Assuming that A & ϕ are independent.

Also: $x(t,\phi) = A\sin(\omega_c t + \phi)$, as here ϕ is varying. → Angle of frequency.

Computing: $E[x(t,\phi)] = E[A\sin(\omega_c t + \phi)]$

$$= E[A]\,E[\sin(\omega_c t + \phi)] \quad \text{(as } A \text{ \& } \phi \text{ are ind)}$$

$$= \mu_A \int_{-\pi}^{\pi} \underbrace{\frac{1}{2\pi}}_{\text{(Normalizing)}} \sin(\omega_c t + \phi)\, d\phi$$

$$= \frac{-\mu_A}{2\pi}\Big[\cos(\omega_c t + \phi)\Big]_{-\pi}^{\pi} = \frac{-\mu_A}{2\pi}$$

$$= \frac{-\mu_A}{2\pi}\{\cos(\omega_c t + \pi) - \cos(\omega_c t - \pi)\}$$

$\cos(A) - \cos(B) = 2\sin\left(\frac{A+B}{2}\right)\sin\left(\frac{A-B}{2}\right)$

$$= \frac{-\mu_A}{2\pi}\, 2\sin\left(\frac{2\omega_c t}{2}\right)\sin\left(\frac{2\pi}{2}\right)$$

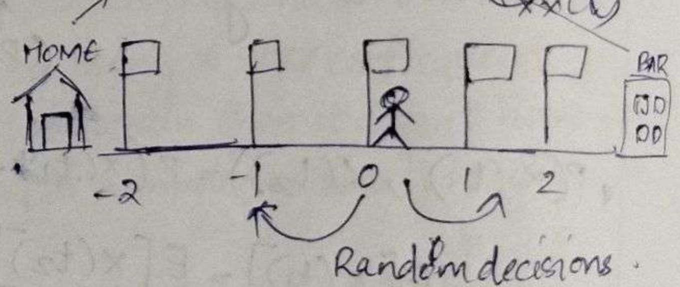
$\sin\left(\frac{2\pi}{2}\right) = \sin\pi = 0.$

$E[x(t,\phi)] = \underline{\underline{0}}$. (Try for $R_{xx}(t)$)

~~$R_{xx}(t)$~~

Drunkards Problem

Random
* Transitions of state.
* Random process



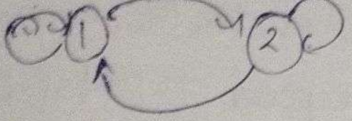
Assumption of multiple variables,

$P(x,y) = P(x)\,P(y)$ are x & y are independent ~~they~~ then they are PDF. → ~~Any~~ process (Random process)

For dependency also, ~~the prob things~~ there will be PDF associated.

* Andrey Markov proved this (dep with PDF)
  By bringing dependency & converging (run for several time)

No. of bl = 3000
N. of white = 2000
$P(w) = \frac{2000}{5000}$



Black bead , white bead

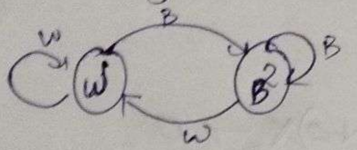
Markov chain

No. of black beads = 3000   No. of white beads = 2000.

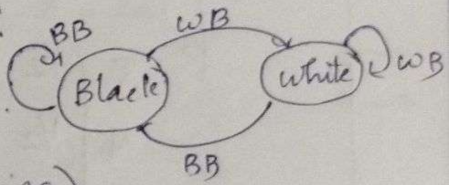
Here $P(w) = \frac{2000}{5000}$, even if we take any one of white beads.

so here its converging to a probability, but there is no dependency

Now bringing ~~day~~ dependency (Acc to Markov)
After mixing white & Black beads in 2 bowls, now we ~~cla~~ separate them by: ~~If~~ In Bowl I, if we take black bead transfer to Bowl II and if its white keep in the same bowl, and vice versa in Bowl II. So now we will get a dependency and converging by running this process for n times.



This transition chain is called as 'Markov chain'.



13/08 To find how it converges in long term.
(Program) to find probability (Find converging).

~~set~~ (Gaussian Process :- Random variable process applying on Gaussian variables)

Random Process / Stochastic Process

Specify a Markov chain

* A set of states $S = \{S_1, S_2 \ldots S_n\}$.
* The process starts from any state S_i and moves (step) to another state S_j.
* Each move is called as a step.
* Current state is S_i and move to S_j, Probability becomes P_{ij}

$P_{ij} \rightarrow$ transition probabilities

* Probability does not depend on which states the chain was before the current state (does not depend on past states)

Land of Oz (Not blessed with good weather)

* Never have two nice days in a row
* If they have a nice day then they are just as likely to have Snow as Rain
* If they have Snow or rain they have even chance of having the same next day.
* If there is change from snow or rain only half of the time is this a change to a nice day. (Increasing this pblty converging more)

SYRSN (Rain Snow Nice)

$$P \quad \begin{array}{c|ccc} & R & N & S \\ \hline R & 1/2 & 1/4 & 1/4 \\ N & 1/2 & 0 & 1/2 \\ S & 1/4 & 1/4 & 1/2 \end{array}$$

$(N \to N \cancel{0}) X \to 0$

$P = P_{ij} \to$ one of entry in transition matrix

$P_{ij}^{(2)} \to$ one transition, then another transition from it

$$P_{13}^{(2)} = P_{11}P_{13} + P_{12}P_{23} + P_{13}P_{33} = \frac{1}{2}\times\frac{1}{4} + \frac{1}{4}\times\frac{1}{2} + \frac{1}{4}\times\frac{1}{2}$$

$$= \frac{3}{8}$$

Multiplying P again with P.

(dot product → P_{13})

$$\begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

$\cancel{\frac{1}{4} + \frac{1}{8} + \frac{1}{16} \quad \frac{1}{8} + \frac{1}{16}}$

$= \frac{1}{4} + \frac{1}{8} + \frac{1}{16} \qquad \frac{1}{8} + \frac{1}{16} \qquad \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$

$\frac{1}{4} + \frac{1}{8} \qquad \frac{1}{8} + \frac{1}{8} \qquad \frac{1}{8} + \frac{1}{4}$

$\frac{1}{8} + \frac{1}{8} + \frac{1}{8} \qquad \frac{1}{16} + \frac{1}{8} \qquad \frac{1}{16} + \frac{1}{8} + \frac{1}{4}$

4+2+1

$= \frac{7}{16} \quad \frac{3}{16} \quad \frac{3}{8}$

$\frac{3}{8} \quad \frac{2}{8} \quad \frac{3}{8} \quad \frac{3}{8} \quad \frac{3}{16} \quad \frac{7}{16}$

19/08 Let transition matrix be P, where ij^{th} entry $P_{ij}^{(n)}$ $(P \cdot P = P^2)$

P^n be probability that Markov chain starting in state S_i will be in state S_j after 'n' steps (like P^2)

or Power Iterations (Multiplying matrix 'n' times).

* Let u (S_i) be initial state of Markov chain, probability vector.

$\therefore$ ~~$u^n = uP^n$~~ $\therefore \boxed{u^n = uP^n}$

$$P' = \begin{array}{c} \begin{array}{ccc} R & N & S \end{array} \\ \begin{bmatrix} 0.50 & 0.25 & 0.25 \\ 0.50 & 0.00 & 0.50 \\ 0.25 & 0.25 & 0.50 \end{bmatrix} \end{array} \begin{array}{c} R \\ N \\ S \end{array}$$

$$P^2 = P' \cdot P' = \overset{P'}{[\,]}\overset{P'}{[\,]} = \begin{bmatrix} 0.438 & 0.188 & 0.375 \\ 0.375 & 0.250 & 0.375 \\ 0.375 & 0.188 & 0.438 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0.406 & 0.203 & 0.391 \\ 0.375 & 0.250 & 0.375 \\ 0.375 & 0.188 & 0.438 \end{bmatrix} \quad " \begin{bmatrix} 7/16 & 3/16 & 3/8 \\ 3/8 & 2/8 & 3/8 \\ 3/8 & 3/16 & 7/16 \end{bmatrix}$$

Let Initial vector, $u = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}$

$$u^3 = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} \overset{P^3}{[\;]} = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} \underbrace{\begin{bmatrix} 0.406 & 0.203 & 0.391 \\ 0.375 & 0.250 & 0.375 \\ 0.375 & 0.188 & 0.438 \end{bmatrix}}_{3\times 3}$$

$P^4 = \ldots \quad P^5 = \ldots \quad P^6 = \ldots \quad \ldots P^n =$ At a point it may converge or probability may get fixed to a point.

H H ... H — Next toss may or maybe H or T. but most chance
1 2 ... 49 — is for T, because to satisfy (maintain) the ratio 1:1 (1/2 : 1/2) (1/2 H : 1/2 T)

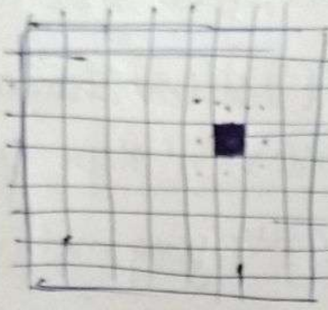
$$u^3 = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix} \Big[$$

Game of Life

$n \times n$

Assume grids with

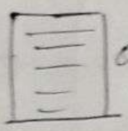
→ Green
→ Blue
→ Red → White



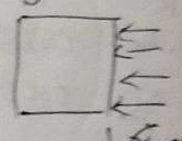
Part of cellular automata (Rotating)

→ Randomly chosen, then neighbourhood should be same colour

* WWW as Chain
* Considered as "Random Surfer" Model (From one page to another reaching to same more than once)
* Page Rank : Algorithm



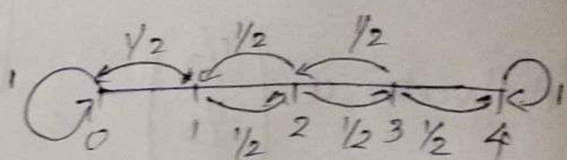
outer links



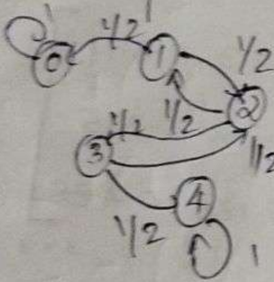
incoming links (imp)

02/09 Drunkards walk / Random walk

Absorbing Markov chain:-



* can also be represented as states.



* arrows rep state transition probability
* circles rep states
* Reach at home (0) or bar (4)

* To find the probability, convert this into P(T) matrix and do power iterations, we will converge to a probability.

$$P = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1/2 & 0 & 1/2 & 0 & 0 \\ 2 & 0 & 1/2 & 0 & 1/2 & 0 \\ 3 & 0 & 0 & 1/2 & 0 & 1/2 \\ 4 & 0 & 0 & 0 & 0 & 1 \end{array}$$

1, 2, 3 → transient states (move either from 1, 2, 3)

0, 4 → absorbing states. (finally remain that state)

$\therefore u^n = u P^n$ (u → initial vector (start from anywhere))

(Consider initial state from transient states, if absorbing states are taken, it remains there itself)

03/09 Drunkard's Walk / Random walk

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{l} \to 1 \\ \to 1 \end{array} \longleftarrow \text{Stochastic Matrix}$$

(contains probability values)
(Row wise adding getting 1)
Row stochastic matrix)

(col stoch) not 1's

Canonical form

$$P = \begin{array}{c} \\ Tr \\ Abs \end{array} \left(\begin{array}{c|c} Tr & Abs \\ Q & R \\ \hline O & I \end{array} \right)$$

$$P^n = \left(\begin{array}{c|c} Q^n & * \\ \hline 0 & I \end{array} \right)$$

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 0 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccc|cc} 1 & 2 & 3 & 0 & 4 \end{array} \\ \left[\begin{array}{ccc|cc} 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] \end{array}$$

a) what is the probability that the process will end up in a given absorbing state?

b) on an average, how long will it take for the process to be absorbed?

c) On an average, how many times will the process be in each transient state?

Theorem: In an absorbing Markov chain, the probability that the process will be absorbed is 1.

ie, $Q^n \to 0$ as $n \to \infty$ (surely in absorbtion state).

Fundamental Matrix

For an absorbing Markov chain, the matrix $I-Q$ has an inverse N,

$$N = I + Q + Q^2 + \cdots$$

$N = (I-Q)^{-1}$ is called the fundamental matrix

n_{ij} of N gives the expected number of times that the process is in transient state S_j if it is started in S_i.

$$Q_{(\text{transient})} = \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix} \quad \therefore N = (I - Q)^{-1}$$

$$= \cancel{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} - \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix}\right]^{-1}$$

$$= 1\left(1 - \frac{1}{4}\right) + \frac{1}{2}\left(\frac{-1}{2}\right) \qquad = \begin{pmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & -1/2 & 1 \end{pmatrix}^{-1}$$

$$= 1 \times \frac{3}{4} + \frac{-1}{4} = \frac{2}{4} = \underline{\underline{\frac{1}{2}}}$$

$\frac{3}{4}$

$a_{11} = 1 - \frac{1}{4} = \frac{3}{4}$

$a_{12} - \cancel{-1}(-1 \times (\cancel{1} - 1/2) = 1/2$ $\qquad a_{31} =$

$a_{13} = 1\left(\frac{1}{4}\right) = 1/4$

$a_{21} = -1(-1/2) = 1/2$

$a_{22} = 1(1) = 1$

$a_{23} = -1(-1/2) = 1/2$

$$N = (I - Q)^{-1} = \begin{array}{c|c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \end{array}$$

On an average, if the man is in 2, the no. of times the person in 1 is 1, in 2 is 2 and in 3 is 1

Time to Absorption

* Given that the chain starts in state S_i, what is the number of steps before the chain is absorbed?

Let 't_i' be the expected number of steps before the chain is absorbed given that the chain is in S_j and let t be the column vector whose ~~ith~~ i^{th} entry is t_i, then $\underline{\underline{t = Nc}}$

← vector
1's

$t = \underline{\underline{Nc}}$ where c is vector with all ones. $\frac{4}{2}$ 2

$\frac{4}{2}^2 + 1$

$$t = \begin{bmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3/2 + 1 + 1/2 \\ 1+2+1 \\ 1/2 + 1 + 3/2 \end{bmatrix} = \begin{bmatrix} 3 \\ \cancel{8}\,4 \\ 3 \end{bmatrix}$$

3 → 0 or 4 absorbing (no transient states)

4 → 4 times in transient

3 → 0 or 4 absorbing.

23/09 Descriptive statistics (Module III).

CSV → comma separated values

Data

Getting Data in

* format ~~getting~~: CSV, Excel, json
opensource DB (Database can be used)

*

Data helps to
- visualize the data
- Get insights.

Preprocessing of Data :- arranging of data if any data is missing or misplaced.
(cleaning of data)

Visualizing data :- Done by tools for visualising.
* Requires huge computing power.

Get insights : finding what is present in data
- statistics of data (avg, mean, std deviation, variance).

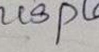
R-programming tool : helps to statistics.

(Download) RStudio.

[To learn R programming : https://r02pro.github.io]

(Kaggle). set of call for competition

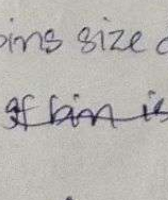
* Histogram :- has 1D data → easy to get insight.
↳ x axis - buckets / bins
→ y axis → frequency.



bins

* bins size decreases resolution increases.

* ~~If bin is closer to~~ * Discrete : graph is PMF } depending on bin size.
continuous : graph is PDF.



inorder to make ~~as~~ centered we calculate deviation (from centre to data

$\sqrt{x - \bar{x}}$

$(x-\bar{x})^2$ → variance.

Order Statistic : To order the data.

Find GDP data in Kaggle (GDP/Person)

* Quantiles * Percentiles

Let $x_1 \leq x_2 \leq \dots \leq x_n$ denote ordered elements of set {$x_1, x_2 \dots x_n$}.

* The quantile of the data for $0 < q < 1$ is $x_{[q(n+1)]}$ where $q_{(n+1)}$ is the result of rounding $q(n+1)$ to closest integer the $100p$ quantile is known as p percentile.

* 0.25 and 0.75 quantiles are known as first and third quartiles

* 0.5 quantile is known as sample median.

Median: Even: $\frac{x_{n/2} + x_{n/2+1}}{2}$ } → 0.5 quantile.

Odd: $\frac{x_{n/2+1}}{2}$

Median
Even: $\frac{n}{2} + \frac{n+1}{2}$
Odd: $\frac{n+1}{2}$

HW
R-programming
Take data from kaggle & plot mean, mode, median using R-prog

30/9. In R-Studio

* we use library package 'r02pro'.
* Other dataset is faithful
* If faithful is entered, then all the values of eruptions & waitings will be displayed.
* If head(faithful) is entered, then certain values (ie important) will be displayed.
* summary(faithful) gives you min, max, median, mean, mode of eruptions & waitings (values)

To compute mean (faithful$eruptions)

faithful dataset

boxplot(faithful$eruptions) to find outlayers (that fitting data) (if circle present)

head(faithful) some values

summary(faithful) mean, median

mean square error stops when meanse tends to zero

$y = mx + c$

model ← lm(faithful) regression fn (to fit a model) as y=mx into a data set

scatter smooth (graph)

bimodal data

2D → slope
n-D → coefficient

fit: check distance b/w point & line (minimum distance line)

$y = ax_1 + bx_2 + cx_3$ → gives a line (coefficient)

* Scatter.smooth (faithful) will display a bimodal graph, ie the values concentrated at 2 places in a graph.

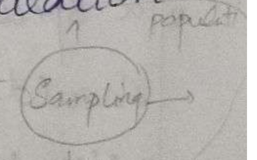
* To compute mean, mean(faithful$eruptions)
* Boxplot (faithful$eruptions) displays you a box graph, and it is used to find outlayers (ie to find the value not fitting the graph, displayed as circle)

~~To compute mode~~

* mode ← lm(faithful) is a regression function, ie used to fit a model into a dataset. Fit check is the difference between distance of point & line (min dist is considered). we follow y = mx, m → is the slope in 2D, for Multi-Dimensional it will be coefficients, $y = ax_1 + bx_2 + cx_3$ - (gives a line fitting in the plot)
* Mean-Square error stops when MSE tends to 0.
  (MSE)

01/10. Hypothesis Testing

* Hypothesis is a claim or statement about a population parameter, we will do sampling



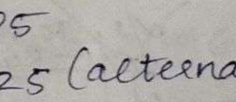
Ex: Hypothesis: Average age of students of this university is 25.

Checking this hypothesis is hypothesis testing.

Let H_0 (null hypothesis) , $H_0 : \mu = 25$
when hypothesis having =, ≥, ≤
$H_1 : \mu \neq 25$ (alternative hypothesis)

Two-tailed: Age of students will be less than 25 or age greater than 25 (no equal)

Alternative hypothesis: $\mu > 25$, $\mu < 25$



∴ $H_1 : \mu < 25$

$H_0 : \mu \geq 25$ → Null hypothesis
(equal or oppo inequal)

$H_1 : \mu > 25$
$H_0 : \mu \leq 25$ — Null hypothesis

* Alternative hypothesis determines the tail of the test.
* ~~altern~~ Alternative hypothesis never contain equal sign. and Null hypothesis always contain inequality.
* when we consider hypothesis, always we wanted to either Reject or Failed to Reject the hypothesis.

~~If Ho Ho Reject, then~~ - Reject Ho → cannot support · Failed to reject H_1

→ Failed to Reject Ho · Reject H_1

↓

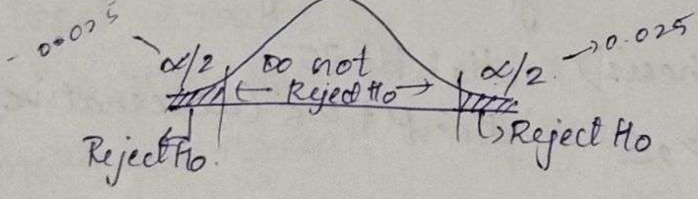
we do not have evidence to support alternative.

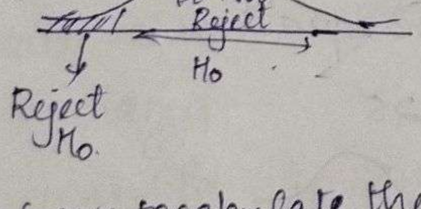
* Significance level (α) , $\alpha = 0.05$ (5%)

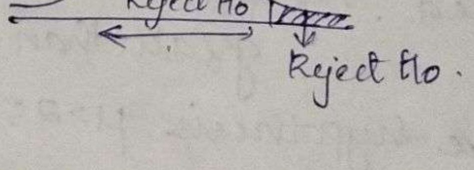
$= 0.10$ or 0.01

* α specifies the size of the region where null hypothesis should be rejected.
* Also known as critical/rejection region.

* For a two-tailed test α is divided into two.

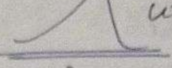




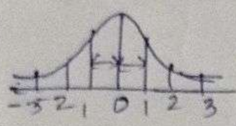


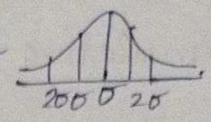
* We have to calculate the test statistic ∴ Z → test & t-test (2 tests) when data is very large.

* For a single population mean test we usually conduct either Z test (σ is known) or t-test (σ is not known) (σ → Standard deviation)

Let Standard normal distribution  with mean μ.

* z table says that each point how (σ) far sigma far from mean

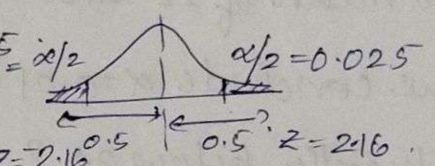
z table: 

(each will be 1σ far) 

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Example:

In recent years, mean age of all university students has been 23. A random sample of 42 students revealed a mean age of 23.8. Suppose ages are normally distributed, with a population standard deviation $\sigma = 2.4$. Can we infer at $\alpha = 0.05$ that population mean changed?



$H_0 : \mu = 23$, $H_1 : \mu \neq 23$ (alt)

0.080
0.025
0.015

$n = 42$ $\mu = ?$ $\alpha = 0.05$

$\bar{X} = 23.8$ $\sigma = 2.4$

$$z = \frac{23.8 - 23}{2.4/\sqrt{42}} = \underline{\underline{2.16}}$$

$$\boxed{z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}$$

= ~~0.5 − 0.025 = 0.475~~ → find z corresponding this value.

-0.025 → corresponding z value :- -1.96 (from cumulative less than table)

Decision :- Reject H_0 if $z < -1.96$ or $z > 1.96$.

∴ Reject H_0, ie, it supports H_1 ($H_1 : \mu \neq 23$)

If α is changes the value z changes. Depends on α.

07/10

one tailed test: 

If H_0 is null hypothesis.

Reject H_0 ; when we do not have sufficient evidence the support H_0.

Support H_1 :(ie have evidence to support H_1)

Statistical Tables :- For calculating value

* Cumulative less than table

* Cumulative from mean table

Here $2.16 > 1.96$ $\therefore$ we reject H_0.

we do not ~~en~~ have sufficient evidence to support H_0 at $\alpha = 0.05$.

May change, when α changes

~~Ho~~

Q. A random sample of 27 observation from a large population has a mean of 22 and standard deviation of 4.8. Can we conclude $\alpha = 0.01$ that population mean is significantly below 24?

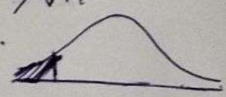
7/10

$n = 27$ $\quad \sigma = 4.8$

$\bar{x} = 22$ $\quad \alpha = 0.01$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$H_0 : \mu \geq 24$ - * One tailed test



$H_1 : \mu < 24$

(Using t-test)

1. Degrees of Freedom $df = n - 1$

$\therefore df = 27 - 1 = \underline{\underline{26}}$

Value from table :-

2.479 (one tail)

$$t = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \cancel{22 - 24} \; \frac{22-24}{4.8/\sqrt{27}} = \frac{-2}{\frac{\cancel{4.8}\;1.6}{3\sqrt{3}}} = \frac{-2 \times \sqrt{3}}{\cancel{1.6}\;0.8} = \frac{-1\sqrt{3}}{0.8}$$

$$= \frac{-\sqrt{3}}{0.8} = \underline{\underline{-2.16}}$$

$\cancel{-2.16}$ $t < -2.479$

$0.8 \overline{)1.73}$ $\quad 8\overline{)17.3}$ 2, 16

$\therefore$ Failed to reject H_0. $\quad \therefore -2.16 < -2.479$. ✗

H_1 rejected.

~~2.16 >~~

∴ ~~Ma~~ There is no evidence that the population mean is less than 24.

Type I & Type II Errors

	H_0 true	H_0 false
Reject H_0	Type I errors [false +ve]	Correct decision [power].
Fail to reject H_0	Correct Decision [Confidence]	Type II errors ~~[take~~[false -ve].

* Probability of Type I error is known as α (significance).
* Probability of Type II error is known as β.
* Probability of complement of $(1-\beta)$ is called power

Ex: Is average uber drivers age different from 40?

$H_0: \mu = 40$. Type I error:- avg age diff from 40.

$H_1: \mu \neq 40$. Type II error:- Age is 40.
concluding that the avg age is not different from 40, when infact it is different from 40.

Estimation Techniques: → Paramet
→ Non-paramet.

Parameter Estimation

* We know some parameters & distribution itself

[Frequentist Approach] → frequency taken
~~Biased Approach~~
Bayesian Approach → From prior info, can be updated

Frequentist Approach ~~* Mainly~~
(Mean Square Error)
MSE of an estimator Y that approximates a deterministic quantity $\gamma \in \mathbb{R}$.
(avg) → calculated).

$$MSE(\gamma) = E((Y-\gamma)^2) \quad \text{(Expectation)}.$$

* Calculated by Sampling

~~Q When $Y = \gamma$?~~

when $Y == \gamma$?

$$MSE(Y) = E\left((Y - E(Y))^2\right) + (E(Y) - \gamma)^2$$

Variance   Bias   ~~Bias~~ 0

If Bias $= 0$, then $E(Y) = \gamma$; $Y == \gamma$

Theorem:- Sample mean is an unbiased estimator of the mean of an iid sequence of random variables. (Sample mean can be taken from mean if its iid)

iid
~~inde~~ identically independent distribution.

10/10

$$MSE(\underset{\text{Estimator}}{Y}) = E\left[(Y - E[Y])^2\right] + (E[Y] - \gamma)^2$$

Variance   Bias

Ex: of an estimator sample mean ~~approaches to population mean~~ (un~~biased~~)

unbiased estimator $=0$ sample mean approaches to γ

consider the sample mean of an iid sequence $\tilde{x}$ with mean μ.

$y = h(x_1, x_2 \ldots x_n)$
↳ sequence

$$\tilde{Y}(n) = \frac{1}{n} \sum_{n=1}^{n} \tilde{x}(i)$$

(sample mean)   ↳ μ

$$E(\tilde{Y}(n)) = \frac{1}{n} \sum_{i=1}^{n} E[\tilde{x}(i)]$$

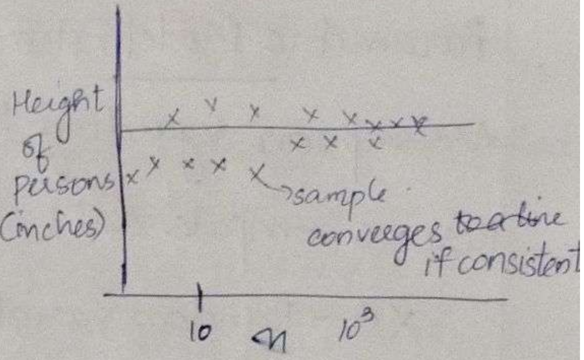
(mean)

$$= \mu$$

Consistency of the data: when more and more data is available, the estimate converges to the true value. (If diverges then it is not consistent)

Confidence Interval

* can conclude the data lie within the interval and check.



* A $1-\alpha$ confidence interval I for $\gamma \in R$ satisfy $\underbrace{P(\gamma \in I)}_{\text{within the interval}} \geq 1-\alpha$

$Y+C$ & $Y-C$ $\quad$ $c \leq C \to 1-\alpha$

$$I_n = \left[Y_n - \frac{b}{\sqrt{\alpha n}}, Y_n + \frac{b}{\sqrt{\alpha n}} \right] \rightarrow \text{confidence interval for mean.}$$

$\sigma^2 \leq b^2$ where $0 < \alpha < 1$

Estimation Technique: Non parametric Estimation :- Do not know the distribution.

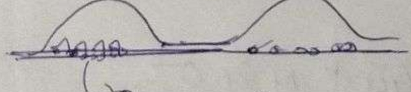
Kernal Density Estimation (to understand non-parametric estimation)

14/10 Kernal Density Estimation

* Assumption is that more data points in a sample that occur around a location

$$\hat{P}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} \underbrace{K\left(\frac{x_i - \text{obseravation}}{h} \right)}_{\text{Kernal function.}}$$

x_i - fixed location.
$K(x) \rightarrow$ The kernal function
$h \rightarrow$ Bandwidth.



The sum of ~~should~~ this would be the graph.

$$\frac{1}{nh} \sum_{x=1}^{n} K\left(\frac{x - x_i}{h} \right)$$

~~Epanechnikov K~~

Keenel	Function
uniform	$K(x) = 1/2$
Epanechnikov	$K(x) = \frac{3}{4}(1-x^2)$
Gaussian	$K(x) = \frac{1}{\sqrt{2\pi}} e^{-1/2 x^2}$

$\int K(x)\,dx = 1$, $K(x) \geq 0$.

(Total Probability should be 1 and not negative).

Parametric Model Estimation : Goal: Compute model parameters like μ, σ etc

Assumption regarding the type of distribution.

Here we compute data randomly

(Frequentist Approach: data is fixed)

$X \leftarrow$ Random variable.

$(x_1, x_2, \dots x_n) \rightarrow$ realisation of data

$(X_1, X_2 \dots X_n) \rightarrow$ Random variables

(X_1 can generate any (x_1, x_2))

1. Method of Moments (MM)

Adjust parameters of the distribution, so that the moments of the distribution coincides with the sample moments of the data.

(moments $\rightarrow$ Mean, standard deviation, variance ..)

Exponential distribution: $\mu = 1/\lambda$ (Exponential distribution parameter is λ)

Assuming we have 'n' (iid) samples $x_1, \dots x_n$.

$$\lambda_{MM} = \frac{1}{av(x_1 \dots x_n)}$$

Here we have to adjust λ.

For Gaussian distribution, we have to adjust μ & σ.
$(\mu \rightarrow av(x_1 \dots x_n))$ (σ is difference)

Maximum Likelihood Estimate (MLE)

Likelihood function is the joint pdf of the data, interpreted as ~~the~~ a function of the unknown parameters.

$x_1, x_2, x_3 \dots x_n$ are observed data which ~~the~~ are realisation of $X_1, X_2 \dots X_n$ (random variables).

Parameter (vector) → θ $\begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$

Joint probability, $P_\theta(x_1, x_2 .. x_n)$ (data which parameterized by θ)

Definition: Given realisation of $x_1 \dots x_n$ of set of discrete random variables $X_1 \dots X_n$ with joint pmf $P_{\vec{\theta}}$, ~~θ~~ $\vec{\theta} \in R^n$ is a vector of parameters.

The likelihood function: $L_{x_1 .. x_n}(\vec{\theta}) = P_{\vec{\theta}}(x_1 .. x_n)$

For iid sequence; joint probability is the product

$$L_{x_1 .. x_n}(\vec{\theta}) = P_{\theta_1}(x_1) P_{\theta_2}(x_2) \dots P(x_n)$$

Log Likelihood (for joint probability)

$$\log P(X, Y) = \log P(X) + \log P(Y)$$

$$\ln L_{x_1 .. x_n}(\vec{\theta}) = \log P(x_1) + \log P(x_2) + \dots \log P(x_n)$$

$\log P(X, Y)$
$= \log P(X) +$
$\log P(Y)$

0	1			
5	10	2	3	4

argmax array$[x] = 1$ (max)

Maximum Likelihood Estimator

$$\vec{\theta}_{ML}(x_1 .. x_n) = \arg\max_{\theta} L_{x_1 .. x_n}(\theta)$$

$(P_\theta(x_1)$ → maximum in the distribution

(maximum value/point of the distribution is θ_{ML})

Maximum log likelihood $= \arg\max_{\theta} \log L_{x_1 .. x_n}(\theta)$

Log likelihood and Likelihood functions are same because log is monotonic function (~~chang~~ everything is same) (no changes are drastic)

* Maximum likelihood estimator is consistent (more & more data is available, and converge) deal

Maximum Likelihood estimator for Bernoulli

$$L_{x_1 .. x_n}(\theta) = P_\theta(x_1, x_2 .. x_n)$$

Parameter θ (success or failure)

$= \prod_i \left(\mathbb{I}_{x_i=1}\, \theta + \mathbb{I}_{x_i=0}(1-\theta) \right)$ (since iid sequence)

(indicator function)

$= \theta^{n_1}(1-\theta)^{n_0}$ $\quad n = n_1 + n_0$ (no. of 1s + no. of 0s) (Total no. of flips)

$\mathbb{I} \to$ Indicator func

log Likelihood, $\log L_{x_1 \dots x_n}(\theta) = n_1 \log\theta + n_0 \log(1-\theta)$

$$\theta_{ML} = \arg\max_\theta \left(n_1 \log\theta + n_0 \log(1-\theta) \right)$$

To find max:

$$\frac{d}{d\theta}\left(\log L_{x_1 \dots x_n}(\theta)\right) = \frac{n_1}{\theta} - \frac{n_0}{(1-\theta)}$$

$\frac{n_1}{\theta} - \frac{n_0}{1-\theta} = 0$

$n_1(1-\theta) - n_0\theta = 0$

$n_1 - n_1\theta - n_0\theta = 0$

$n_1 - \theta(n_1+n_0) = 0$

$n_1 = \theta(n_1+n_0)$

$\theta = \frac{n_1}{n_1+n_0}$

$$\left.\frac{d^2}{d\theta^2}\left(\log L_{x_1 \dots x_n}(\theta)\right) = \frac{-n_1}{\theta^2} - \frac{n_0}{(1-\theta)^2} < 0 \right\}$$ to check function concave

If $\left(\frac{n_1}{\theta} - \frac{n_0}{1-\theta}\right) = 0$, then $\theta_{ML} = \frac{n_1}{n_0+n_1}$. $\left(\frac{\text{success}}{\text{Total no. of flips}}\right)$

$\frac{n_1(1-\theta) - \theta(n_0)}{\theta(1-\theta)}$

$= \frac{n_1 - n_1\theta - \theta n_0}{\theta(1-\theta)}$

$= n_1 - \theta(n_1+n_0)$

$= 0$

$\frac{n_1}{n_1+n_0} = \theta$

[Check this θ_{ML} with Binomial distribution & Gaussian distribution]

Bayesian Statistics / Bayesian Estimation

Prior (P from knowledge)

Likelihood (P from data)

Posterior (actual P from Prior & Likelihood)

* Assumption of Prior, Likelihood

Posterior $\propto$ Likelihood $\times$ Prior

$$\underbrace{P(\theta \mid \text{Data})}_{\text{Posterior}} = \frac{\overbrace{P(\text{Data} \mid \theta)}^{\text{Likelihood}}\ \overbrace{P(\theta)}^{\text{Prior}}}{P(\text{Data})}$$

→ fun

$P(\text{Data}\mid\theta)$ is a conditional probability

$\downarrow$

Model

$P(\text{Data})$ (All possible Data)

(Evidence) (Normalisation constant) gives num

$P(\text{Data}\mid\theta) \to$ Likelihood : what is the probability that is this model (θ) can generate data.

$P(\theta)$: prior

$P(\text{data}) = \int P(\text{data}\mid\theta)\, P(\theta)\, d\theta$

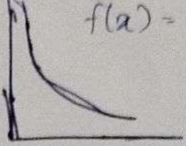
$P(\theta|Data)$: with given data, the probability of generating by model.

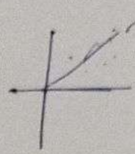
Both $P(\theta|Data)$ & $P(\theta)$ is distribution
If both distribution comes from same family, then its called as conjugate prior (Prior)

ex (β & Bernoulli ~~from~~ same family)

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

5/10 $f(x) = \lambda e^{-\lambda x}$ we fit by adjusting λ value.





* Model fitting for predicting.

Purpose: Model fitting.

* Start with a known model.
For that first one's finding likelihood, prior is the experience.
Posterior: with given evidence, updating the ~~data~~ model

Aposteriori Estimate (MAP) — Maximum Aposteriori

$$\theta_{MAP}(\vec{x}) = \arg\max_{\theta} P_{\theta|x}(\theta|\vec{x})$$

MLE
$$\arg\max_{\theta} P_{x|\theta}(x|\theta)$$

'MAP ≡ MLE, when the Prior is a constant.
(Not constant additional term in MAP)

$$\frac{d}{d\theta} P(\theta|x) = \frac{d}{d\theta}\left[\left(P(x|\theta) \cdot P(\theta)\right)\right]$$