Estimation.

↳ mean and σ are sufficient statistics of $N \varepsilon (\mu, \sigma^2)$

↳ Estimating these parameters from sample of the data is a demanding task.

Data distribution → Statistical model → Estimation.

↳ Can be parametrised → And then estimated

BAYES, MLE, OLS

uses Prior. Largest likelihood. ↳ Closest to the data

One describes the parameteric model for the data → The equation for these parameters → Estimator

→ We observe the real data → Calculate parameters.
using the equation + Observation → Estimate.

Estimator → Statistic → RV → Will have prob over range of outcomes.

↳ ~~Calculated~~ using present data (params)

B → only using x_i's, y_i's. → Design.

u → only using x_i's.

OLS → calculation for the estimator (Parameters) ↳ Ex. wts.

↳ Use this calculation to evaluate observation → compare L2 diff between observation and real data. → deviation → Reduce

$x_i \in X$ (n observations)

$E(x_i) = u$ then the deviance must ~~be~~ minimum for all x_i's ~~at~~ ↳ ~~about~~ u.

↳ $E(Y|X_i) = BX$ → Deviance must be low

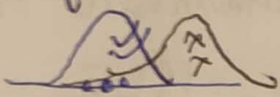
↳ $\sum_X (y_i - BX_i)^2$ → We can use all x's (Linearly linked data) $\sum (y_i - BX_i)^2$.

MLE → DATA

↳ Distribution as a function of parameters.

↳ Chk which set of params gives the best explanation of distribution

↳ We know if P↑ then distribution is more likely (→ suggested)



therefore we chk which param dist gives the highest likelihood

AIC → How much information is loss. If we use all the data points ln(Likelihood) ↑max. If we use params info will be lost

↳ $2K - 2\ln(\text{Likelihood})$

↳ params

All data points

$$\frac{\ln[\text{Best\_model}]}{\ln[\text{likeli}(\text{model})]}$$

Likeli

LR → $P(y|X)$ → Follows. $N(B_i X_i, \sigma^2)$ → error

↳ $0 + N(0, \sigma^2)$

↳ likelih = Product for each x_i.

BAYES → We have prior distribution for the parameters (→ RV's.)

↳ We update these priors wrt new data.

$P(\theta's \mid Data) \times P(Data) = P(Data, \theta's) = P(Data \mid \theta) \times P(\theta)$

↳ The space will change as we now have more data.

** Regularization in MLE → $P(y|X) = P(y_i|B_i X_i) \times P(B_i)$ (→ $N(B_i X_i, 0)$.)

Sampling properties of estimator.

↳ For given X → we try to predict $\bar{y}$ — y_{mean},

↳ We ~~get~~ have observations.

↳ for given X they act as real y_{mean}

Mean of Y given X_i — β dependent

⇒ , MSE of $\bar{y}$ = $E(y_{pred} - y)^2$ = $E(\bar{y} - y)^2$

↳ iid → 1 dataset

MSE of any $\bar{\theta} = E(\bar{\theta} - \theta)^2$.

↳ $\frac{1}{n}(\beta x_i - y_i)^2$.

↳ That how a normal E is calculated

⇒ Bias of $\bar{y}$ = $E(\bar{y}) - y_{true}$

↳ observations are considered as y_{true}

This Expectation is over datasets.

⇒ Var of $\bar{y}$ = $E[\bar{y} - E(\bar{y})]$

↳ Variance of mean of Y given X.

↳ This requires multiple sampling

→ In LR ever $(y_i|X)$ is expected to be true value

∴ bias = $E(\beta X_i - y_i)$ → How much expectation is far from true value.

→ Variance of estimate this βX_i is ~~still~~ computed over datasets.

MSE = Var + Bias^2

↳ measure of efficiency.

CENTRAL LIMIT THEOREM

↳ Markov inequality. ↱ Average

$$P(|X| \geq a) \leq \frac{E(|x|)}{a}$$ ↳ variable

Conceptual Proof :-

10 people → 95% (at least) are younger than the mea

Age [Is it possible] ??

1, 1, 1, 1, 1, 1, 1, 1, 1, 100 → Possible.

$$P(x \geq E|x|) \leq \frac{E(X)}{E(x)}$$

→ $P(x > \text{avg}) \leq 1$ ↳ Can be c to

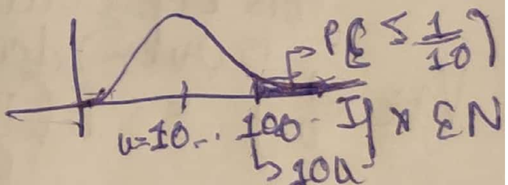
$\frac{9+n}{10} = n. \rightarrow 9n = 9, n = 1$

10 people :- more than 50% are older than 2× Avg age

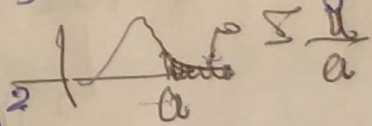
$$P(x \geq 2 \times E[x)]) \leq \frac{E(x)}{2 \times E(x)}$$ $\leq \frac{1}{2}$ → less than = 5

↳ $n \times E(x) \leq \frac{1}{n \times E(x)}$

↳ Not possible.



↳ conceptually they will drive the average u



⇒ Chebyshev :- $P(|X-u|^2 \geq a^2) \leq \frac{E(X-u)^2}{a^2}$

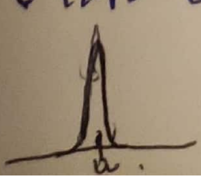
$$\leq \frac{Var(X)}{a^2}.$$

⇒ Law of Large Number :-

⇒ ~~Prob~~ For any $c > 0$

↳ ~~$P(|\bar{X}_n - u|) \geq 0$ as $n \to \infty$~~

↳ $P(|\bar{X}_n - u| > c) \to 0$ as $n \to \infty$



Proof. $P(|X_n - u| \geq C) = \frac{Var(X_n)}{C^2}$ (Chebyshev)

Popu Mean

$X_n = \frac{1}{n} \sum X_i$ → Samples ↳ $\frac{\sigma^2}{nC^2}$ → 0 If $n \to \infty$

$Var(X_n) = \frac{1}{n^2} \sigma^2 * n = \frac{\sigma^2}{n}$

↳ To find the distribution we multiply $(\bar{X}_n - u)$ with n → $n^{(1/2)} * (\bar{X}_n - u)$

↳ lim ∞

Tend to 0

If we normalise $\frac{(\bar{X}_n - u)}{\sigma^2} * n^{1/2} \to N(0,1)$

↳ This is CLT

Just $n \to \infty$

X → can be any RV.

(Proof on Lec 29 Harvard LLN)

$P\left(a < \left(\frac{\bar{X}_n - u}{\frac{\sigma^2}{n^{1/2}}}\right) < b\right) \to P(a < Z < b)$

↳ Z statistic

→ used in residual correlation

↳ $N(0, \sigma^2)$.

↳ Chi-Square RVs :-

↳ Squaring normal RV → χ^2_1

↳ squaring + adding n normal RV's → χ^2_n

Mean = n

Var = $2n$.

** $\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2}$ follows. χ^2_{n-1}

→ sample mean

population variance

↳ If n samples.

↳ If ~~x_i~~ $\frac{\bar{X}-u}{\sigma/\sqrt{n}} \rightarrow N(0, \frac{\sigma^2}{\sqrt{n}})$ then

$$\left(\frac{\bar{x}-u}{\sigma/\sqrt{n}}\right)^2 \rightarrow \chi_1^2$$

↳ $x_i \sim N(\bar{x}, \sigma^2)$ then $\frac{x_i-\bar{x}}{\sigma}$ (Z stat) $\rightarrow \left(\frac{x_i-\bar{x}}{\sigma}\right)^2 \rightarrow$ will follow χ_1^2

Now if we take n samples from pop $-N(u, \sigma^2)$

$\sum_{i=0}^{n-1}\left(\frac{x_i-u}{\sigma}\right)^2 \rightarrow chi_n^2$ but $\sum_{i=0}^{n-1}\frac{(x_i-\bar{x_i})}{\sigma^2}$ chi_{n-1}^2

↳ This is normal ↳ Not Normal.

$$\sum\frac{(\bar{x}-u)^2}{\sigma^2}$$

↳ $\left[\frac{(x_i-\bar{x})}{\sigma^2} + \frac{(\bar{x}-u)]^2}{\sigma^2}\right.$ ↳ $\frac{(\bar{x}-u)^2}{\sigma^2/n} \rightarrow Chi_1^2$

→ 0 about $\bar{x}$

$+2(x_i-\bar{x})(\bar{x}-u)$

$-2x_iu + x_i\bar{x}^2 + u\bar{x}$

$\frac{\sum(x_i-\bar{x})^2}{\sigma^2}$ ↳ $\left(\frac{\bar{x}-u}{\sigma/\sqrt{n}}\right)^2 \rightarrow$ CLT ↳ $N(u, \sigma/\sqrt{n})$

$E(x_i-\bar{x})^2 \rightarrow$ MSE $\frac{MSE}{\sigma^2 \text{ of population}} \rightarrow Chi_{n-1}^2$

or $E(\bar{y} - y_{true}) \rightarrow$ MSE

$E\left(\frac{\sum(x_i-\bar{x})^2}{\sigma^2}\right) = n-1$ as $E(chi_n^2) = n$.

Sampe Variance $\rightarrow \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{\sigma^2}{n-1} \times \frac{\sum(x_i-\bar{x})^2}{\sigma^2}$ (Chi_{n-1}^2)

$E(\text{Sample variance}) \rightarrow \frac{\sigma^2}{n-1} \times E(Chi_{n-1}^2) = \sigma^2$.

Student ~ t RVs. → Definition.

↳ $\frac{\text{Normal RV}}{(\text{Chi}^2 \text{ RV} / \text{D of freedom of Chi}^2)}$

Ex. X_1, X_2 → Random sample. of Normal dists $(0, \sigma^2)$

$$\frac{X_1}{\sqrt{X_2^2/1}}$$ → T distribution → 1 d of freedom

** $x_1, x_2 \ldots x_n$ → R samples $N-(u, \sigma^2)$.

↳ N samples.

1. $$\frac{\bar{X}-u}{\sigma/\sqrt{N}}$$ → CLT → $N(u, \frac{\sigma^2}{n})$.

∴ If

2. $$\frac{\sum^{N}(X_i - \bar{X})^2}{\sigma^2}$$ → Chi^2_{n-1}.

∴ $$\frac{[(\bar{X}-u)/\sigma/\sqrt{N}]}{\sqrt{\frac{\frac{\sum(X_i-\bar{X})^2}{\sigma^2}}{n-1}}}$$ → T dist d.of F_{n-1}.

↳ $$\frac{\bar{X}-u}{s/\sqrt{N}}$$ → $$\frac{(\bar{X}-u)/\sqrt{N}}{\text{Sample Variance}}$$

↳ Hypothesis test fundamental.

** ∴ If σ is unknown we can perform t-test with sample variance.

F-stat → We have a dist table for this

↳ $\frac{(\text{Ratio of } Chi^2_a \text{ RVs} / a)}{(\text{Ratio of } Chi^2_b \text{ RVs} / b)}$

↳ Thus we can hypothesize. if the N, D are Chi^2 and thus if the inside RVs (lets say error) are Normal

~~Total error =~~

Remember $\frac{\sum^{n} (x_i - \bar{x})^2}{\sigma^2}$ → Follow Chi^2_{n-1}

~~Chi^2~~

↳ $\frac{x_i - \bar{x}}{\sigma}$ is not Normal

~~→ We do~~

But $\sum_{i=1}^{N} \left(\frac{x_i - u}{\sigma}\right)^2$

↳ Chi^2_n and $\left(\frac{\bar{x} - u}{\sigma / \sqrt{n}}\right)^2$ → Chi^2_1

$\frac{\sum^{n} (x_i - \bar{x})^2}{\sigma^2}$ → $Chi^2_n = ( ) + Chi^2_1$

$( ) \to Chi^2_{n-1}$.

Similarly $\frac{\sum (x_i - (\beta x_i + \beta_0))^2}{\sigma^2}$ → Follows Chi^2_{n-2}

→ Will lead to Chi^2_1 each.

Numerator = MSE

Denominator = σ^2 of the entire population

↳ If we consider the entire dataset Chi^2_1

put back

→ We want it = 1 if $\beta_1 = 0$.

→ Population base

$\frac{\text{New explanability} / \sigma^2}{\text{Old explanability} / \sigma^2} = \frac{Chi^2_{n-2}}{Chi^2_1}$ → Fstat