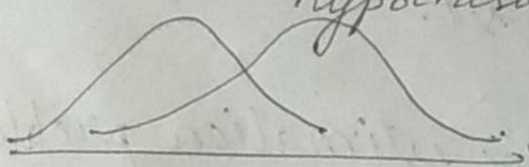
Date

Statistics 2/4/18

Statistical inference (estimation ~~of~~ & testing of hypothesis)

$N(\mu, \sigma^2)$



$\{X_1 = x_1, \dots, X_n = x_n\}$

e.f. $F_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1\{x_i \leq t\}$ - frequency analysis.

let $(X_1, X_2, \dots, X_n)$ be random samples from $f_\theta()$ or $F_\theta()$ where f is the p.d.f and F is the CDF of X. Here random sample $X_1, X_2 \dots X_n$ are i.i.d $f_\theta()$ i.e

joint pdf of $(X_1, X_2, \dots, X_n)$ at $(x_1, x_2 \dots x_n)$ is

$f(x_1, x_2 \dots x_n) = \prod_{i=1}^{n} f_\theta(x_i)$

The family of distribution is defined as

$\mathcal{F} = \{ f_\theta() \mid f_\theta() - \text{pdf of } X \text{ \& } \theta \in \mathcal{X} \}$

$\mathcal{X} = \{ F_\theta() \mid F_\theta() \text{ is cdf of } X \text{ \& } \theta \in \mathcal{X} \}$

e.g.

$= \left\{ \binom{n}{x} p^x (1-p)^{n-x} \mid n \in \mathbb{N}, p \in [0,1] \right\}$

parameter space $\mathcal{X} = \mathbb{N} \times [0,1]$

$N = \left\{ \frac{e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\sigma} \mid (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+ = \mathcal{X} \right\}$

· In parametric estimation problem we want to identify the parameter values of the distribution from given data set.

· In nonparametric estimation problem we want to identify pdf/cdf directly from the data without any parametric set up.

Point estimation in parametric set-up

(1) Dfn of estimation
(2) good prop. of estimation
(3) methods of estimation
(method of moment , MLE)

Statistic - is a function of data but it is free from any unknown parameter.

let $x_1, \dots x_n$ be random sample from $N(\mu, \sigma^2)$
example
(1) μ known : $\Sigma(x_i - \mu)^2$ ✓
(2) μ unknown : $\Sigma(x_i - \mu)^2$ ×
(3) $(x_i - \bar{x})^2$ is a statistic ✓
(4) μ σ^2 unknown, $\Sigma \frac{(x_i - \bar{x})^2}{\sigma^2}$ ×

family of distribut^n = degree of freedom

$X \sim N(0,1)$

$X^2 \sim \chi^2_1 = G(1/2, 1/2)$

$$\sum_{i=1}^{n} (x_i - \mu)^2 \sim \sigma^2 \chi^2_n = \sigma^2 G\left(\frac{n}{2}, \frac{1}{2}\right)$$

$$\sum (x_i - \bar{x})^2 \sim \sigma^2 \chi^2_{n-1} = \sigma^2 G\left(\frac{n-1}{2}, \frac{1}{2}\right).$$

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \sim \chi^2_{n-1} = G\left(\frac{n-1}{2}, \frac{1}{2}\right)$$

Let $x_1, \ldots x_n$ iid $f_\theta \in \mathcal{F} \{ f_\theta \mid \underset{\sim}{\theta} \in \alpha \}$

we may be interested to estimate a function of $\underset{\sim}{\theta}$ say $g(\theta)$ from data

$g(\mu, \sigma^2) = \mu$ for $N(\mu, \sigma^2)$

$g(\mu \ \sigma^2) = \sigma^2$ for

coeff of variation $g(\mu, \sigma^2) = \frac{\sigma^2}{\mu}$ if $\mu \neq 0$

$$\left.\begin{aligned} g(\alpha, \lambda) &= \frac{\alpha}{\lambda} \quad G(\alpha, \lambda) \\ g(\alpha, \lambda) &= \frac{\alpha}{\lambda^2} \quad \text{"} \\ g(\alpha, \lambda) &= \lambda \quad \text{"} \end{aligned}\right\}$$

estimator: A statistic $T(\underset{\sim}{X})$ when used to estimate a parametric function $g(\underset{\sim}{\theta})$ is called an estimator of $g(\underset{\sim}{\theta})$.

As $T(\underset{\sim}{X})$ is a function of r.v. $\underset{\sim}{X} = (x_1, x_2, \ldots x_n)$ it is also a r.v.

For a given data set, $(X_1 = x_1, X_2 = x_2 \dots, X_n = x_n)$ the value of $T(x)$ is known as an estimate of $g(\theta)$

$g(\theta) \hat{=} T(X)$ or $\widehat{g(\theta)} = T(x) \Rightarrow g(\theta)$ is estimate of $T(x)$

e.g. $\mu \hat{=} \bar{x}$ or $\hat{\mu} = \bar{x}$

$\sigma^2 \hat{=} \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$

E.g. let $X_1, X_2 \dots X_n$ be iid random variables with $E(X_i) = \mu$ $Var(X_i) = \sigma^2$

μ can be estimated by ① $T_1(\underset{\sim}{X}) = X_5$

② $T_2 = (5X_1 + 3X_2)/8$

③ $T_3 = \sum_{i=1}^{n} a_i X_i$ s.t. $\sum a_i = 1$

④ $T_n = \frac{1}{n} \sum \cancel{x} X_i = \bar{X}$

e.g. σ^2 can be estimated by

$S_1^2(\underset{\sim}{X}) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{x})^2 \leftrightarrow$ data scattered around mean)

$= \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - X_j)^2$

$S_2(\underset{\sim}{X}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

HW: $E(S_2^2(\underset{\sim}{X})) = \sigma^2$

in particular if

$\sum (X_i - \bar{x})^2 \sim \sigma^2 \chi^2_{n-1}$

then $E\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right] = \sigma^2 (n-1)$

$E\left(\frac{1}{n-1} \sum (X_i - \bar{X})^2\right) = \sigma^2$

Properties of an estimator {Ch 4 Hoel Port Stone

Unbiased estimator: let $T(\underset{\sim}{X})$ be an estimator of $g(\theta)$ which satisfies

$$E[T(X)] = g(\underset{\sim}{\theta}) \quad \forall \ \underset{\sim}{\theta} \in H \text{ then}$$

$T(X)$ is said to be an unbiased estimator of $g(\theta)$

- let $X_1 \sim N(0,1)$, $T(\underset{\sim}{X}) = 3X_1$
  is $T(\underset{\sim}{X})$ an unbiased estimator of mean?

Ans. NO bcoz $N(\mu, \sigma^2)$ doesnt satisfy these all parameter condit^n ($\forall \theta \in H$)

Bias of an estimator is defined as

$$B_{g(\theta)}(T(x)) = E(T(x) - g(\theta)) = E[T(x)] - g(\theta)$$

- e.g. T_1, T_2, T_3, T_4 given before
- e.g. show that $S_2^2(\underset{\sim}{X})$ is an unbiased estimator of σ^2 but not $S_1^2(X)$.

Asymptotically unbiased estimator
- if the bias of an estimator $T(\underset{\sim}{X})$ goes to zero when sample size $n \to \infty$,

ex: $S_1^2(X)$ is an asymptotically unbiased estimator of σ^2 in ex ②

Mean squared Error (MSE)

if $g(\theta)$ is an parametric fn & its estimated by $T(X)$ then MSE of $T(X)$ is def. as.

$MSE(T(x)) = E[T(X) - g(\theta)]^2$

$MSE[T(X)] = Var[T(X)] + [Bias(TX)]^2$ //

Note: if $T(X)$ is an unbiased estimator of $g(\theta)$

then $bias(T(X)) = 0$

$\Rightarrow MSE[T(X)] = Var(T(X))$

But in general, $MSE(T(X)) \geq Var(T(X))$

$E(T(X) - g(\theta))^2$

$= E[T(X) - E(T(X)) + E(T(X)) - g(\theta)]^2$

$= E[T(X) - E(T(X))]^2 + E[T(X) - g(\theta)]^2$

$= Var[T(X)] + Bias(T(X))^2$

<u>Consistent</u> <u>Estimator</u>

An estimator $T_n(\underset{\sim}{X}) = T(X_1 X_2 \ldots X_n)$ of $g(\theta)$ is said to be a consistent estimator if

$\lim_{n\to\infty} P(|T_n(X) - g(\theta)| > \epsilon) = 0$

$\lim_{n\to\infty} P(|T_n(X) - g(\theta)| < \epsilon) = 1$

Ex 4: if MSE of $T_n(X)$ for $g(\theta)$ is going to zero as $n \to \infty$ then $T_n(X)$ is consistent estimator of $g(\theta)$

a) $T_n = \sum_{i=1}^{n} a_i X_i \qquad \sum_{i=1}^{n} a_i = \mu$

for what values of a_i, T_3 has min. variance?

$$\mathrm{var}(T_n) = \sigma^2 \sum_{i=1}^{n} a_i^2 = \frac{\sigma^2}{n}\left(\sum_{i=1}^{n} a_i^2\right) n$$

$$= \frac{\sigma^2}{n}\left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} 1\right)$$

$$\geq \frac{\sigma^2}{n}\left(\sum_{i=1}^{n} a_i \cdot 1\right)^2$$

[since
$\sum a_i^2 \sum b_i^2 \geq \left(\sum a_i b_i\right)^2$.
'=' holds when
$a_i \propto b_i$ i.e. $a_i = k b_i$]

$$\geq \frac{\sigma^2}{n} \quad (\because \sum a_i = 1)$$

'=' holds if $a_i \propto 1$, $a_i = k$

$$\sum a_i = nk$$

$$1 = nk \Rightarrow k = \frac{1}{n}$$

$\therefore$ Min. unbiased estimator of μ is

$$\frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

3/4/18

Estimation:

→ Estimator / statistic

→ Properties of estimator

→ Unbiased estimator

- Asymptotically unbiased estimator

- Consistency of an estimator

- MSE (mean square error) — if MSE = 0, estimator is consistent

Methods of estimation

1. Method of moment estimator

observation $(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)$ from a population with say pdf $f_\theta(\ )$

we are interested to estimate $(\theta_1, \theta_2 \ldots \theta_k)$

→ compute

(i) Theoretical moments from pdf

(ii) empirical " " data

(iii) Construct k equations if you have k unknown parameters

(iv) Solve for k parameter values.

$E(X) = \mu_1' =$ 1st order raw moments

$E(X - \mu_1) = 0$

$E(X^2) = \mu_2'$ or $var(X) = E(X-\mu)^2 = \mu_2$

$\mu_1' = \frac{1}{n} \sum_{i=1}^{n} x_i$

$\mu_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$\mu_3 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$ or $\mu_3' = \frac{1}{n} \sum x_i^3$

k many equations for k many parameter values

Gamma(α, λ)

$\mu_1' = \frac{\alpha}{\lambda} = \bar{x}$ — (1)

$\mu_2 = \frac{\alpha}{\lambda^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$ solve for α and λ

Ex: $X_1, \ldots X_n \overset{iid}{\sim} (\mu, \sigma^2)$

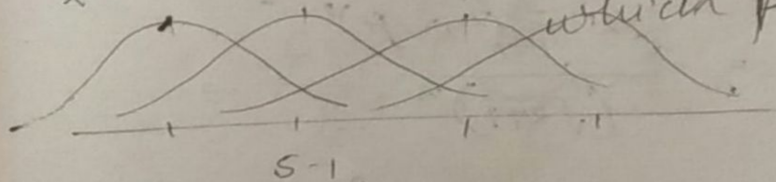
we cannot use methods of moment estimation for those cases where theoretical moments does not exist

eg. $E(|X|)$ $E(X^2)$ as not finite

<u>Maximum likelihood method</u>

$X \sim N(\mu, 1)$ $\mu \in \mathbb{R}$ identify possible value of μ for

$X = x = 5.1$ which $P(X = x = 5.1)$ is maxm.



$\frac{e^{-1/2(x-\mu)^2}}{\sqrt{2\pi}} \sim 1$

$X_1 = x_1, X_2 = x_2 \ldots X_n = x_n$

$f(\underset{\sim}{x}) = \prod_{i=1}^{n} \frac{e^{-1/2(x_i-\mu)^2}}{\sqrt{2\pi}}$

$\hat{\mu}$ = maxima of $f(\mu, x)$

↓

argmax

Let $x_1, x_2 \ldots x_n$ be iid $f_\theta()$ then the joint pdf of $x_1 \ldots x_n$ is

$$f(\underset{\sim}{x}) = \prod_{i=1}^{n} f_{\underset{\sim}{\theta}}(x_i)$$

the likelihood of $x_1 \ldots x_n$ is defined as

$$l(\underset{\sim}{\theta}) = \prod_{i=1}^{n} f(\theta, x_i) \text{ for known } x_i \text{ values}$$

Maximum likelihood estimators of

$$\underset{\sim}{\theta} = \underset{\underset{\sim}{\theta} \in H}{\arg\max}\, l(\underset{\sim}{\theta}) \qquad \theta \in H$$

$$= \underset{\underset{\sim}{\theta} \in H}{\arg\max} \left(\log(l(\theta))\right)$$

ex: $x_1 \ldots x_n$ iid $N(\mu, 1)$ find the MLE of μ.

$$l(\mu) = \prod_{i=1}^{n} \frac{e^{-1/2 (x_i - \mu)^2}}{\sqrt{2\pi}} = \frac{e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2}}{(\sqrt{2\pi})^n}$$

$$L(\mu) = \log l(\mu) = c - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$\boxed{\hat{\mu}_{MLE} = \bar{x}}$$

eg. $x_1, \ldots x_n \sim N(\mu, \sigma^2)$

find MLE of μ and σ^2

eg. $x_1 \ldots x_n \sim \Gamma(\alpha, \lambda)$

find MLE of α, λ

Properties of MLE

i. MLE need not be an unbiased estimator

ii. " is always a consistent "

iii. under some regularity conditions MLE follows normal distribution upto some location and s.— parameter.

— (i) sample space should be free from parameter

(ii) 3rd order moment must be finite

(iii) MLE need not be unique.

1. $X_1 \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

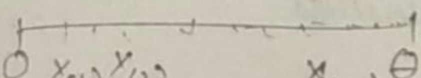
$$\sigma^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$= \frac{(n-1)}{n}\left[\frac{1}{(n+1)}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

let $X_1 \dots X_n \overset{iid}{\sim} U(0, \theta)$ $\quad \frac{\theta}{2} = \bar{x}$

$$\hat{\theta}_{MME} = 2\bar{X}.$$

method of moment estimator

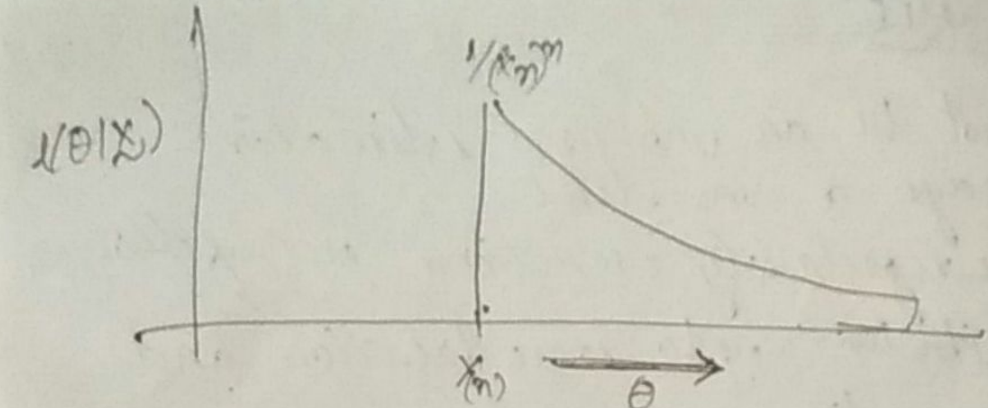


$$0 \le X_{(1)} \le X_{(2)} \dots X_{(n)} \le \theta$$

joint pdf of $X_1 \dots X_n$ is

$$f_{\theta}(\underset{\sim}{x}) = \begin{cases} \frac{1}{\theta^n} & 0 < x_i < \theta \ \forall i \\ 0 & o/w \end{cases}$$

$$l(\theta | \underset{\sim}{x}) = \begin{cases} \frac{1}{\theta^n} & \text{if } \theta \geq X_{(n)} \\ 0 & o/w \end{cases}$$



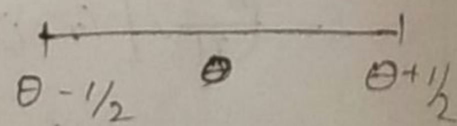
$$\hat{\theta}_{MLE} = X_{(n)}$$

→ $X_{(n)}$ is not an unbiased estimator but $X_{(n)}$ is an
→ asymptotically " " & $X_{(n)}$ is a
→ consistent estimator

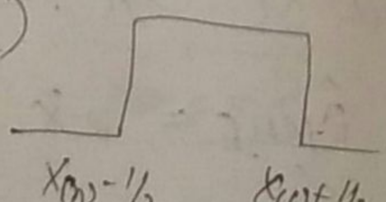
if $Y = X_{(n)}$

$$f(y) = \begin{cases} \dfrac{n y^{n-1}}{\theta^n} & \text{} \; 0 < y < \theta \\ 0 & \text{o/w} \end{cases}$$

Ex: let $X_1 \ldots X_n \overset{iid}{\sim} (\theta - 1/2, \theta + 1/2)$



$\hat{\theta}_{MLE}$ attains maximum on an interval $(X_{(n)} - 1/2, X_{(1)} + 1/2)$



joint densities

check condn: $X_{(n)} < \theta + 1/2$

$Y_{(n)} > \theta - 1/2$

No. 9/4/18
Date

Statistics

Descriptive analysis (not predictive).
try to find a distributⁿ that best fits the given data.

Random Sample (r.s.): Collection of independent & identically distributed iid r.v. $X_1 - - X_n$.

Realizations of $X_1 X_2 - - X_n$ as $x_1 x_2 - - x_n$ is also called random sample.

Data Reduction:

Statistic (no's) : A functⁿ of $X_1 X_2 - - X_n$

Eg. of statistic : $\bar{X} = \frac{X_1 + - - X_n}{n}$ - Sample mean

$s^2 = \frac{1}{n-1} \sum_1^n (X_i - \bar{X})^2$: sample variance.

$R_s = \max_{i=1,2-n}\{X_i\} - \min_{i=1,2-n}\{X_i\}$: sample range

Let 0.3, 2.1, -1.3, -0.72, 3.3, -0.76, 1.37, -1.9, -0.69, 0.01 → sample
mean & Variance = ?
0.171 ; 2.4023 / 2.64

0.016641

21.6

Let $X_1 X_2 - X_n$ → r.s. with $x_1 x_2 - x_n$: realization.
Let $\bar{X}$ denote sample mean. Then, from data $\bar{x} = \sum_1^n \frac{x_i}{n}$ is the realizatⁿ of $\bar{X}$.

→ Can I say $\bar{x}$ is the mean of X_is ??

$X_i \sim$ Bernoulli($p = 1/2$). Let $X_1 X_2 - X_{10}$ be indep. Bernoulli ($p = 1/2$) r.v. $X_1 - - X_{10}$: r.s.
Let 1, 1, 1, 0, 0, 1, 0, 0, 0, 0 be realizatⁿ of $X_1 - X_{10}$
mean = 0.4. But mean of Bernoulli = 0.5

No.
Date.

→ when we speak of r.v. we should never speak of realizat^n but of density. $\bar{X}$ itself is a r.v. & obvio it's realizat^n will be diff^r for diff^r people (but density same)

Let $X_1 X_2 \ldots X_n$ be a random sample. Let $\mu = E(X_i)$. Define $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
↓ sample mean (also r.v.)

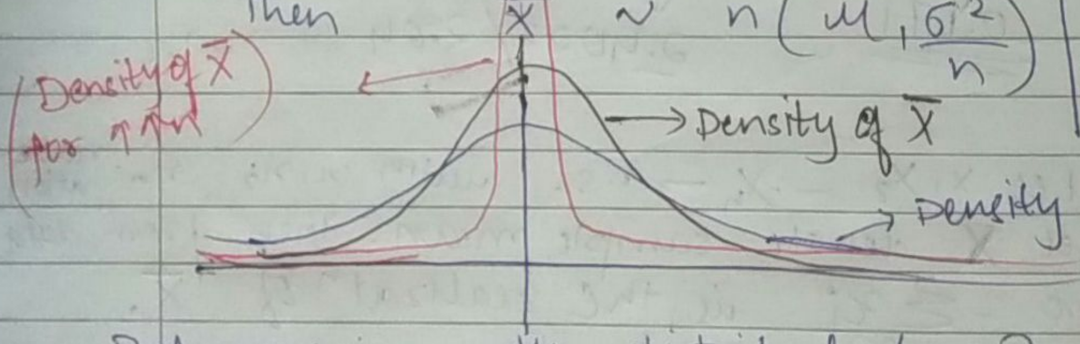
Then, notice that
$$E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n\mu = \mu.$$
$$\boxed{\therefore E(\bar{X}) = \mu}$$

- Estimator → r.v.
- Unbiased estimator. — the realized value of estimator

Q - How much 'confidence' we have in this estimator?

→ Let $X_1 X_2 \ldots X_n$ be a r.S. from normal distribut^n with mean μ & variance σ^2

Then $\bar{X} \sim n\left(\mu, \frac{\sigma^2}{n}\right)$ | $X \sim n(\mu, \sigma^2)$
Var of X
$X > \bar{X}_2$
$\sigma^2 > \frac{\sigma^2}{n}$



Prob mass is mostly distributed as
$$P\left(\mu - \frac{3\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + \frac{3\sigma}{\sqrt{n}}\right) = 0.9995$$
(n ↑↑ : $\frac{3\sigma}{\sqrt{n}} \approx 0$ i.e. ↓↓ mass concentrated in neighbourhood & huge spike at centre)
↓
i.e. variability in $\bar{X} \approx 0$ (↓↓)

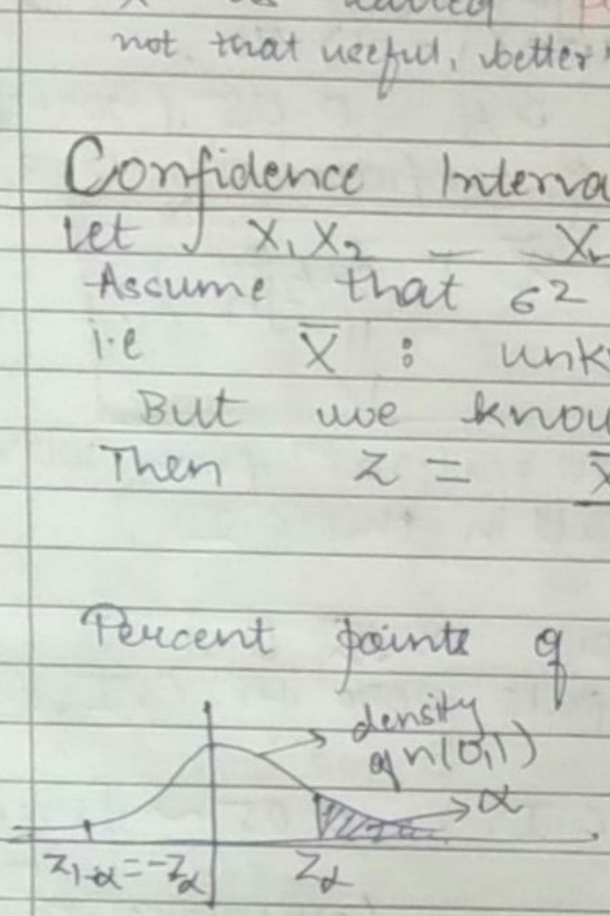
μ : unbiased estimator. → we don't consider this to find confidence
Variance & Bias decides confidence interval & as $n \uparrow \uparrow$

- In prev. eg. $\overline{X}$: estimator of μ. & $\bar{x} = 0.17$: estimate of μ.
  $\overline{X}$ is called point estimator.
  not that useful, better to estimate in an interval (conf interval).

Confidence Interval on μ.

Let $X_1 X_2 \ldots X_n$ be a r.s. from normal(μ, σ^2)
Assume that σ^2 is known (this is hypothetical for now)
i.e $\overline{X}$: unknown σ^2 : known.
But we know $\overline{X} \sim n(\mu, \sigma^2/n)$
Then $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim n(0,1)$.

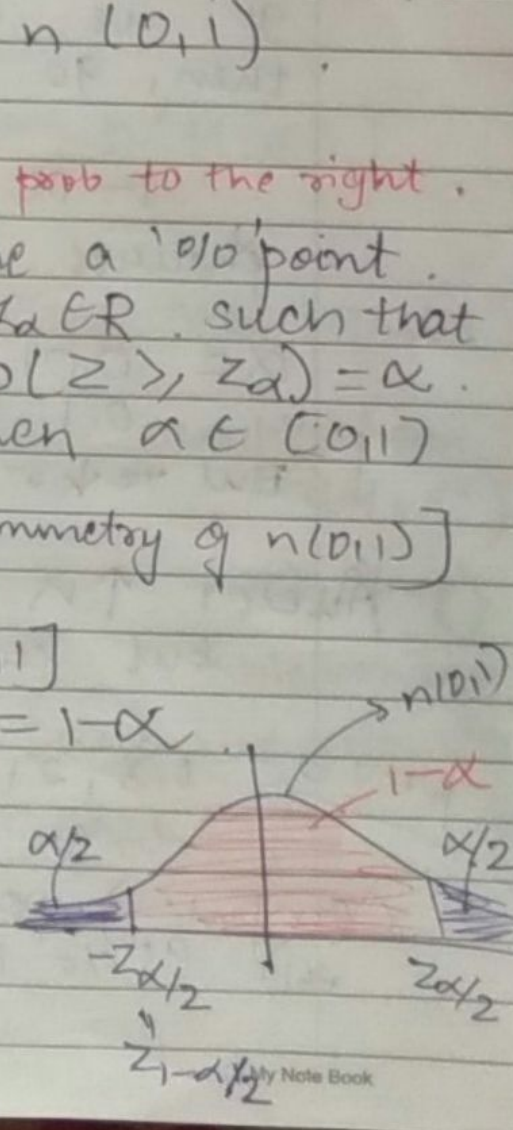
Percent points of $n(0,1)$ ⟶ prob to the right.



Define a % point. $z_\alpha \in \mathbb{R}$ such that $\text{Prob}(Z \geq z_\alpha) = \alpha$. For given $\alpha \in [0,1]$

$z_{1-\alpha} = -z_\alpha$ [b'cos of symmetry of $n(0,1)$]

Then, for some no. $\alpha \in [0,1]$
$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$.



No.
Date

$$\Rightarrow P\left(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right)$$

$$\Rightarrow P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

Confidence interval for μ (σ^2 known) $100(1-\alpha)\%$.

Choose $\alpha = 0.1 \Rightarrow \alpha/2 = 0.05$ (90% of $100(1-\alpha)\%$ of confidence interval

$$\boxed{\left(\underbrace{\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{= L},\ \underbrace{\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{= U}\right)}$$

If we consider 100 realizatn from 100 rv then, 90 μ will fall in this C.I.

$1-\alpha \approx 1$ or $\alpha \approx 0$.
ensures that μ falls 100% in C.I.

$\alpha =$ 0.01 ~ 99% C.I. 0.05 ~ 95% CI
0.1 ~ 90% -

But $\downarrow\sigma$ to zero of no use (otherwise $\mu \to \mathbb{R}$ interval)

Alter: $\uparrow n$ so that $\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \longrightarrow \bar{x}$.
but can't $\uparrow n$ until one limit.

prev. eg - 0.3, 2.1, −1.3, −0.72, 3.3, −0.76, 1.37, −1.9, −0.69, 0.01. $\mu = 0.171$ $\sigma^2 = 2.64$.

Compute i) point estimate of μ = 0.171.
ii) 95% C.I on μ.

No.
Date.

$\rightarrow \left(\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} , \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 100(1-\alpha)\%$

$\alpha = 0.05 \longrightarrow 95\%$ CI $\quad$ 95 %.

$\alpha/2 = 0.025$. $\quad Z_{\alpha/2} = Z_{0.025} = 1.96$

$0.025 \to$ prob $\quad 1.96 \to$ z factor

Given: $\bar{x} = 0.17$.

$\sigma = 1$, $n = 10$.

$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 0.17 - 1.96 \left(\frac{1}{\sqrt{10}} \right) = -0.449$

$\bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \underline{\underline{0.79}}$.

95% of CI on $\mu = (-0.448, 0.79)$.

Confidence Interval:

~~Let $X_1 X_2 - - X_n$ be r.s. from $n(\mu, \sigma^2)$~~

case i)

Estimation of Variance in normal populat^n.

- Let $X_1 X_2 - - X_n$ be a random sample from $n(\mu, \sigma^2)$

$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$: sample mean
(unbiased estimator of μ)

$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$: sample variance
(unbiased estimator of σ^2)

$E(s^2) = \sigma^2$

We already know $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim n(0,1)$.

No.
Date.

Aim : To know distribution of s^2 or some f^n of s^2

- I can construct CI on μ

$\bar{x}$ = point estimator of μ.

Result - $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$

$$\frac{(n-1)s^2}{\sigma^2} = \frac{(n-1)\cdot\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}$$

$$= \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\sigma^2}$$

Obs : $\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2} \sim \chi^2_n$

$$\sum_{i=1}^{n}\frac{(x_i-\bar{x}+\bar{x}-\mu)^2}{\sigma^2} = \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}$$

$$= \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\sigma^2} + \sum_{i=1}^{n}\frac{(\bar{x}-\mu)^2}{\sigma^2} + \frac{2\sum(x_i-\bar{x})(\bar{x}-\mu)}{\sigma^2}$$

$0 =$

$$= \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2} = \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\sigma^2} + \frac{(\bar{x}-\mu)^2}{\sigma^2/n}$$

$$\underbrace{\sum_{1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}}_{\chi_n} = \underbrace{\frac{(n-1)s^2}{\sigma^2}}_{\chi^2_{n-1}} + \underbrace{\frac{(\bar{x}-\mu)^2}{\sigma^2/n}}_{\chi_2}$$

s^2 & $\bar{x}$ are indep. r.v.

$\bar{X} \sim N(\mu, \sigma^2/n)$ $\quad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$\frac{(\bar{X}-\mu)^2}{\sigma^2/n} \sim \chi^2_1.$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{\underbrace{n-1}_{\text{d.of.}}}$$ (Additive prop. of χ^2

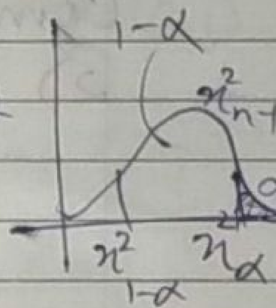
Estimatⁿ of σ^2 for normal populatⁿ.

- We already know: s^2 is an unbiased estimator of σ^2.

We want to construct a confidence interval for σ^2

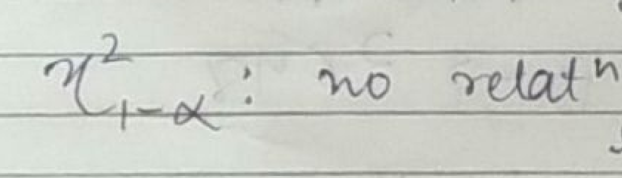
Note: $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$.

Percent point: for $0<\alpha<1$
$\chi^2_\alpha \in \mathbb{R}$ is such that
$P(\chi^2_{n-1} \geq \chi^2_{\alpha,n-1}) = \alpha$.



$\chi^2_{1-\alpha}$: no relatⁿ b/w χ^2_α due to lack of symmetry

For a given level $\alpha \in (0,1)$,
$100(1-\alpha)\%$ C.I. for σ^2.



$\chi^2_{1-\frac{\alpha}{2}, n-1}$

$$P\left(\chi^2_{1-\frac{\alpha}{2},n-1} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2},n-1}\right) = 1-\alpha.$$

No.
Date.

$$\Rightarrow P\left[\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right] = 1-\alpha$$

$\Rightarrow$ $100(1-\alpha)$ % CI on σ^2 is

$$\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)$$

Example : let 0.3, 2.1, −1.3, −0.72, 3.3, −0.76, 1.37, −1.9, −0.69 & 0.01 be 9 random sample from normal populatⁿ with mean μ & variance σ^2

Compute 1) point estimate of $\sigma^2 = s^2$
2) 90 % CI for σ^2

$$\sigma^2 = \frac{\sum (X-\bar{x}^2)}{n}$$

$$\left(\frac{9S^2}{\cancel{16.92}}\right.$$

1) → 2.648

2) (1.409, 7.859)