$P(\text{system works}) = P(A_1 \text{ works and } A_2 \dots A_n \text{ works})$$

$$= P(A_1 \cap A_2 \dots \cap A_n)$$

$$= P(A_1) \cdot P(A_2) \cdot P(A_3) \dots P(A_n)$$

2) Parallel system



$$P(\text{system works}) = P(A_1 \text{ or } A_2 \text{ or } \dots A_n)$$

$$= P(A_1 \cup A_2 \dots \cup A_n)$$

$$= 1 - P(\text{system doesn't work})$$

$$= P(A_1' \cap A_2' \cap \dots A_n')$$

$$= P(A_1') P(A_2') \dots P(A_n') = (1 - P_1)(1 - P_2) \dots (1 - P_n)$$

$$\therefore P(\text{system works}) = 1 - [(1 - P_1)(1 - P_2) \dots (1 - P_n)]$$

Series

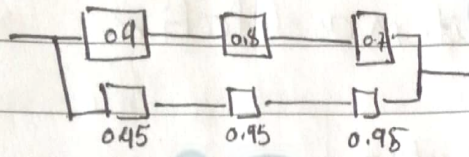
$$P(\text{system works}) = \prod_{i=1}^n P_i$$

Parallel

$$P(\text{system works}) = 1 - \prod_{i=1}^n (1 - P_i)$$

P-2-136 | Circuit operates iff there is a functional path from left to right. Assume that devices fail/function independently, what is the prob that the circuit operates

Sunday
13-Oct-19



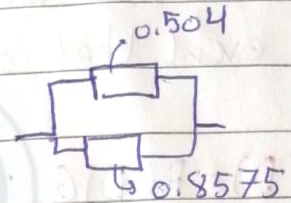
fail/function independently, what is the prob that the circuit operates

A: upper side works

B: lower side works

$$P(A) = 0.9 \times 0.8 \times 0.7 \Rightarrow \text{series} = 0.504$$

$$P(B) = (0.95)^3 \Rightarrow \text{series} = 0.8574$$



$$P(\text{circuit operates}) = 1 - \prod_{i=1}^2 (1 - P_i) \\ = 1 - [(1 - 0.504)(1 - 0.8574)] = 0.9293$$

Bayes's Theorem

P(event given a condition)

what if we wanted to know if the condition was present when an outcome (event) has occurred

- using conditional probs and law of total prob

$$P(A \cap B) = P(A|B) P(B) = P(B|A) P(A)$$

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

if $E_1, E_2, E_3, \dots, E_n$ are mutually exclusive and exhaustive events $\Rightarrow$

$$P(E_j|B) = \frac{P(B|E_j) P(E_j)}{\sum_{i=1}^n P(B|E_i) P(E_i)}$$

P2-142 | $P(A|B) = 0.7$ $P(A) = 0.5$ $P(B) = 0.2$ find $P(B|A)$

$$P(B|A) = \frac{P(A|B) P(B)}{P(A)} = 0.28$$

$$P(A) = P(A|B) P(B) + P(A|B') P(B')$$

2-148] Inspector at a production line, takes samples
 99% correctly identifying a defective item
 0.5% incorrectly classifying a good item as defective.
 0.9% of items are defective

A) what is the prob that the item is defective

D: event that an item is defective

G: event that an item is correctly identified as good

$$P(D) = P(D|G)P(G) + P(D|G')P(G') \\ = 0.005 \times (1 - 0.009) + 0.99(0.009) = 0.013865$$

B) If the item is classified as non-defective what is the prob that it is indeed good

$$P(D|G') = \frac{P(G \cap D')}{P(D')} = \frac{P(D'|G)P(G)}{P(D')} \\ = \frac{(1 - 0.005)(1 - 0.009)}{1 - 0.013865} = 0.999$$

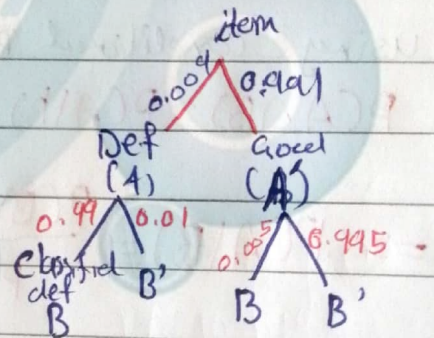
Method 2] A: item is defective

B: item is classified defective

$$P(A) = 0.009$$

$$P(B|A) = 0.99$$

$$P(B|A') = 0.005$$



A) $P(\text{classified as defective}) P(B) = P(B|A)P(A) + P(B|A')P(A')$
 $= 0.99 \times 0.009 + 0.005 \times 0.991 = 0.013865$

B) $P(A'|B') = \frac{P(A' \cap B')}{P(B')} \xrightarrow{\text{common sense}} \frac{1 - P(A \cup B)}{1 - [P(A)P(B) - P(A \cap B)]}$
 $\xrightarrow{1 - P(B)} \frac{1 - [P(A)P(B) - P(A \cap B)]}{1 - P(B)} = \frac{1 - [0.009 \times 0.99 - 0.009 \times 0.99]}{1 - 0.013865} = 0.999$

Final answer = 0.999

→ random variable

Chapter 3 Discrete R.V. & Prob distribution Tuesday 15-oct-19

Random Variables: a function defined on the sample space or a function that assigns a real number to each outcome in the sample space. is used to summarize the outcome from a random experiment by a simple number.

Random variables $\left\{ \begin{array}{l} \rightarrow \text{discrete (finite or countably infinite)} \\ \rightarrow \text{continuous (finite or interval)} \end{array} \right.$

Ex Throw 2 4-face dice

X : Sum of values of the two dice

$S = \{(1,1), (1,2), \dots, (4,4)\}$

my y is the probability at x

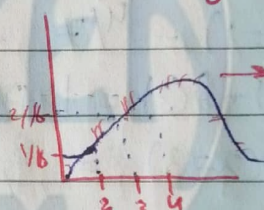
1st \ 2nd	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

$X = \{2, 3, 4, 5, 6, 7, 8\}$

$$p(2) = 1/16$$

$$p(3) = 2/16$$

$$p(4) = 3/16$$



do not connect the dots because we don't have any values for x in between

Problem 3-4 Batch of 500 machined parts, Contains 10 - non-conforming to requirements

R.V. = number of parts in a sample space of 5 parts that do not conform to customer requirements

$X: \{0, 1, 2, 3, 4, 5\}$

$p(x=0)$ = number of samples that contain no defection

Total number of samples from 500

$$= \frac{{}^{490}C_5}{{}^{500}C_5}$$

${}^{490}C_5 \rightarrow$ places 5 non-defective (ND) parts in a sample of 5

$$p(x=1) = \frac{10 \times {}^{490}C_4}{{}^{500}C_5}$$

$10 \rightarrow$ 10 defective (D) parts, ${}^{490}C_4 \rightarrow$ places 4 non-defective (ND) parts in a sample of 5

Mapping (associating) probabilities at different values at the R.V. $\Rightarrow$ Distribution

definition of distribution

For a discrete random variable X with possible values $x_1, x_2, \dots, x_n$ the prob mass function is $f(x_i)$

1) $f(x_i) \geq 0$

2) $\sum f(x_i) = 1$

3) $f(x_i) = P(X=x_i)$

Prob 3-16 $f(x) = \frac{8}{7} \times (\frac{1}{2})^x$ $x=1, 2, 3$ is this prob mass?

sub to see if $\sum = 1$ ~~$\frac{8}{7} \times \frac{1}{2} + \frac{8}{7} \times \frac{1}{4} + \frac{8}{7} \times \frac{1}{8}$~~ =

$\frac{8}{7} \times (\frac{1}{2}) + \frac{8}{7} \times \frac{1}{4} + \frac{8}{7} \times \frac{1}{8} = \frac{7}{7} = 1$ then it's prob mass function
+ no prob is = 0

a) $P(X \leq 1) = P(X=1) = \frac{4}{7} = f(1)$

b) $P(X > 1) = f(2) + f(3) = 1 - P(X \leq 1) = \frac{3}{7}$

c) $P(2 < X < 6) = P(X=3) = \frac{8}{7} \times \frac{1}{8} = \frac{1}{7} = f(3)$

d) $P(X \leq 1 \text{ or } X > 1) = P(X=1) + P(X=2) + P(X=3)$

Prob 3-26 Soil analysis instrument

successful, moderately successful, unsuccessful

Thursday $\heartsuit$
17-10-1990

0.3	0.65	0.05
Revenue \$10M	\$5 million	\$1 million

Prob distribution for the yearly Revenue

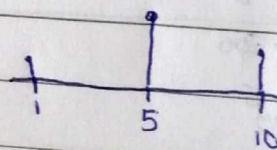
X : annual Revenue

Prob mass function

$P(X=10) = 0.3$

$P(X=5) = 0.65$

$P(X=1) = 0.05$



Cumulative distributive functions

finding prob of having a value as the $RV \leq$ to a specific value $P(X \leq x) \Rightarrow$ CDF

$$P(X \leq x) = F(x) = \sum_{x_i \leq x} f(x_i) \quad \text{given that } x_1, x_2, x_3, \dots, x_n \text{ are mutually exclusive}$$

Cumulative Prob also can be used to find Probability function Pmf

$$P(X = x_i) = P(X \leq x_i) - P(X \leq x_{i-1})$$

Problem 3-35

x	-2	-1	0	1	2
$f(x)$	$1/8$	$2/8$	$2/8$	$2/8$	$1/8$

it's Pmf $\sum = 1$ and all are less than 1 greater than 0

Determine the CDF $F(-2) = P(X \leq -2) = f(-2) = 1/8$

$$F(-1) = 3/8, F(0) = 5/8, F(1) = 7/8, F(2) = 1$$

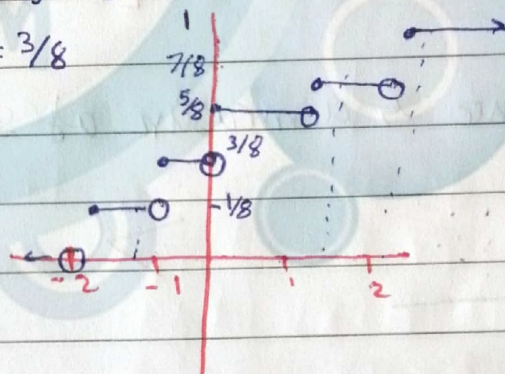
a) $P(X \leq 1.25) = 7/8$

b) $P(X \leq 2.2) = 1$

c) $P(-1.1 < X \leq 1) = P(X \leq 1) - P(X \leq -1.1) = 7/8 - 1/8 = 6/8$

d) $P(X > 0) = 1 - P(X \leq 0) = 1 - 5/8 = 3/8$

$$F(x) = \begin{cases} 0 & x < -2 \\ 1/8 & -2 \leq x < -1 \\ 3/8 & -1 \leq x < 0 \\ 5/8 & 0 \leq x < 1 \\ 7/8 & 1 \leq x < 2 \\ 1 & 2 \leq x \end{cases}$$



Hw. solve ^{problem} example 42) thickness of wood

Mean and variance of discrete R.V.

mean and variance are two measures used to summarize a prob disk

mean: expected value of a discrete R.V.

$\mu = E(X) = \sum x f(x)$ weighted average of x with weights equal probabilities it measures the central tendency of the disk and it is called first moment of the disk.

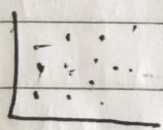
$$\sum x f(x) = \frac{2+3+5+4+6}{5} \\ = 2 \times \frac{1}{5} + 3 \times \frac{1}{5} + 5 \times \frac{1}{5} + \dots$$

Sunday
20-Oct-19

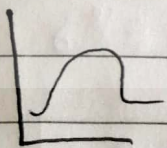
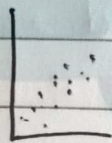
Variance of X : denoted σ^2 $\overset{\text{variance}}{=} \overset{\text{mean}}{\mu}$ $\overset{\text{variance}}{=} \overset{\text{mean}}{E(X)}$

$$\begin{aligned} \sigma^2 &= E[(x - \mu)^2] = V(X) \\ &= \sum (x - \mu)^2 f(x) \\ &= [\sum x^2 f(x)] - [\mu^2] \\ &= E[x^2] - [E(X)]^2 \end{aligned}$$

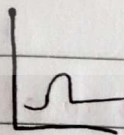
variance is a measure of scatterness of the possible variables.



>



>



equally likely.

Ex	outcome	a	b	c	d	e	f
(dimens)	x	0	0	1.5	1.5	2	3
	$f(x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

→ outcome
→ variable
→ probability.

what is the mean?

$$\mu = E(x) = \sum x f(x)$$

$$= 0 \times \frac{2}{6} + 1.5 \times \frac{2}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6}$$

$$= 1.333$$

what is the variance?

$$V(x) = \sum x^2 f(x) - \mu^2$$

$$= 0^2 + \frac{2}{6} + 1.5^2 \times \frac{2}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} - (1.33)^2$$

$$= 1.139 \text{ units}^2 = 30^2 \rightarrow \text{take root to remove } x^2$$

$$\sqrt{V(x)} = \sqrt{6^2} = 6 = \text{std dev.}$$

Tuesday
22-oct-2019

Expected value of a function

for a r.v. X with pmf $f(x)$ the expected value of a function $h(x)$

$$\sum [h(x)] = \sum h(x) f(x) \Rightarrow \text{variance} = E[(X - \mu)^2]$$

Ex) two product designs are to be evaluated on basis of profit

Design A : revenue \$3 million

Design B : revenue \$7 million with prob = 0.3

or revenue \$2 million with prob = 0.7

prob at $(X = \$3 \text{ million}) = 1$

Design A $\Rightarrow E[\text{profit}] = E[h(x)]^2 =$

$$E[0.4x] = 0.4 \times 3 + 1 = 1.2 \text{ M\$}$$

x	7	2
$f(x)$	0.3	0.7

Design B $\Rightarrow E(h(x)) = 0.4 \times 7 \times 0.3 +$

$$0.4 \times 2 \times 0.7 = 1.4 \text{ \$ million}$$

profit = 0.4 from revenue

$$h(x) = 0.4x$$

$$\sigma_A^2 = E(h(x) - \mu)^2 f(x) = (0.4 \times 3 - 1.2)^2 \times 1$$

$$= 0^2 \times 1 = 0$$

$$\sigma_B^2 = (0.4 \times 7 - 1.4)^2 \times 0.3 + (0.4 \times 2 - 1.4)^2 \times 0.7$$

$$= 0.84$$

$$V_B = \sqrt{0.84} = 0.916$$

Rules for $E[X]$ and $V[X]$

$$\begin{aligned} y &= ax + b & a, b \text{ constants} \\ \mu_y = E[Y] &= E[ax + b] = \sum (ax + b) f(x) \\ &= \sum ax f(x) + \sum b f(x) \\ &= a \sum x f(x) + b \sum f(x) \\ &= a E[X] + b(1) \end{aligned}$$

$$\begin{aligned} V[Y] &= E[(Y - \mu_Y)^2] \\ &= E[(ax + b) - (a\mu_X + b)]^2 \\ &= E[(ax - a\mu_X)^2] = E[a^2(x - \mu_X)^2] = a^2 E[(x - \mu_X)^2] \\ &= a^2 V[X] \end{aligned}$$

$$\begin{aligned} Y &= \sum a_i X_i \\ E[Y] &= \sum a_i E[X_i] \\ V[Y] &= \sum a_i^2 V(X_i) \end{aligned}$$

Common discrete distribution and the situation they model

1) Discrete uniform distribution $\rightarrow$ equal prob for each outcome.

$$f(x_2) = \frac{1}{n} \Rightarrow \text{is range } a, a+1, \dots, b$$

range of values contains $b - a + 1$

$$\Rightarrow \text{prob} = \frac{1}{b-a+1}$$

$$\Rightarrow \mu = E[X] = \frac{b+a}{2}$$

$$\sigma^2 = \frac{(b-a+1)^2 - 1}{12}$$

Solve 3-14 example

2) Bernoulli trial $\rightarrow$ dist

R.V. $\Leftarrow$ Bernoulli $\Rightarrow$ 3 conditions

1) two outcomes for each trial "once"

$\hookrightarrow Y, N, H, T, T, F$

2) Prob of success = P is not changed "constant"

3) Trials are independent

let Y be a R.V that represent Bernoulli experiment with values at $1 \Rightarrow$ success $0 \Rightarrow$ fail

$$Y = \{0, 1\}$$

$$f(1) = P(Y=1) = P$$

$$f(0) = P(Y=0) = 1-P$$

$$E[Y] = \sum Y f(Y) = 1 \times P + 0(1-P) = P$$

$$V[Y] = \sum Y^2 f(Y) - \mu^2 = 1 \times P + 0 - P^2 = P - P^2 = P(1-P)$$

$Y \overset{\text{distributed as}}{\sim} \text{Bernoulli}(P)$

3) Binomial distribution $\rightarrow$ Series of Bernoulli trials.

R.V. represents a count at the number of trials that meet specific criteria "Success"

ex: flipping a coin 10 times. Count H

Y_i : outcome from i^{th} trial n : trials

ex: $n = 5$ trials $1, 0, 1, 1, 0$

$$X = \sum_{i=1}^n Y_i = 1 + 0 + 1 + 1 + 0 = 3$$

X : sum of success.

X : # of successes in n independent Bernoulli trials having the same prob of success " p "

Possible values of X : $0, 1, 2, 3, \dots, n$

for k successes

$$\underbrace{1, \dots, 1}_k$$

$$\underbrace{0, 0, \dots, 0}_{n-k}$$

$$p \dots p$$

$$(1-p) \dots (1-p)$$

this sequence has the prob of $p \times p \times \dots p \times (1-p) \times \dots (1-p)$
 $p^k \times (1-p)^{n-k}$

We have C_k^n at k and $n-k$

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

if $p=0$ the $P(X=0) = 1$

↳ if there is no success then it will always fail then the prob at fail is $1 = \binom{n}{0} 0^0 \cdot 1 = \frac{n!}{0!n!} = 1 \cdot 1 = 1$

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \Rightarrow x = 0, 1, \dots, n$$

$$E[X] = np$$

$\Rightarrow$ average of successes "expected number of successes"

$$V[X] = np(1-p)$$

$$X \sim \text{Bin}(n, p)$$

end of quiz material

Sunday

27-Oct-2019

3-75 | a, f Binomial

3-77 | $X \sim \text{Bin}(10, 0.1)$

$$a) \text{Prob}(X \leq 2) = \sum_{x=0}^2 \binom{10}{x} 0.1^x (1-0.1)^{n-x}$$

$$= \binom{10}{0} 0.1^0 (0.9)^{10} + \binom{10}{1} 0.1^1 (0.9)^9 + \binom{10}{2} 0.1^2 (0.9)^8 = 0.9298$$

$$b) P(X > 8) = 1 - P(X \leq 8) = P(X=9) + P(X=10) =$$

$\Leftarrow$ (very low prob)

$$c) P(X=4) = \binom{10}{4} 0.1^4 \times 0.9^6 = 0.0112$$

$$d) P(5 \leq X \leq 7) = P(X=5) + P(X=6) + P(X=7) = 0.0016$$

4) Geometric distribution

Let X denote a r.v. that represents the number of trials until first success in a sequence of Bernoulli trials

$$X = 1, 2, 3, \dots$$

$$X \sim \text{Geo}(p) \quad P(X=x) = f(x) = P(1-p)^{x-1}$$

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

$$E(X) = 1/p$$

$$\sigma^2 = V[X] = \frac{1-p}{p^2}$$

3-99 | $X \sim \text{geo}(0.5)$ a) $P(X=1) = 0.5(1-0.5)^0 = 0.5$

* when first trial success $\Rightarrow$ Bern

b) $P(X=4) = 0.5(1-0.5)^3 = 0.0625$

c) $P(X=8) = 0.5(1-0.5)^7 = 0.0039$

d) $P(X \leq 2) = P(X=1) + P(X=2) = 0.75$

e) $P(X > 2) = 1 - P(X \leq 2) = 0.25$

3-101 | geo , $p=0.2$

a) what is the expected number of trials to obtain 1st success?

$$X \sim \text{geo}(0.2) \Rightarrow E[X] = 1/p = 1/0.2 = 5 \text{ trials}$$

b) After the 8th success occurs, what is the expected number of trials to obtain the ninth success?

$\rightarrow$ because it is independent we don't look at the eighth so the expected value is still 5 $\Rightarrow 1/p$ "Because lack of memory" property

5) Negative Binomial Dist

"Generalization of geometric"

X : number of trials until the r^{th} success

$$P(X=x) = f(x) = \binom{x-1}{r-1} (1-p)^{x-r} p^r$$

$$X: r, r+1, r+2, \dots, \infty$$

if $r=1 \Rightarrow$ geometric

$$\mu = E[X] = r/p$$

$$\sigma^2 = V[X] = r(1-p)/p^2$$

$$X \sim \text{Neg Bin}(r, p)$$

$$P(X < r) = 0$$

Figure 3-10

const r

$\uparrow p \Rightarrow x \downarrow$

const p

$\uparrow r \Rightarrow x \uparrow$

Binomial r.v: Count number of successes in n Bernoulli Trials $n \Rightarrow$ ^{r.v} predetermined

Tuesday
29-10-2019

Neg Binomial: count # of Bernoulli trials to obtain a n # of successes
_{r.v.} ^{predetermined}

#3-106 In a clinical study, volunteers are tested for a gene this gene $\Rightarrow$ disease Prob a person carried the gene is 0.1

a) what is the prob that 4 or more people will have to be tested before two with the gene be detected?

X : number of people who are tested (trials) 2 successes

$X \sim \text{Neg Bin}(n, p) = (2, 0.1)$ number of trials $>$ # of success

$$P(X \geq 4) = 1 - P(X < 4) = 1 - [P(X=2) + P(X=3)]$$
$$= 1 - \left[\binom{1}{1} (1-0.1)^0 (0.1)^1 + \binom{2}{1} (1-0.1)^1 (0.1)^2 \right] = 0.972$$

$\Rightarrow$ the prob that i would at least need 4 to get 2.

b) How many people are expected to be tested ~~for neg~~ before two with the gene are detected?

$$E[X] = \frac{n}{p} = \frac{2}{0.1} = 20 \Rightarrow \text{i will expect to test 20}$$

6) Hypergeometric Distribution $\Rightarrow$ sampling is done without replacement

N objects, K classified as successes

$N-K$ failure $\Rightarrow$ sample of size n randomly w/o replacement.

ex) 850 parts, 50 parts do not conform (defective)

2 parts are selected randomly "sample"

X : # of defective parts in the sample

$$P(X=0) \Rightarrow \text{prob both are good} = \frac{800}{850} \times \frac{799}{849} = 0.886$$

$$P(X=1) \Rightarrow \text{one is defective} = \left(\frac{800}{850} \times \frac{50}{849} \right) + \left(\frac{50}{850} \times \frac{800}{849} \right) = 0.111$$

A: 1st def B: 2nd def

$$P(X=2) \Rightarrow \text{both are defective} = P(A \cap B) = P(A) \cdot P(B|A) = \frac{50}{850} \times \frac{49}{849} = 0.003$$

$$f(x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}$$

X : # success in the sample

$X \sim \text{Hypergeo}(K, N, n)$

$$f(x=0) = \frac{\binom{50}{0} \binom{850-50}{2-0}}{\binom{850}{2}} = 0.886$$

$n+K > N \rightarrow n+K-N \neq 0$

$x = [0, n+K-N] \rightarrow \max$

$x = [K, n] \rightarrow \min$

$$p = \frac{K}{N}$$

$$f(x=1) = \frac{\binom{50}{1} \binom{800}{2-1}}{\binom{850}{2}} = 0.111$$

$$E[X] = nP$$

$$V[X] = nP(1-P) \frac{N-n}{N-1}$$

$N-n$ = correction factor without replacement.

if indirect $N \leftarrow V[X] = nP(1-P)$

Sunday

3-Nov-2019

without replacement at very big population

$\Rightarrow$ binomial

$$N=100, n=10, K=11$$

$$\max = \min\{10, 11\} = 10 \quad \min \Rightarrow \max\{0, 21-100\} = 0$$

P-3-122] 36 Bacteria cells, 12 not capable at replication
you examine 3 cells w/o replacement $\rightarrow$ Hypergeo

a) pmf that can replicate

$$X \sim \text{Hypergeo}\left(\overset{K}{24}, \overset{N}{36}, \overset{n}{3}\right), f(x) = \frac{\binom{24}{x} \binom{36-24}{3-x}}{\binom{36}{3}}$$

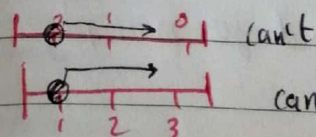
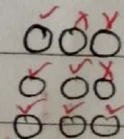
b) mean and variance that can replicate

$$E[X] = nP = n \times \frac{K}{N} = 3 \times \frac{24}{36} = 2$$

$$V[X] = nP(1-P) \frac{N-n}{N-1} = 2 \left(1 - \frac{24}{36}\right) \left(\frac{36-3}{36-1}\right) = 0.629$$

c) Prob at least one can't replacte

$$P(X \leq 2) = 1 - P(X=3) = 0.717$$



Tuesday
5-Nov-2019

7) Poisson dist

Model prob at arrivals within
an interval at time

number of customers showing up in a given time interval

number of phone calls arriving to a call center

number of flaws in a length unit

general: number at x in an interval at y

$\lambda \Rightarrow$ rate at arrivals $x \Rightarrow$ number at arrivals

$$\lambda = x/y$$

divide interval into n subintervals

$$n \rightarrow \infty$$

in each of these n subintervals an arrival could or could not
happen $\Rightarrow$ now it's Bernoulli

no. at arrivals in the large interval (n subintervals)
 $= nP$

λ : mean no. at arrivals

$$\lambda = nP \Rightarrow P = \frac{\lambda}{n}$$

$$\binom{n}{x} P^x (1-P)^{n-x} \Rightarrow \lim_{n \rightarrow \infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$$E[X] = V[X] = \lambda$$

average # of customer arrivals per hour = 2.6

per 10 hours = 26

$$P(X=x) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}$$

where t is what you multiply
by λ

ex: $\lambda = 2.3$ flaws/mm $P(X=10)$ in 5 mm

$$P(X=10) = \frac{e^{-2.3 \times 5} (2.3 \times 5)^{10}}{10!}$$

[3-135] # of stars in a given volume space is a poisson r.v with density = 1 star/16 cubic light years

a) Prob of 3 or more stars in 16 ly³

$$P(X \geq 3) = 1 - P(X < 3) \\ = 1 - [P(X=2) + P(X=1) + P(X=0)]$$

$$\frac{e^{-\lambda} \lambda^x}{x!} \Rightarrow \frac{e^{-1}}{2!} + \frac{e^{-1}}{1!} + \frac{e^{-1}}{0!}$$

b) $P(X \geq 1) \approx 0.95 \Rightarrow$ what is λ

$$= 1 - P(X=0) = 1 - \frac{e^{-\lambda}}{0!} \geq 0.95$$

$$1 - 0.95 \geq \frac{1}{e^{\lambda}}$$

$$e^{\lambda} \geq 20$$

$$\lambda \geq \ln 20$$

$$\lambda \approx 3$$

$$16 \times 3 = 48 \text{ cubic light years}$$

10-11-19

Sunleey

Continuous distributions

ch4

Continuous R.V. $\left\{ \begin{array}{l} \text{infinite } [0, \infty) (-\infty, 1] \\ \text{finite interval } [1, 2] \end{array} \right.$

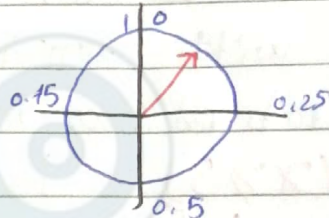
(interval of real number sample length, time, Area, ...)

Candyland spinner

-arrow can stop anywhere

X : the proportion of the total circumference it lands on

$X \in [0, 1]$



$$P(0.25 \leq 0.75) = \frac{\text{Area in range}}{\text{total Area}} = \frac{0.5}{1} = 0.5$$

$$P(X=0.5) = \frac{0}{\pi} = 0 \rightarrow \text{no area}$$

For any individual outcome, x , $P(X=x) = 0$

Prob. associated with a continuous random variable must be computed for interval of X

For a continuous R.V. X , prob density function is a function such that

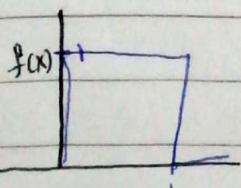
$$1) f(x) \geq 0$$

$$2) \int_{-\infty}^{\infty} f(x) dx = 1 \rightarrow \text{total Area}$$

$$3) P(a \leq x \leq b) = \int_a^b f(x) dx$$

For the spinner, Pdf

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \leq x \leq 1 \\ 0 & x > 1 \end{cases}$$



$$P(0.25 \leq x \leq 0.75) = \int_{0.25}^{0.75} f(x) dx = \int_{0.25}^{0.75} 1 \cdot dx = x \Big|_{0.25}^{0.75} = 0.75 - 0.25 = 0.5$$

$P(0.25 < x < 0.75)$ same value $\Rightarrow < = \leq \Rightarrow$ the equal sign does nothing

ex 4-1] $f(x) = e^{-x} \quad x > 0$

a) $p(1 < x) = \int_1^{\infty} e^{-x} dx = e^{-x} \Big|_1^{\infty} = -e^{-\infty} + e^{-1} = \frac{1}{e}$

b) $p(1 < x < 2.5) = \int_1^{2.5} e^{-x} dx = -e^{-x} \Big|_1^{2.5} = e^{-1} - e^{-2.5} = 0.2858$

c) $p(x=3) = 0$

d) $p(x < 4) = \int_0^4 e^{-x} dx = 1 - e^{-4} = 0.9817$

e) $p(3 \leq x) = \int_3^{\infty} e^{-x} dx = e^{-x} \Big|_3^{\infty} = e^{-3} = 0.0498$

f) $p(x < x) = 0.1 = \int_x^{\infty} e^{-x} dx = 0.1 = e^{-x} \Big|_x^{\infty} = e^{-x} = 0.1$
 $-x \ln e = \ln 0.1 = -2.3 = x$

g) $p(x \leq x) = 0.1 = \int_0^x e^{-x} dx = 0.1 \Rightarrow 1 - e^{-x} = 0.1 \Rightarrow -x \ln e = \ln 0.9$
 $x = 0.1054$

4-2] $f(x) = 3(8x - x^2)/256 \quad 0 < x < 8$

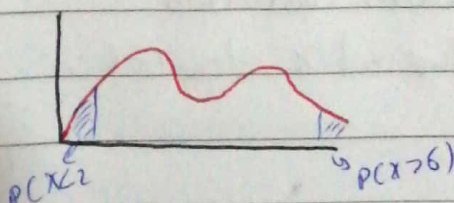
a) $p(x < 2) = 0.1563$

b) $p(x < 9) = 1$ All the area

c) $p(2 < x < 4) = 0.3438$

d) $p(x > 6) = 0.1563$

e) $p(x < x) = 0.95 \Rightarrow x^3 - 12x^2 + 243.2 = 0$



Cumulative distribution function

The CDF of a continuous r.v. X is

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

example: give the following PDF

$$f(x) = 0.25 \quad 1 \leq x \leq 5$$



define the CDF for the r.v. X

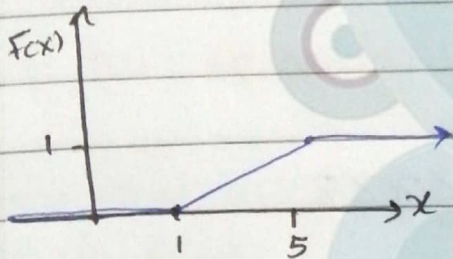
1) greater than zero $\Rightarrow \checkmark$

2) integration $\checkmark$ $\frac{1}{4}((5-1) \times 1) = 1$

$$3) F(x) = \int_{-\infty}^x f(u) du = \int_1^x 0.25 du = 0.25 u \Big|_1^x = 0.25(x-1)$$

$$F(x) = \begin{cases} 0 & , -\infty < x < 1 \\ 0.25(x-1) & , 1 \leq x < 5 \\ 1 & , 5 \leq x \end{cases}$$

$$\int_{-\infty}^1 0 du + \int_1^5 (0.25) du + \int_5^{\infty} 0 du = 1$$



When given $F(x)$ then $f(x) = \frac{dF(x)}{dx} \rightarrow$ differentiation (reverse it)

- The mean and variance of a continuous Random variable

Suppose that X is a continuous random variable with PDF $f(x)$ the mean or expected value of X denoted as μ or $E(X)$ is

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

The variance of x denoted as $V(x)$ or σ^2 is

$$\begin{aligned} V(x) &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2 \end{aligned}$$

The standard deviation of x is $\sigma = \sqrt{V(x)}$

continued at previous example

Tuesday
19-11-19

$$E(x) = \mu = \int_1^5 x f(x) dx = \int_1^5 0.25 x dx = \left. \frac{0.25 x^2}{2} \right|_1^5 = \frac{(25-1)(0.25)}{2} = \boxed{3}$$

$$\begin{aligned} V(x) = \sigma^2 &= \int_1^5 x^2 f(x) dx - \mu^2 \\ &= \int_1^5 0.25 x^2 dx - (3)^2 = \left. \frac{0.25 x^3}{3} \right|_1^5 - 3^2 \\ &= \frac{0.25(5^3 - 1^3)}{3} - 3^2 = 1.33 \end{aligned}$$

$$\sigma = \sqrt{V(x)} = \sqrt{1.33}$$

If x is a continuous R.V. with PDF $f(x)$ then

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

$$\begin{aligned} \text{ex: } f(x) &= 0.25 \quad 1 \leq x \leq 5 \quad \text{find } E(x^2 - 2) \\ &= \int_1^5 (x^2 - 2)(0.25) dx = \int_1^5 (0.25 x^2 - 0.5) dx = \left. \frac{0.25 x^3}{3} - 0.5 x \right|_1^5 \end{aligned}$$

Continuous uniform distribution: A continuous R.V x with PDF $f(x) = \frac{1}{b-a}$, $a \leq x \leq b$ is called a continuous uniform Random variable

$$\mu = \int_a^b x f(x) dx = \int_a^b x \left(\frac{1}{b-a} \right) dx = \frac{x^2}{2(b-a)} \Big|_a^b$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \boxed{\frac{a+b}{2}}$$

$$V(x) = \int_a^b x^2 f(x) dx - \mu^2$$

$$= \int_a^b x^2 \left(\frac{1}{b-a} \right) dx - \mu^2$$

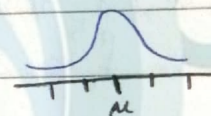
$$= \frac{x^3}{3(b-a)} \Big|_a^b - \left(\frac{a+b}{2} \right)^2$$

$$= \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4}$$

$$= \boxed{\frac{(b-a)^3}{12}}$$

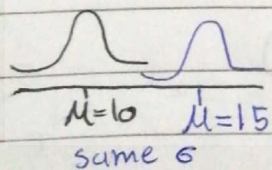
2) Normal Distribution

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

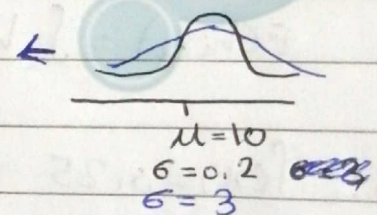


- to find the P you need to integrate x but that's hard

the bigger the σ
the wider the graph.



→ the difference in μ makes
me shift the curve

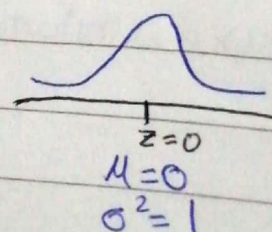


Standard Normal Random Variable

⇒ Normal R.V $\mu=0, \sigma^2=1$

- there is only one standard curve

use appendix to get values.

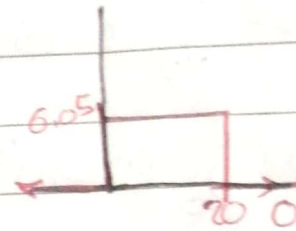


24-11-19

Sunday

ex) X : current in a thin wire range of $X = [0, 20]$ mA

$$pdf = f(x) = \begin{cases} 0 & x < 0 \\ 0.05 & 0 \leq x \leq 20 \\ 0 & x > 20 \end{cases}$$

CDF for $x < 0$

$$f(x) = P(X \leq x) = \int_{-\infty}^x f(u) du = \int_{-\infty}^x 0 du = 0$$

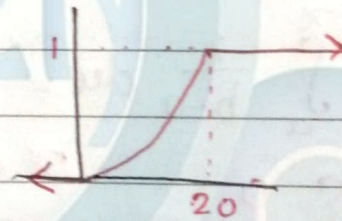
CDF for $0 \leq x \leq 20$

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du = \int_{-\infty}^0 0 du + \int_0^x 0.05 du = 0.05x$$

CDF for $x > 20$

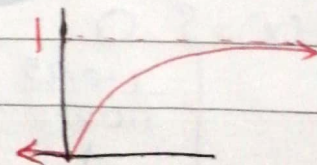
$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du = \int_{-\infty}^0 0 du + \int_0^{20} 0.05 du + \int_{20}^x 0 du = 0 + 0.05 \Big|_0^{20} + 0 = 1$$

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.05x & 0 \leq x \leq 20 \\ 1 & x > 20 \end{cases}$$

Determine CDF for $f(x) = e^{-x}$ $x > 0$ for $x < 0 \Rightarrow F(x) = \int_{-\infty}^x 0 du = 0$ not identified

$$\text{for } x > 0 \Rightarrow F(x) = \int_0^x e^{-u} du = -e^{-u} \Big|_0^x = -e^{-x} + e^0 = 1 - e^{-x}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x > 0 \end{cases}$$



$$CDF \quad f(x) = \begin{cases} 0 & x < -2 \\ 0.25x + 0.5 & -2 \leq x < 1 \\ 0.5x + 0.25 & 1 \leq x < 1.5 \\ 1 & 1.5 \leq x \end{cases} \Rightarrow \text{differentiable}$$

$$f'(x) = \begin{cases} 0.25 & -2 \leq x < 1 \\ 0.5 & 1 \leq x < 1.5 \end{cases}$$

$$f(x) = 0.25 \quad 0 < x < 4 \quad \mu \text{ and } \sigma^2 \text{ for } X$$

$$\mu = E(x) = \int_0^4 0.25x dx = \frac{0.25x^2}{2} \Big|_0^4 = 2$$

$$\sigma^2 = v(x) = \int_0^4 (x-2)^2 0.25 dx = 0.25 \frac{(x-2)^3}{3} \Big|_0^4 = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$$

For a function $h(x)$ $E[h(x)] = \int_{-\infty}^{\infty} h(x)f(x) dx$

1) Continuous uniform dist

X : Continuous uniform r.v. $f(x) = \frac{1}{b-a}$

$X \sim \text{uniform}(a, b)$

$$\mu = E[X] = \frac{a+b}{2}$$

$$\sigma^2 = v[X] = \frac{(b-a)^2}{12}$$

$$F(x) = \int_a^x \frac{1}{b-a} du = \frac{x-a}{b-a}$$

$$f(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & 0 \leq x \leq b \\ 1 & x > b \end{cases}$$

4-41

example) thickness at a flange is unif dist $[0.95, 1.05]$

$$a) \text{ CDF} = F(x) = \begin{cases} 0 & x < 0.95 \\ \frac{x-0.95}{1.05-0.95} & 0.95 \leq x \leq 1.05 \\ 1 & x > 1.05 \end{cases}$$

$$b) P(X > 1.02) = 1 - P(X \leq 1.02) = 1 - (10 \times 1.02 - 9.5) = 0.3 = F(1.02)$$

c) Thickness exceeded by 90% at flange

$$P(X > x) = 0.9 = 1 - F(x)$$

$$F(x) = 0.1$$

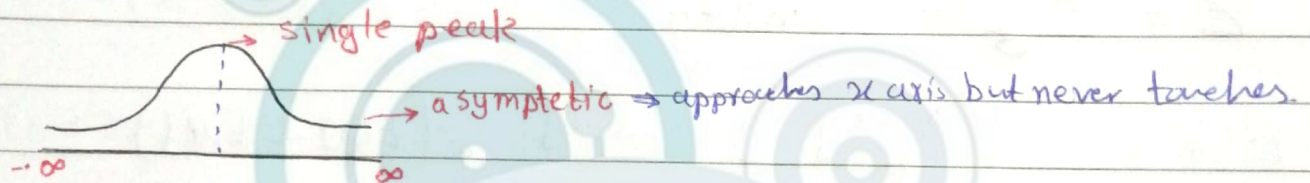
$$10x - 9.5 = 0.1 \Rightarrow x = 0.96$$

d) $E(x) = \frac{1.05 + 0.95}{2} = 1$ units

$V(x) = \frac{(1.05 - 0.95)^2}{12} = 0.0083 \text{ units}^2$

2) Normal distribution $\leadsto$ Gauss dist

Tuesday
26-11-19

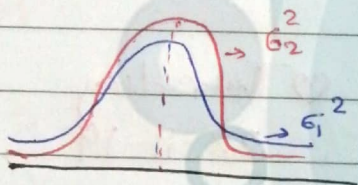


single peak: mean, medium, and mode

* always symmetric

* never touches the x-axis (asymptotic)

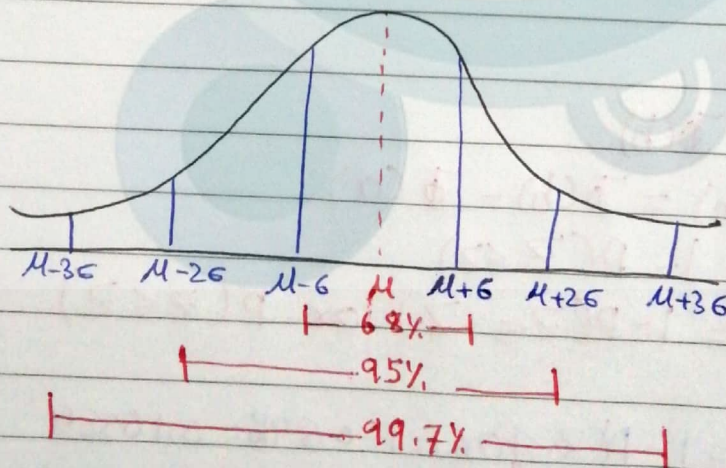
pdf $\Rightarrow f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$



$\sigma_1^2 > \sigma_2^2$
the bigger the σ the flatter it becomes.

$E[x] = \mu \Rightarrow$ determines center "location" of the distribution

$V[x] = \sigma^2 \Rightarrow$ determines the width or dispersion



$P(\mu - \sigma < x < \mu + \sigma) = 68\%$

$P(\mu - 2\sigma < x < \mu + 2\sigma) = 95\%$

$P(\mu - 3\sigma < x < \mu + 3\sigma) = 99.7\%$

1st test $\Rightarrow \mu = 63 \quad \sigma = 18$

$$P(63 - 18 < x < 63 + 18) = (45 < 81) = 68\%$$

Standardization

$$Z = \frac{X - \mu}{\sigma} \quad \text{"unitless"}$$

$$E[Z] = E\left[\frac{X - \mu}{\sigma}\right] = \frac{1}{\sigma} E[X - \mu] = \frac{1}{\sigma} (E[X] - E[\mu]) \\ = \frac{1}{\sigma} (\mu - \mu) = 0$$

$$V[Z] = V\left[\frac{X - \mu}{\sigma}\right] = \frac{1}{\sigma^2} V[X - \mu] = \frac{1}{\sigma^2} (V[X]) = \frac{\sigma^2}{\sigma^2} = 1$$

Z: standard normal r.v.

$$\text{CDF} = \Phi(z) = P(Z \leq z) = \int$$

Ex: Monthly salaries are normally distributed

♥ Thursday
28-11-19

$$\mu = \$2000 \quad \sigma = \$200$$

$$x_1 = 2200 \$ \Rightarrow \frac{2200 - 2000}{200} = 1 = z_1, \Phi(1) = P(Z \leq 1)$$

$$x_2 = 1700 \$ \Rightarrow \frac{1700 - 2000}{200} = -1.5 = z_2, \Phi(-1.5) = P(Z \leq -1.5)$$

General Rules:

$$1) P(Z \leq b) = \Phi(b)$$

$$2) P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$$

$$3) P(Z \geq z) = 1 - P(Z \leq z)$$

$$4) P(Z \geq -z) = 1 - P(Z \leq -z) \text{ or } P(Z \leq z)$$

$$1) P(Z > 1.26) = 1 - P(Z \leq 1.26) = 1 - 0.8944 = 0.1056$$

$$2) P(Z < -0.86) = 0.1949$$

$$3) P(Z > -1.37) = 1 - P(Z \leq -1.37) = P(Z \leq 1.37) = 0.91465$$

$$4) P(-1.25 < Z < 0.37) = \Phi(0.37) - \Phi(-1.25)$$

5) $P(Z \leq -4.6) = 0$ because it's smaller than the least value

6) $P(Z > z) = 0.05 = 1 - \Phi(z) = 0.05 \Rightarrow \Phi(z) = 0.95 \quad z = 1.65$

7) $P(-z \leq Z \leq z) = 0.99$ the area left is only 0.1 so 0.005 on each side so $z = \Phi^{-1}(0.005) \Rightarrow -z = -2.57$

8) $P(-1 < Z < 1) = P(Z < 1) - P(Z > 1)$
 $= P(Z < 1) - [1 - P(Z < 1)]$

9) $P(0 < Z < 1) = P(Z < 1) - P(Z < 0) = 0.5$

Percentiles of a continuous dist

Let $0 \leq p \leq 1$ The $(100p)$ percentile of the dist of a continuous r.v. X is x_p and is defined by.

$$P = P(X \leq x_p) = F(x_p) = \int_{-\infty}^{x_p} f(u) du$$

x_p is the value on the measurement axis such that the area under the graph of $f(x)$ lies to the left

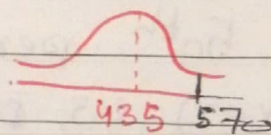
Ex: students receive 570 on the verbal SAT
SAT scores in general are ND (435, 10,000)

- Find student's percentile = $P(X \leq 570)$

$$= P\left(\frac{X - \mu}{\sigma} \leq \frac{570 - \mu}{\sigma}\right) = P\left(Z \leq \frac{570 - 435}{\sqrt{10,000}}\right)$$

$$= P(Z \leq 1.35) = \Phi(1.35) = 0.9115$$

Sunday
1st December



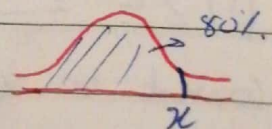
- Prob of student scoring higher than 435 = 0.5

- If a school accepts students who are above the 80th percentile?

$$P(X \leq x) = 0.8 \Rightarrow P(Z \leq z) = 0.8 = \Phi(0.8)$$

$$\text{at } z = 0.8416 \quad \Phi = 0.8 \Rightarrow z = \frac{x - \mu}{\sigma} \Rightarrow x = \mu + z\sigma$$

$$= 435 + 0.8416 \times 100 = 519$$



ex) Diameter at a shaft $\sim N(0.2508, 0.0005^2)$

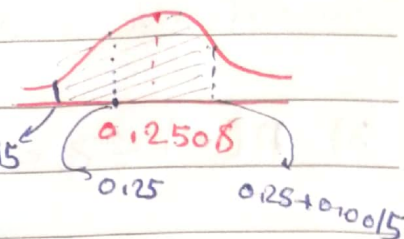
specifications 0.2500 ± 0.0015 inches

proportion at shafts conform to specifications.

$$\sigma = 0.0005$$

$$P(0.2485 \leq x \leq 0.2515)$$

$$P\left(\frac{0.2485 - 0.2508}{0.0005} \leq z \leq \frac{0.2515 - 0.2508}{0.0005}\right) = P(-0.46 \leq z \leq 0.14)$$



$$P(-4.6 \leq z \leq 1.4)$$

$$\Phi(1.4) - \Phi(-4.6)$$

$$= 0.91924 - 0$$

because it does not exist on the graph.

$$= 1 - 0.91924$$

Ex prob 4-56 Time until recharge for a battery is normally distributed with $\mu = 260$ min $\sigma = 50$ min

a) prob that battery last more than 4 hours.

X: time until recharge $\Rightarrow X \sim N(260, 50^2)$

$$P(X > 4 \times 60) \Rightarrow P(Z > \frac{240 - 260}{50}) = P(Z > -0.4)$$

$$= 1 - P(Z < -0.4) = 0.655422$$

b) 25th percentile: lower quartile Q_1

$$P(X \leq x) = 0.25, \Phi(z) = 0.25$$

$$\Phi^{-1}(0.25) = -0.675 \Rightarrow \frac{x - 260}{50} = -0.675$$

$$x = 226.25 \text{ min}$$

50th percentile: middle quartile Q_2 (median)

$$P(X \leq x) = 0.5, \Phi(z) = 0.5, \Phi(0.5) = 260 \text{ min}$$

75th percentile: upper quartile Q_3

$$P(X \leq x) = 0.75, P(Z \leq z) = 0.75, \Phi^{-1}(0.75) = 0.675$$

$$= 293.75 \text{ min}$$

$$c) P(X > x) = 0.95 \Rightarrow P(X \leq x) = 0.05$$

$$P(Z \leq \frac{x - 260}{50}) = 0.05, \Phi^{-1}(0.05) = -1.645$$

$$\text{convert to } x \Rightarrow 177.75 \text{ min}$$

*Normal Approximations

Tuesday
3rd of Dec.

1) If $X \sim \text{Binomial}(n, p)$ is large enough $np > 5, n(1-p) > 5$
Rule of thumb

$\Rightarrow$ it is feasible to approximate the Binomial by a normal distribution.

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0, 1)$$

$$P(X \leq x) = \sum_x \binom{n}{x} p^x (1-p)^{n-x} \approx P\left(Z \leq \frac{x - np}{\sqrt{np(1-p)}}\right)$$

Bin is a discrete r.v. so we use continuity correction

$$P(a \leq X \leq b) = P\left(a - \frac{1}{2} \leq X \leq b + \frac{1}{2}\right)$$

$$\Rightarrow P(X \leq x) = P(X \leq x + 0.5) \approx P\left(Z \leq \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

$$\Rightarrow P(X \geq x) = P(X \geq x - 0.5) \approx P\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}} \leq Z\right)$$

ex 77) $X \sim \text{Bin}(200, 0.4)$

$$np = 80$$

$$\frac{80}{2} = 40$$

$$a) P(X < 70) = P\left(Z \leq \frac{70 + 0.5 - 200 \times 0.4}{\sqrt{200 \times 0.4 \times 0.6}}\right)$$

$$= P(Z \leq -1.37) = 0.0853$$

$$b) P(70 \leq X \leq 90) = P(69.5 \leq X < 90.5)$$

$$= P\left(\frac{69.5 - 80}{\sqrt{48}} \leq Z \leq \frac{90.5 - 80}{\sqrt{48}}\right)$$

$$= P(-1.516 \leq Z \leq 1.516) \rightarrow \text{from the table}$$

$$c) P(X = 80) = P(79.5 \leq X \leq 80.5)$$

$$= P\left(\frac{79.5 - 80}{\sqrt{48}} \leq Z \leq \frac{80.5 - 80}{\sqrt{48}}\right)$$

Poisson distribution is basically the limit of binomial when number of trials n goes to ∞ Thursday 4th December

$$X \sim \text{Pois}(\lambda) \text{ if } \lambda > 5 \quad \sqrt{EX} = E[X] = \lambda$$

$$\lambda = np \quad Z = \frac{X - \lambda}{\sqrt{\lambda}} \sim N(0, 1)$$

Ex: 4-78 | $X \sim \text{Poi}(6) \quad E[X] = 6 \quad \sigma = \sqrt{6}$

a) Exact $P(X < 4) = P(X \leq 3) = \sum_{i=0}^3 \frac{e^{-6} 6^i}{i!} = 0.1512$

b) approx $P(X < 4) \approx P\left(Z < \frac{4-6}{\sqrt{6}}\right) = P(Z < -0.82) = 0.20618$

* but these numbers are far from each other so we use continuity correction

$$P(X < 4) = P(X \leq 3) = P\left(Z \leq \frac{3+0.5-6}{\sqrt{6}}\right) = P(Z \leq -1.02) = 0.1539$$

c) approx $P(8 < X < 12) = P\left(\frac{8-6}{\sqrt{6}} < Z < \frac{12-6}{\sqrt{6}}\right)$

$$= P(0.82 < Z < 2.45) = 0.199$$

$$P(9 \leq X \leq 11) = P\left(\frac{9-0.5-6}{\sqrt{6}} \leq Z \leq \frac{11+0.5-6}{\sqrt{6}}\right) = P(1.02 \leq Z \leq 2.25) = 0.1416$$

Exponential dist 'opposite of poisson' to model the time until an event occurs

ex: time until next customer arrives.

- time until next failure
- time until car crossing traffic light.

In poisson we are interested in the number of arrivals in t time units $N \sim \text{pois}(\lambda)$

$$E[N] = \lambda t \quad \lambda \rightarrow \text{mean per time unit}$$

$t \rightarrow \text{time units}$

$$P(N(t)=n) = \frac{e^{-\lambda t} (\lambda t)^n}{n!}$$

Time for next arrival $\equiv$ the interval has no arrivals.

$$P(N(t)=0) = \frac{e^{-\lambda t} (\lambda t)^0}{0!} = e^{-\lambda t} \quad \text{prob of no arrivals in the interval } (0, t)$$

X : time until an arrival from any point

$$\{N(t)=0\} = \{X > t\}$$

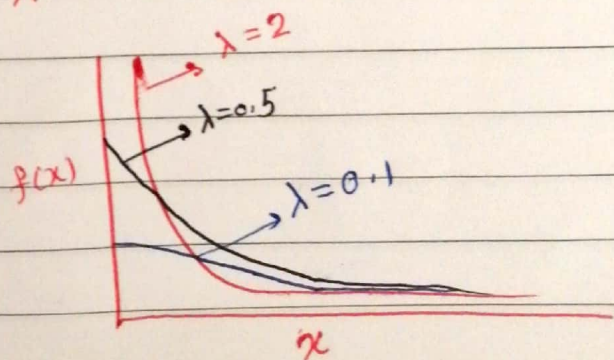
$$P(N=0) = P(X > x) = e^{-\lambda x}$$

$$P(X \leq x) = 1 - P(X > x) = 1 - e^{-\lambda x} \quad \text{CDF for } X$$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x \geq 0 \\ 0 & , \text{otherwise} \end{cases}$$

If X is a r.v. represents the time between events

$$X \sim \text{exp}(\lambda) \quad E[X] = 1/\lambda \quad \text{Var}[X] = 1/\lambda^2$$



4-90] $X \sim \exp(2)$

- a) $P(X \leq 0) = \int_0^0 \lambda e^{-\lambda x} dx = 0$
- b) $P(X \geq 2) = \int_2^{\infty} 2e^{-2x} dx = -e^{-2x} \Big|_2^{\infty} = e^{-2 \cdot 2} = e^{-4} = 0.0183$
- c) $P(X \leq 1) = \int_0^1 2e^{-2x} dx = -e^{-2x} \Big|_0^1 = -e^{-2} + 1 = 0.8647$
- d) $P(1 < X < 2) = \int_1^2 2e^{-2x} dx = -e^{-2x} \Big|_1^2 = -e^{-4} + e^{-2} = 0.117$
- e) $P(X \leq x) = 0.05 = \int_0^x 2e^{-2t} dt = -e^{-2t} \Big|_0^x = 1 - e^{-2x} = 0.05$

$$e^{-2x} = 0.95 \Rightarrow -2x \ln e = \ln 0.95 \quad x = 0.0256$$

改善

KAIZEN

TEAM

4-90] $X \sim \exp(2)$

- a) $P(X \leq 0) = \int_0^0 \lambda e^{-\lambda x} dx = 0$
- b) $P(X \geq 2) = \int_2^{\infty} 2e^{-2x} dx = -e^{-2x} \Big|_2^{\infty} = e^{-2 \cdot 2} = e^{-4} = 0.0183$
- c) $P(X \leq 1) = \int_0^1 2e^{-2x} dx = -e^{-2x} \Big|_0^1 = -e^{-2} + 1 = 0.8647$
- d) $P(1 < X < 2) = \int_1^2 2e^{-2x} dx = -e^{-2x} \Big|_1^2 = -e^{-4} + e^{-2} = 0.117$
- e) $P(X \leq x) = 0.05 = \int_0^x 2e^{-2t} dt = -e^{-2t} \Big|_0^x = 1 - e^{-2x} = 0.05$
 $e^{-2x} = 0.95 \Rightarrow -2x \ln e = \ln 0.95 \Rightarrow x = 0.0256$

Consistent units between x & λ

Sunday
Dec 8th

4-93] Counts recorded by a Geiger counter follow a poisson dist with avg 2 counts/min

a) prob at no counts in 30 seconds interval

X_i denotes the time between arrivals

$X \sim \exp(\lambda)$

$$P(X > 0.5) = \int_{0.5}^{\infty} 2e^{-2x} dx = -e^{-2x} \Big|_{0.5}^{\infty} = 0.3679$$

b) prob that first count occurs in less than 10 seconds

$$P(X < \frac{10}{60}) = \int_0^{\frac{1}{6}} 2e^{-2x} dx = 1 - e^{-\frac{1}{3}} = 0.2835$$

c) prob 1st count occurs between 1 & 2 mins

$$P(1 < X < 2) = 0.1170$$

4-21] Computer usage log ons pois(25) $\lambda = 25$ log-on/hr

X : time between log-ons $X \sim \exp(25)$

a) prob at no log-ons in the first 6 minutes

$$P(X > \frac{6}{60}) = P(X > 0.1) = \int_{0.1}^{\infty} 25e^{-25x} dx$$

b) prob at next log-on between 2 & 3 minutes

$$P(\frac{2}{60} < X < \frac{3}{60})$$

c) Interval of time such that prob that no log-ons occurs is 0.9

$$P(X > x) = 0.9 = \int_x^{\infty} 25e^{-25u} du = 0.9$$
$$= e^{-25x} = 0.9 \Rightarrow -25x \ln e = \ln 0.9$$
$$x = 0.00421 \text{ hours} = 0.25 \text{ minutes}$$

d) mean & variance

$$\mu = \frac{1}{\lambda} = \frac{1}{25} = 0.04 \text{ hr} = 2.4 \text{ min expected time b/w 2 arrivals.}$$

$$\sigma = \frac{1}{\lambda} = \frac{1}{25} = 0.04 \text{ hr}$$

- Lack of memory property - *fish memory*

it doesn't matter where you are ρ is constant.

Starting point doesn't matter

$$P(x < t_1 + t_2 \mid x > t_1) = P(x < t_2)$$

Given that you already waited t_1 , the prior dist is exactly the same.

proof:

$$= P(t_1 < x < t_1 + t_2 \mid x > t_1) = \frac{F(t_1 + t_2) - F(t_1)}{1 - F(t_1)}$$
$$= \frac{[1 - e^{-(t_1 + t_2)\lambda}] - [1 - e^{-t_1\lambda}]}{1 - [1 - e^{-t_1\lambda}]}$$
$$= \frac{e^{-t_1\lambda} - e^{-t_1\lambda} \cdot e^{-t_2\lambda}}{e^{-t_1\lambda}} = \frac{e^{-t_1\lambda} [1 - e^{-t_2\lambda}]}{e^{-t_1\lambda}} =$$
$$= F(t_2) - P(X < t_2) = F(x) = 1 - e^{-x\lambda}$$

98) let x denote the time until the arrival of a taxi

$x \sim \exp(\lambda)$ average time between arrival is 2 minutes

$\lambda = 1/2$ arrival per minute

a) prob that taxi arrives after 2 minutes

DC $X >$

Tuesday

10-12-19

4-98 Time between arrival of messages at a computer is exponentially dist with mean of 3hrs

a) what is the prob that you do not receive a message during 3 hrs period?

$$X \sim \exp(1/3) \quad P(X > 3) = \int_3^{\infty} \frac{1}{3} e^{-1/3 x} = -e^{-x/3} \Big|_3^{\infty} = e^{-1} = 0.3679$$

b) "open book" $\equiv$ part a because lack of memory

$$P(X > 7 | X > 4) = P(X > 3)$$

c) expected time between 5th and 6th messages
 $= 3 \text{ hrs}$ $E(X) = 3 \text{ hr}$ constant

* exponential distributions is the continuous analogue to geometric distributions

Erlang dist Generalization of exponential

expo $\Rightarrow$ until first arrival

Erlang $\Rightarrow$ until r arrival

ex) assume number of failures of CPU's at a large computer system is governed by a poisson process
rate $\Rightarrow \lambda = 0.001$ failures per hour, what is the prob that the fourth failure occurs after 40,000 hr

$P(X_4 > 40,000) \equiv$ that means we already had max of 3 failures before 40,000 $P\{X_4 > 40,000\} \equiv P\{N(40,000) \leq 3\}$

$$P\{N(40,000) \leq 3\} = \sum_{n=0}^3 \frac{e^{-\lambda t} (\lambda t)^n}{n!} \quad \lambda=0.0001, t=40,000 = 0.433$$

In general

$$\{X_r > t\} = \{N(t) \leq r-1\} = \sum_{n=0}^{r-1} \frac{e^{-\lambda t} (\lambda t)^n}{n!} = P(X_r > t)$$

$$CDF \Rightarrow P(X_r \leq t), \quad \frac{d}{dt} F(t) = f(t) = \frac{\lambda^r t^{r-1} e^{-\lambda t}}{(r-1)!} \Leftarrow pmf$$

if $r=1$ Erlang dist $\Rightarrow$ expo dist

$$X \sim \text{Erlang}(r, \lambda) \quad E[X] = \frac{r}{\lambda} \quad V[X] = \frac{r}{\lambda^2}$$

Gamma dist same as Erlang but r is continuous

$$X \sim \text{gamma}(r, \lambda) \quad E[X] = \frac{r}{\lambda} \quad V[X] = \frac{r}{\lambda^2}$$

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}$$

$\lambda =$ scale parameter $r =$ shape parameter

if r is integer $\Gamma(r+1) = r \Gamma(r), \Gamma(r) = (r-1)!$

- know the parameters and $E(x)$ and $V(x)$ only for the rest at dist as concepts only and not to solve on them

$$\text{Beta} \quad E[X] = \frac{\alpha}{\alpha+\beta} \quad V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$\text{log} \quad X = e^W \rightarrow N(\mu, \sigma^2) \quad \log \text{norm} \Rightarrow X \quad \underline{\ln X = W}$$

ch 6

♥ Thursday
12-12-19

Descriptive Statistics

statistics can be divided into 3 categories

1] Probability likelihood 'chance' of an event
ch 2, ch 3, ch 4

experiments → sample space → events →
→ counting technique → probability →
→ observing r.v. → represent data
by a distribution.

2] Descriptive stat

Data → Describe sample

→ Summary of data

Data a) Graphical: data plot, stem & leaf
histogram, Box plot

b) numerical summary statistics

1) like mean: typical, expected

2) Median: half of data to left

3) Mode: most common value

4) Range: Gives an idea about spread of data

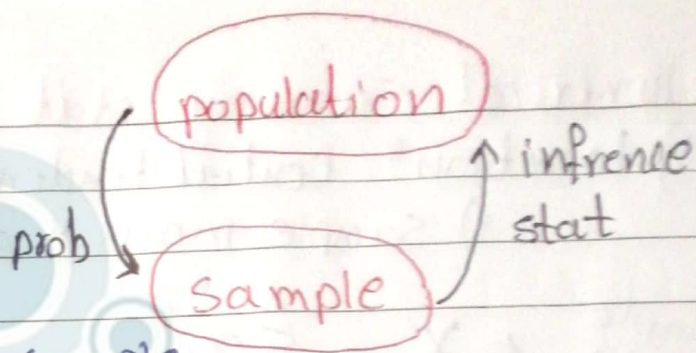
5) standard deviation: " "

6) correlation: see how they are connected

3] Data analysis

- confidence intervals } inferential
- hypothesis test } stat

inferential stat: Draw conclusions and make predictions about the population based on the sample



calculated measures $\Rightarrow$ population = "true value" μ

calculated measures $\Rightarrow$ sample = "statistics" estimates $\bar{X}$

X length of the bag

data 1.5m, 1.510m, 1.4805, 1.48m } observed value.

Types of Data

- 1) Quantitative
 - 1) interval $^{\circ}\text{C}$ Temp = 0.0°C
 0 has no meaning ex: temp/time
 - 2) Ratio: Temp Kelvin 0
 0 has a meaning
ex: temp, time, length, height.
- 2) Qualitative
 - a) Nominal: categories of objects without being ordered
ex: types of PCs : Hp, ASUS, DELL, Apple
 - b) ordinal: categories but the order is meaningful
ex: sale 1-7, college year.

Numerical Summary stat 'Descriptive'

1) Location "central tendency" measure

1) sample mean "arithmetic average"

$$\text{sample mean } \bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\text{population mean } \mu = \sum x_i f(x_i) = \int x f(x)$$

if equally likely $f(x_i) = \frac{1}{N}$

sample mean $\bar{X}$ is a reasonable estimate of the population mean μ $\hat{\mu} = \bar{X}$

2) mode: commonly used with nominal data (yes, no)

3) median: " " " " ordinal, interval

Variability scatter / spread of the data

Sunday

15-12-19

measures of variability

1) variance: for observations $x_1, x_2, \dots, x_n$

$$\text{sample variance } S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}, \sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

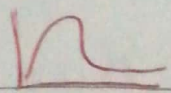
$n \rightarrow$ sample size, $N \rightarrow$ population size

x_i tends to be close to the mean $\bar{x}$ so we divide by $n-1$ to make an inflammation of S

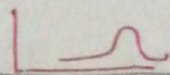
2) Range: max - min Range $\leftrightarrow$ median variance $\leftrightarrow$ mean

3) skewness: measure of the lack of symmetry of the distribution

$$\hat{\beta}_3 = \frac{\sum (x_i - \bar{x})^3}{S^3}$$



Right skewed



left skewed

4) Order statistics: sort ascending from smallest to largest $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)}$

i^{th} order element $x_{(i)}$

$\min(x_i) = x_{(1)}$, $\max(x_i) = x_{(n)}$, $\text{Range} = x_{(n)} - x_{(1)}$

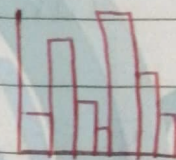
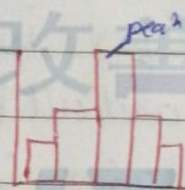
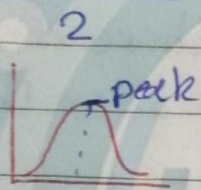
median $\tilde{x}$ if n is odd

$\tilde{x} = x_{(\frac{n+1}{2})}$ ex: 2, 4, 5, 9, 10 $\tilde{x} = x_{(\frac{5+1}{2})} = x_{(3)} = 5$

median $\tilde{x}$ if n is even

$\tilde{x} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$

5) mode:



peak $\Rightarrow$ mode = mean = median

uni and unimodal

multi

6) Quartile: 1) sort the data before

① first quartile Q_1 : 25% of data is less

② second quartile Q_2 : 50% of data is less "median"

③ third quartile Q_3 : 75% of data is less

④ inter quartile $IQR = Q_3 - Q_1$
Range

Q_1 : observation at position $\frac{n+1}{4}$

Q_3 : observation at position $3(\frac{n+1}{4})$

if the results is not integer $\Rightarrow$ interpolation

ex: $n=10 \Rightarrow Q_1 = \frac{10+1}{4} = 2.75$ $2^{\text{nd}} < Q_1 < 3^{\text{rd}}$

$Q_1 = x_2 + 0.75(x_3 - x_2)$

$Q_3 = 3(\frac{10+1}{4}) = 3 \times 2.75 = 8.25$

ex: \$ X: 30,000 38,000 42,000 57,000 73,000
 \$ Y: " " " " 244,000

outlier ↗

$\bar{X}$: $\frac{\text{add}}{\text{number}} = 48,000 \$$

$\tilde{x}$: 42,000 \$

$\bar{Y}$: 82,000 → weak because of outlier

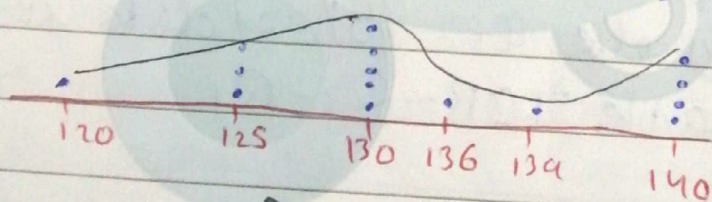
$\tilde{Y}$: 42,000

Solve [6-1] Can $S=0$? yes, when there is no variance and all the data are the same

Solve [6-10] $\bar{X} = \frac{\sum x_i}{n-1} = 126.22 \text{ (kN)}$
 $s^2 = 683.24 \text{ (kN)}^2$

Graphical Summary

1 Dot Plot: ex: 120, 125, 125, 125, 130, 130, 130, 130, 130
 136, 139, 140, 140, 140, 140



2 Stem and leaf: good for visualization with keeping actual data points

Divide each data point into a stem & leaf

Array in order Smallest → largest

ex: 123, 154, 150, 161, 189, 140, 165, 168, 191

sol: always sort before

stem	leaf	count
12		
13	9	1
14		1
15		1
16	0 4	3
17	1 5 8	3 (3) → median is here so stop
18	9	3
19	0 1	2

* if the leaf has too much data then divide the stem into 2 section High and Low

Tuesday
17-12-19

[3] Frequency Histogram

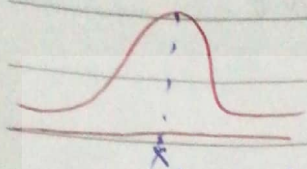
each bar is called a class and it's height is called a frequency

- 1) Divide range of values into class/interval/cell/bin
- 2) number of bins $= \sqrt{n} \Rightarrow (5-20)$ "Rule of thumb"
- 3) width of the interval $= \frac{\text{range}}{\text{\#bin}}$ $[b_0, b_1), [b_1, b_2), [b_2, b_3)$
- 4) count x_i 's that fall in $b_0 \leq x < b_1$ $C_j \Rightarrow$ frequency of interval
relative frequency $h_j = \frac{C_j}{n}$ $\sum C_j = n$, $\sum h_j = 1$

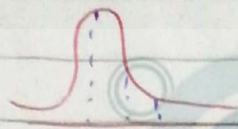
when asked for cumulative: start from the smallest and go up until you reach $F(x) = 1$

advantages: Gives an idea about the following

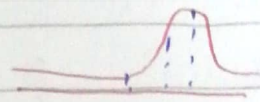
- shape of the dist
- Central tendency
- scatter data



mean, mode, median



mean > median > mode



mode > median > mean

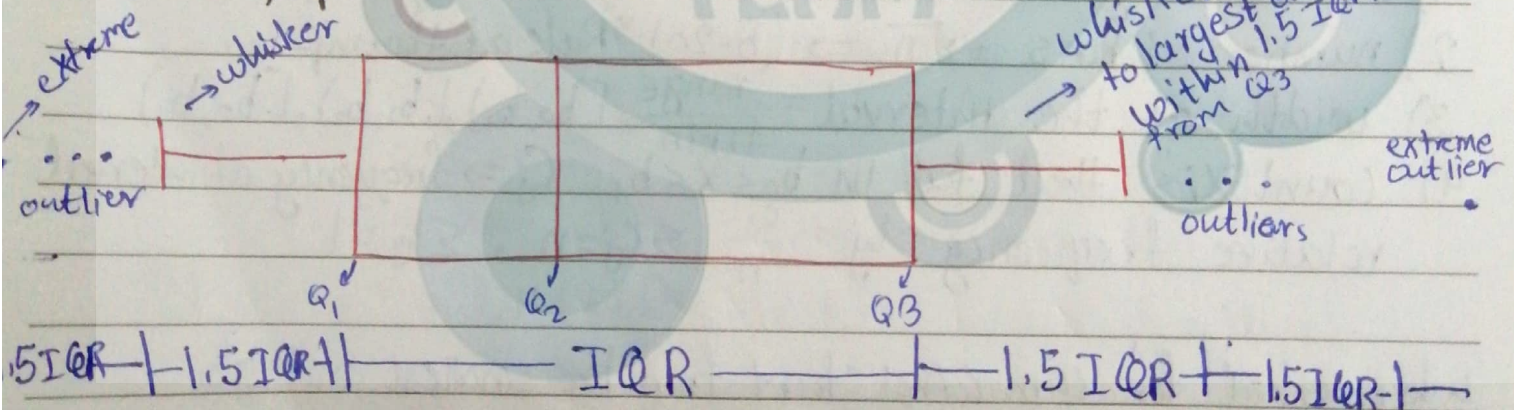
if the bin range are not equal $\text{Height} = \frac{\text{area}}{\text{width}}$

ex:	range	Count	%	width	Height
	0-300	5	5/17	300	29/300
	300-600	8	8/17	300	47/300
	600-1000	4	4/17	400	23/400

4 Box Plot

compact information: min, max, median, Q_1 , Q_3 , outlier
outlier $\rightarrow$ unusual data point

- 1) Order statistics: $x_{(1)}, x_{(2)}, \dots, x_{(n)}$
- 2) Find Q_1, Q_2, Q_3
- 3) plot within 1.5 IQR



5 Time Series

~ will take in other courses, skip ~~for~~ now.

Probability Plots for any dist

Thursday ♥
19-12-19

which dist is a reasonable model for the data?

Assessing Normality:

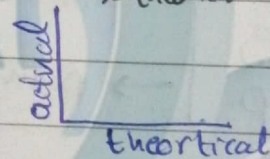
- 1) visual - Histogram, prob plots $\left\{ \begin{array}{l} \rightarrow Q-Q \\ \rightarrow P-P \end{array} \right.$
- 2) statistics - Kolmogorov-Smirnov, Anderson

the procedure: Q-Q Plot

$$\therefore q_i = \frac{i-0.5}{n}$$

- 1) sort data: $x_{(1)}, x_{(2)}, \dots, x_{(n)}$
- 2) calculate cumulative frequency $F(x_i) = q_i = \frac{i-0.5}{n}$
- 3) $Z_{q_i} \Rightarrow Z_{q_i} = F^{-1}(q_i)$
- 4) calculate actual Z from data
- 5) plot $Z_{\text{theoretical}}$ vs Z_{actual} or $Z_{\text{theoretical}}$ vs x_{actual}
or q_i vs x_{actual}

if the plot is linear then it's normal dist



P-P Plot

- 1) sort data
- 2) $q_i = \frac{i-0.5}{n}$
- 3) compute cumulative $F(x_i)$
- 4) plot $F(x_i)$ vs q_i

i	x	$\frac{i-0.5}{n}$	$Z_{\text{theoretical}}$ from graph	Z_{actual}
-----	-----	-------------------	--	---------------------

Chapter 7

Sampling Distribution

Sunday
22nd Dec

* only study this chapter as what we will take

Random sample $(X_1, X_2, \dots, X_n)$

- X 's are independent because the population is huge
- identically distributed every X is taken from the same pop

* what we calculate from the sample is called

Statistics 'estimated Quantities' X is r.v.

- $\bar{X}$: is a statistics which is random it has a dist $\Rightarrow$ sampling dist

• experiment $\rightarrow$ sample $\rightarrow$ statistics $\rightarrow$ estimate pop parameters

like μ, σ, ρ to get a good guess

$\hat{\mu} = \bar{X}$ } sample average is a point estimator of μ pop

$\hat{\sigma} = s$

$\hat{\rho} = r$ } $\rightarrow$ point estimators

* Sampling distribution at the sample mean $\bar{X}$

$X_1, X_2, \dots, X_n$ random sampling from a population with μ, σ^2

The distribution at the averages is normal

each observation $X_i \sim N(\mu, \sigma^2)$ $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$

$\bar{X}$ is also normally dist

$$\begin{aligned} \boxed{\mu_{\bar{X}}} &= E\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \frac{E[X_1] + E[X_2] + \dots + E[X_n]}{n} \\ &= \frac{\mu + \mu + \dots + \mu}{n} = \frac{n\mu}{n} = \boxed{\mu} \end{aligned}$$

$$\begin{aligned} \boxed{V[\bar{X}]} &= V\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \frac{1}{n^2} V[X_1 + X_2 + \dots + X_n] = \\ &= \frac{\sigma^2 + \sigma^2 + \dots + \sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \boxed{\frac{\sigma^2}{n}} \end{aligned}$$

Central Limit Theorem

CLT

for $x_1, x_2, \dots, x_n$ random variable from a population with μ, σ^2 $X \sim N(\mu, \sigma^2)$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

this is valid for any dist with $n \geq 30$ or for dist that are nearly normal $n < 30$

7-3 PVC with mean 1.01 in. $\sigma = 0.003$ in. if you took a sample what is the prob the mean is less than 1.012 or greater than 1.009

$$P(1.009 < \bar{X} < 1.012) = P\left(\frac{1.009 - 1.01}{0.003/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{1.012 - 1.01}{0.003/\sqrt{n}}\right) \\ = P(-1 \leq Z \leq 3) = P(Z \leq 3) - P(Z \leq -1) = 0.9186$$

7-5 $\mu_{\bar{X}} = 75.5$ Psi $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3.5}{\sqrt{6}} = 1.429$

$P(\bar{X} \geq 75.75) =$

$$P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{75.75 - 75.5}{1.429}\right)$$

Tuesday
24-12-19

The last lecture ♥

$$P(Z \geq 0.175) = 0.43055$$

Solve 7 $\rightarrow$ H.W.

7-9 $\mu = 137$ $\sigma^2 = 36$, $n?$ $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = 1.2$
 $1.2 = \frac{6}{\sqrt{n}} \Rightarrow n = 25$

if it's 11.12 always take the largest so it becomes 12
but in this example it came to be 25 exact

Approximate sampling dist of a difference

Pop₁ $X_{11}, X_{12}, \dots, X_{1n}$
Pop₂ $X_{21}, X_{22}, \dots, X_{2n}$

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

7-13 $n_1 = 16$ $\mu_1 = 75$ $\sigma_1 = 8$
 $n_2 = 9$ $\mu_2 = 70$ $\sigma_2 = 12$

$P(\bar{X}_1 - \bar{X}_2 \text{ exceeds } 4)$

$$P(\bar{X}_1 - \bar{X}_2 > 4) = P\left(\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > \right)$$

$$P\left(Z > \frac{4 - 5}{\sqrt{\frac{8^2}{16} + \frac{12^2}{9}}}\right)$$

$$= P(Z > -0.2236)$$

$$= 1 - P(Z < -0.2236) = 0.5885$$

$$P(2.5 \leq \bar{X}_1 - \bar{X}_2 \leq 5.5)$$

$$P\left(\frac{2.5 - 5}{\sqrt{\frac{8^2}{16} + \frac{12^2}{9}}} \leq Z \leq \frac{5.5 - 5}{\sqrt{\frac{8^2}{16} + \frac{12^2}{9}}}\right)$$

* End of stat I Fall 2019 *
~ Wadees Rahabeh