September 5, 2018 3:09 PM

Measure ? ← why?

Eg: $P[2^{nd} \text{ year}] = 10\% = \dfrac{10}{|S|}$ ← # of second year student ← all people sitting in the room

you are measuring something.

<u>Axioms</u>

1. $P[A]$ has to be non-negative
2. $P[\emptyset] = 0$
3. $P[S] = 1$

<u>Ex</u> $P[S] = P[\text{uni students}] = 1$

4. $P[\text{Additive}] = P[A_1 \cup A_2 \cup \ldots] = P[A_1] + P[A_2] \ldots$

Eg $\left.\begin{array}{l} A_1 = \text{first year} \\ A_2 = \text{second year} \end{array}\right\}$ disjoint because nothing in common

$$\left.\begin{array}{l} 3 = HHH \\ 2 = HHT \\ 2 = HTH \\ 1 = HTT \\ 2 = THH \\ 1 = THT \\ 1 = TTH \\ 0 = TTT \end{array}\right\} \text{sample space}$$

$P[H] = \frac{1}{2}$

$P[T] = \frac{1}{2}$

$P[3H] = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

Random variable

Maps sample space

$S \rightarrow \mathbb{R}$, $X = \#$ of heads

X:	0	1	2	3
$P[X]$:	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

or add probabilities

table called pmf or probability mass function.

To find $E(X) = \sum x \, p[x]$ → discrete

$\int x \, p[x] \, dx$ → continuous

To find $E[x^2] = \sum x^2 p[x]$ → discrete

$E(x^2] = \int_{-\infty}^{\infty} x^2 p[x]$ → continuous.

Bernoulli:

$$p_X(x) = \begin{cases} \theta \\ 1 - \theta \end{cases}$$

Ex/ $x = \#$ of H $\left\langle \begin{matrix} 1 \\ 0 \end{matrix} \right.$

$P[H] = \theta$

$$P[T] = 1-\theta$$

poisson mean and variance is exactly the same $\Rightarrow \lambda$

Continuous

Uniform, Exp, gamma, normal, Beta.

Next class review function, mean, cdf, pdf etc....

Marginal prob example

	1	2	3	4	5	6	
H	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$= \frac{1}{2}$
T							$= \frac{1}{2}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	

$A = H$
$B = 1$

is $A \perp B$? Yes

Proof $P(\text{Joint } (A \cap B)) = \frac{1}{12}$

$$P(A] = \frac{1}{2}$$

$$P(B) = \frac{1}{6}$$

$P(A \cap B) = P(A)\,P(B)$ $\cdot' \cdot$ $\frac{1}{\ }$

$$P[1 \mid H] = \frac{P(1 \cap H)}{P(H)} = \frac{\frac{1}{12}}{\frac{1}{2}} = \frac{1}{6}$$

X : 0 1 2 3

$P[X]$: $\frac{1}{8}$ $\frac{3}{8}$ $\frac{3}{8}$

Treatment Heart Transplant

Control group

P1 - survied 49 days after died
P28 - surived 118 days - but he is alive

P - Indicator

P1 - wait time for trasplant is 0 days ⇒ Survived 15 days - dead

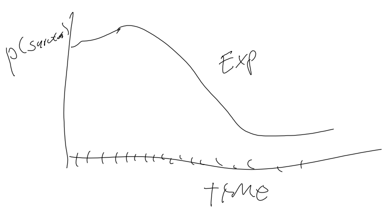
P52 - wait time 5 days → Survived 43 days - alive

30 ppl → mean(t)
↙
Treatment

Control
↑
52 ppl → mean(c)

compare the two values.

Assume 30 is coming from a distribution
eg.



if mean of treatment group is bigger > mean of control group

Interpertation about the population:

⇒ if you go through the treatment you are expected to live longer.

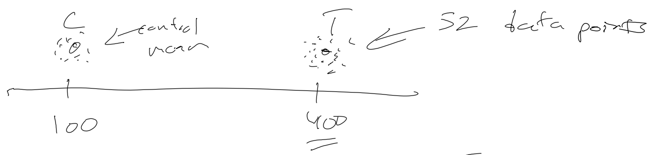
mean(T) > mean(c) are samples, interpertation we want to make is about the entire population.

assume

① $X \sim Exp(\lambda) \rightarrow$
② estimate of λ from these #s (parameter)

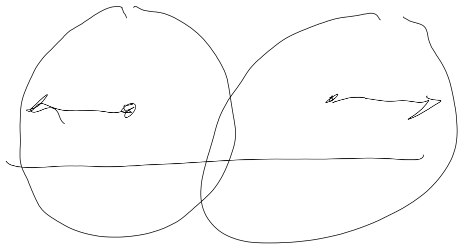
In real life you never know the distribution

Recall
$X \sim exp(\lambda)$
$f_X(x) = \lambda e^{-\lambda x}$



Treatments better

mean is the center of the distribution



radius also matters.

Data - Circle

mean - center of data

Standard deviation - radius

look if they are overlapping or far apart.

5.1.1 $mean(c) = 93.2$
$= \frac{sum(30 \text{ numbers})}{30}$

$mean(t) = 356.2$

you can't just compare the mean, look at SD and then mean.

Ex 5.1.8

$X \sim Exp(\lambda)$

$x_1, x_2, \ldots x_n$

$E(x) = \int_0^\infty x f(x)\,dx$

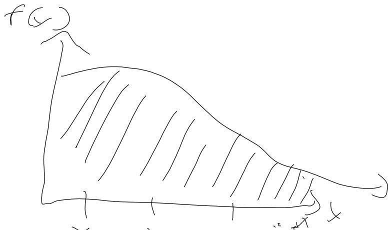
$= \int_0^\infty x \lambda e^{-\lambda x}\,dx$

$= \lambda \int_0^\infty x e^{-\lambda x}\,dx$

difference between X and x_i is

Eg $\lambda = 2$

$f_X(x) = 2e^{-2x}$



$$= \lambda \int_0^\infty x e^{-\lambda x}\, dx$$

$$= \frac{\lambda}{\lambda^2} \underbrace{\int_0^\infty x^{2-1} e^{-\lambda x} \cdot \frac{\lambda^2}{\Gamma(2)}}_{\text{gamma density} = 1}$$ ← divide by λ^2 ← this is 1

$$= \frac{1}{\lambda}(1) = \frac{1}{\lambda}$$

weak law of large #'s says

$x_1, x_2, \ldots, x_n$; $\frac{1}{\lambda}$

$$\Rightarrow \frac{x_1 + x_2 + \ldots + x_n}{n} \xrightarrow{\text{converge}} \frac{1}{\lambda} = E(x)$$

$$= \bar{x} \to \frac{1}{\lambda}$$ ← estimate of mean

$$= \frac{1}{\bar{x}} \to \lambda$$

estimate of lambda

Can take any value X or Y. is the random variable that can take any value

x_1 is sample value

eg $x_1 = 100$ days

$x_2 = 200$ days

Recall: Gamma

$$y \sim G(\alpha, B)$$

$$f(y) = \frac{B^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-By}, \quad y \geq 0$$

problems

$\bar{x} = 0.05$ $\hat{\lambda} = 20$ ← estimate

$\bar{x} = 0.06$ $\hat{\lambda} = 16.66$

if λ is really big eg 200 to contain it you need a large sample size

R-vector: c(1,3,5) => | 1 | 3 | 5 |

c(1,3,5) + 2 = | 3 | 5 | 7 |

variable: x = c(5,7,9,10)

printing: >x

[1] 5 7 9 10

log(x) => log [5], log [7]

$\log(x) \Rightarrow \begin{bmatrix} \log 5 \\ \log 7 \\ \log 9 \\ \log 10 \end{bmatrix}$

1.60943 1.94591...

$\log10(x)$

$\sin(x)$

$\cos(x)$

$x^\wedge 2$

$\left.\begin{array}{l} \log10(x) \\ \sin(x) \\ \cos(x) \\ x^\wedge 2 \end{array}\right\}$ does it for all elements

$y =$

5.1.1 / $x \sim N(0,1)$

$y = x^4 + 2x^3 - 3$

$P(y \in (1,2)) \leftarrow$ prob y between $(1,2)$

$x^2 \to$ Chi-sq

$-\infty$ ∞

Ex 5.2.1

$x =$ life lenght of a machine

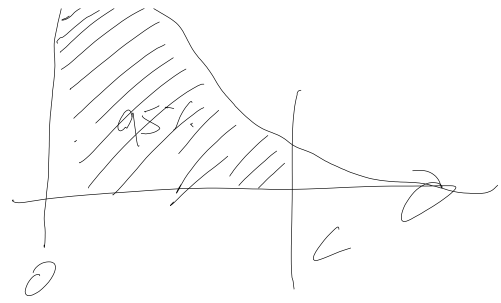
$x \sim Exp(1)$

mean life lenght? $\Rightarrow$ asking for $E(x)$

$Exp(\lambda) = \frac{1}{\lambda} \Rightarrow Exp(1) = \frac{1}{1} = 1$

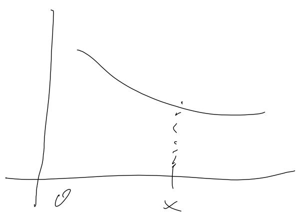
what is the timelenght that 90% of phones will die

What is the timelength that 90% of photons will die



$$0.95 = \int_0^c x f(x)\,dx$$

$$F(x) = \int_0^x f(x)\,dx \leftarrow \text{cdf}$$



$$\int_0^x e^{-x}\,dx$$

$$1 - e^{-c} = 0.95$$

$$e^{-c} = \frac{0.95 - 1}{-1}$$

$$\ln(e^{-c}) = \ln(0.05)$$

$$c = -\ln(0.05)$$

Conditional Distribution

x:	0	1	2	3
P[X=x]	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$\text{mean} = (0)\left(\tfrac{1}{8}\right) + 1\left(\tfrac{3}{8}\right) + 2\left(\tfrac{3}{8}\right) + 3\left(\tfrac{1}{8}\right)$$
$$= 1.5$$

Is this valid Probability Distribution?

Yes because

1) $0 \le P(X=x) \le 1$
2) $\sum P[X=x] = 1$

Condition: cellphone survived 6 month then you won't have the same distribution

x:	~~0~~	1	2	3
P[X=x]	~~$\frac{1}{8}$~~	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

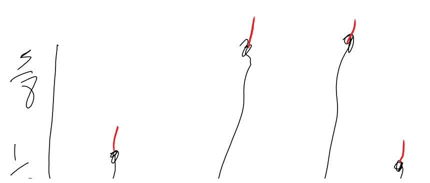
Then 0 isn't an option

This is not a valid probability distribution since it does not add up to 1.

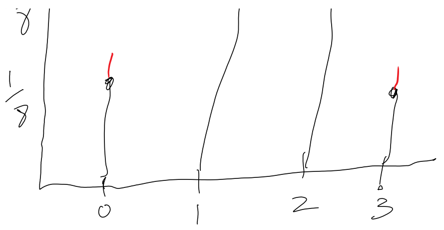
First make it a valid Probability distribution

$P(X=1 \mid X>0)$?

$$\frac{P(X=1 \cap X>0)}{P(X>0)} = \frac{P(X=1)}{P(X>0)} = \frac{\frac{3}{8}}{\frac{7}{8}} = \frac{3}{7}$$



make it a bit bigger



make it a bit bigger for PDF = 1

if $X \sim Exp(1)$

PDF: $f_X(x) = e^{-x}$

$E[X \mid X>1] = \int x \cdot \underline{\text{conditional function}}$

$$f_{X|X>1}(x) = \frac{f_X(x)}{P[X>1]} \Leftarrow \text{condition needed for } \Sigma \text{ to } = 1$$

$$= \frac{e^{-x}}{\int_1^{\infty} e^{-x}\,dx}$$

$$= \int_1^{\infty} x\, e^{-(x-1)}\,dx$$

Review 5.1

Example → Samples
↳ Distribution
↳ Inference → commenting on the calculation

* If X = life length
what is mean/expected life length?
↳ asking for expected value $E(X)$
↳ by what time 95% of the machines

↳ by what time 95% of the machines will fail.

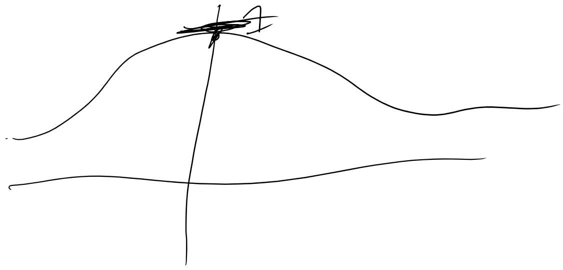
calculate cdf, equal it to 0.95* or value given and solve it.

2nd question: conditional, given $X > 1$

↳ mean/expected life?

$E(x \mid x > 1)$

predict life time distribution; mean or mode



mode = highest point

To find # take derivative set it = 0

max

min

5.2.7 example where you have to calculate the mode.

5.2.8 $x \sim Po(\lambda)$

* Predict the future value? mean / mode

mode

$$P[X=x] = \frac{\lambda^x e^{-\lambda}}{x!}$$, # if discrete no derivative

$$P[X=x+1] = \frac{\lambda^{x+1} e^{-\lambda}}{(x+1)!}$$

$$\frac{P[X=x+1]}{P[X=x]} = \frac{\lambda^{\cancel{x+1}} \cancel{e^{-\lambda}}}{(x+1)!} \div \frac{\cancel{\lambda^x} \cancel{e^{-\lambda}}}{x!}$$

$$= \lambda \frac{x!}{(x+1)!} = \frac{\lambda}{x+1}$$

$$\Rightarrow \frac{P[X=x+1]}{P[X=x]} = \frac{\lambda}{x+1}$$

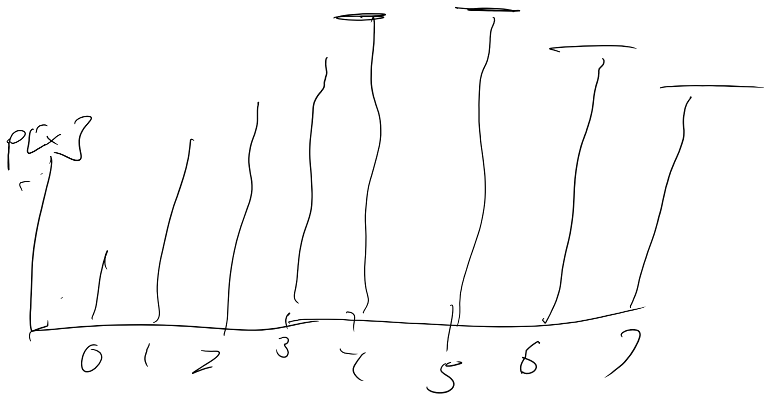
$x=0 \Rightarrow \dfrac{P[X=1]}{P[X=0]} = \lambda$ → if $\lambda = 5$ then its 5x ratio

$x=1 \Rightarrow \dfrac{P[X=2]}{P[X=1]} = \dfrac{\lambda}{2}$ → if $\lambda = 5$ then its 2x ratio

$$\frac{P[X=3]}{P[X=2]} = \frac{5}{3} = 1.666$$

$$\frac{P[X=4]}{P[X=3]} = \frac{5}{4} = 1.25$$

$$\frac{P[X=5]}{P[X=4]} = \frac{\nu}{5}$$



5.3 Statistical Models

① Population: a combination of all the subjects in your subspace. All the outcome.

Sample: A small subset of the population
Sample ⊆ Population

Parameter: Any characteristics of a population is a parameter

Statistic: summary of sample

populations are too big, samples are easier to work with.

Study sample ⇒ make inference about population. That is statistics

Eg $Exp(\underset{\uparrow}{1})$ parameter , $P_\theta(x)$

$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ x is r.v, λ is the parameter.

Statistical model

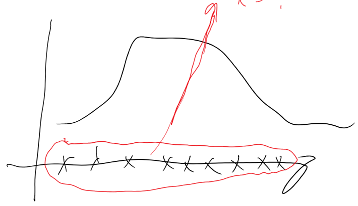
$\{ P_\theta : \theta \in \Omega \}$ — Ω: parameter space; P_θ: function e.g. poisson, normal etc...

θ can take any value in the sample space

$\frac{e^{-5}5^x}{x!} \sim P_o(5)$

$\frac{e^{-100}(100)^x}{x!} \sim P_o(100)$

Eg $x \sim P_o(\lambda = ?\ )$



$\bar{X} \to \frac{1}{\lambda}$

we use $\bar{x}$ to estimate $\frac{1}{\lambda}$. This is an example of point estimation.

Interval estimation – review

Example: Pg 264

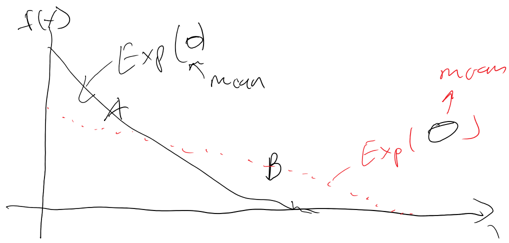
Ex 5.3.2

$P_\theta(x)$ = $f_\theta(x)$

discrete continuous

Sample $(x_1, x_2, \ldots, x_5)$

In this case, we know the parameters as the mean is already there.



← we should be able to distinguish between if a function has a mean of 1 or 1.5.

① $(x_1, \ldots, x_5) = (5.0, 3.5, 3.3, 4.1, 2.8)$

② $(x_1, \ldots, x_5) = (2.0, 2.5, 3.0, 3.1, 3.1)$

} which one belongs with which sample?

① ⇒ Exp(1.5)

② ⇒ Exp(1)

Parameter space in this case is $\Omega = \{1, 1.5\}$

$\Omega = \{A, B\}$

Ex 5.3.1

① Exp ← know function

② $\Omega = \{1, 1.5\}$ ← know parameter space

③ sample ← see sample and pick

Real life

① observe sample

② find the distribution

– known or assumed

② sample ⇐ sel'sample and pick

– known or assumed

③ Parameter - use sample to make inference about parameter

5.3.7

a)

parameter

θ	$P_\theta(x=1)$	$P_\theta(x=2)$	$P_\theta(x=3)$
A	0.5	0.5	0
B	0	0.5	0.5

Ω space $\{A, B\}$

b) if we observe $x=1$ then x is coming from A

if we observe $x=3$ then x is coming from B

if we observe $x=2$ then either a or b.

parameter

θ	$P_\theta(x=1)$	$P_\theta(x=2)$	$P_\theta(x=3)$
A	$\frac{1}{4}$	$\frac{3}{4}$	0
B	0	0.5	0.5

what if $x=2$?

probably A

Notation

$\{f_\theta : \theta \in \Omega\}$

for one sample value what is the statisc model

Ex Pois (x)

$x = 2$

$\left\{ \frac{e^{-\lambda} \cdot \lambda^2}{2!} \right\}$

Ex/ $x \sim Exp(\theta)$

$f_\theta(x) = \theta e^{-\theta x}$

$x = 5 = \theta e^{-\theta(5)}$ ⟵ statical model of one sample.

Joint density of sample

Ex/ $P[x_1, x_2, x_3 \dots x_n]$ if $\perp$ break it down

$\Rightarrow f_\theta(x_1) f_\theta(x_2) \dots f_\theta(x_n)$

$\perp$ is an assumption

$x = \{2, 4, 9\}$

$x = \{2, 4, 9\}$

$\theta e^{-2\theta} \times \theta e^{-4\theta} \times \theta e^{-9\theta}$

Ex 5.3.3

$x \sim Bin(\theta)$ $\quad p_x(x) = \begin{cases} 1 & , \theta \\ 0 & , 1-\theta \end{cases}$

$p_x(x) = \theta^x (1-\theta)^{1-x}$

$x_1, x_2, \dots x_n$

$\theta^{x_1}(1-\theta)^{1-x_1} * (\theta)^{x_2}(1-\theta)^{1-x_2} * \dots \theta^{x_n}(1-\theta)^{1-x_n}$

$= \theta^{x_1+x_2+x_3 \dots x_n} (1-\theta)^{n-(x_1+x_2 \dots x_n)}$

$= \theta^{\sum_{i=1}^n x_i} (1-\theta)^{(n - \sum_{i=1}^n x_i)}$

Ex 5.3.4

$(x_1, \dots, x_n) \sim N(\mu, \sigma^2)$

$\theta = (\mu, \sigma^2) \in R^1 \times R^+ \quad , \quad R^+ = (0, \infty)$

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

$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_2-\mu)^2} \cdot \dots \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_n-\mu)^2}$

$\frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[-\frac{1}{2\sigma^2} \underbrace{\sum (x_i-\mu)^2}\right]}$

$= \frac{1}{\sqrt{2\pi\sigma^2}}$

$$\sum_{i=1}^{n}(x_i-\mu)^2 \qquad a-b$$
$$a-c+c-b$$
$$=\sum(x_i-\bar{x}+\bar{x}-\mu)^2$$
$$=\sum\{(x_i-\bar{x})^2+2(x_i-\bar{x})(\bar{x}-\mu)+(\bar{x}-\mu)^2\}$$
$$=\sum_{i=1}^{n}(x_i-\bar{x})^2+\cancel{2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x})}+\underbrace{\sum_{i=1}^{n}(\bar{x}-\mu)^2}_{n(x-\mu)^2}$$

$$\sum(x_i-\bar{x})$$
$$=\sum x_i-\sum\bar{x}$$
sum of deviation of mean = 0

$$=n\bar{x}-n\bar{x}=0$$

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

$$s^2=\frac{\sum(x_i-\bar{x})^2}{n-1}$$ ~ sample variance

sample mean

x_i	$\bar{x}$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
2		-2	4
4	4	0	0
6		2	4
		= 0	= 8

$$s^2=\frac{\sum(x_i-\bar{x})^2}{n-1}$$
$$=\frac{8}{2}=4$$

$$\frac{1}{var}=\text{percision}$$

$$\lambda=\frac{1}{\sigma^2}$$
precision

$$(2\pi\sigma^2)^{-\frac{1}{2}}\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$
$$=(2\pi)^{-\frac{1}{2}}(\lambda)^{\frac{1}{2}}\exp\left[-\frac{1}{2}*\lambda*(x-\mu)^2\right]$$ ← still normal but in precision intead of variance

Reparameterization: only do it if its a one to one function. precision is $\frac{1}{\text{variance}}$) the change. old parameter to new parameter has to be one to one

old param ⌒func⌒ new param.

Try ex 3.3.5.

Population

Population CDF

$F_X(x) = \frac{\text{Count of } X \leq x}{N}$

Ex 5.4.1

$N = 20$

min = 3

$P(X < 3) = 0$

$P(X < 4) = \frac{3}{20}$

$P(X<4) - P(X<3) = P(X=3)$

5.4 = Population = finite
= Sample — distribution known
Unknown distribution
Parameter

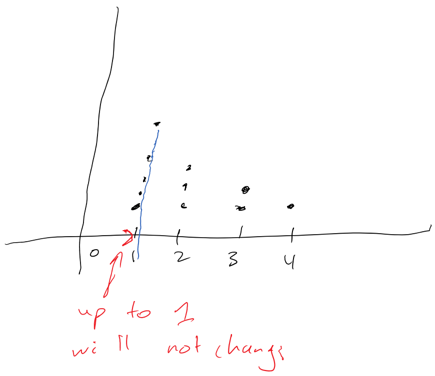
If a population is finite, do you need money / time to study them?

Ex

i	1	2	3	4	5	6	7	8	9	10
$X(\pi_i)$	1	1	2	1	2	3	3	1	2	4

Sort

i	1	1	2	1	2	3	3	1	2	4
$X(\pi_i)$	1	1	1	1	2	2	2	3	3	4



$P[x \le 0] = 0$

$P[x \le 0.999] = 0$

$P[x \le 1] = \frac{4}{10} = 0.4$

$P[x \le 1.999] = \frac{4}{10} = 0.4$

$P[x \le 2] = \frac{7}{10} = 0.7$

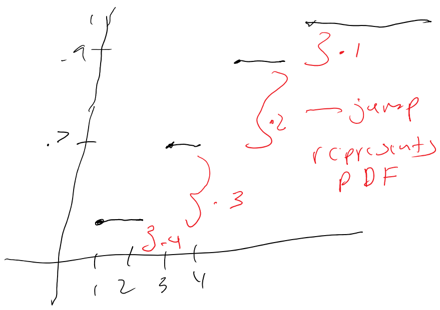
$P[x \le 2.99] = \frac{7}{10} = 0.7$

$P[x \le 3] = \frac{9}{10} = 0.9$

$P[x \le 3.99] = 0.9$

$P[x \le 4] = 1$

$$CDF = F_X(x) = \begin{cases} 0 & x < 1 \\ .4 & 1 \le x \\ .7 & 2 \le x < 3, 4 \\ .9 & 3 \le x \le 4 \\ 1 & x \le 4 \end{cases}$$



CDF on a graph

PMF:

$$f_X(x) = \begin{cases} .4, & x=1 \\ .3, & x=2 \\ .2, & x=3 \\ .1, & x=4 \\ 0, & o/w \end{cases} \Rightarrow$$

X	1	2	3	4
P[x]	.4	.3	.2	.1

To calculate small f, pdf just calculate the proportion.

CDF $= F_X(x) = \begin{cases} 0 & x<1 \\ .4 & 1\le x \\ .7 & 2\le x<3 \\ .9 & 3\le x\le 4 \\ 1 & x\le 4 \end{cases}$

population

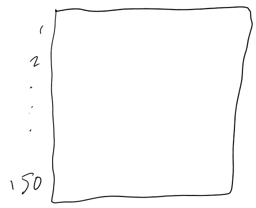
Same calculation but if its sampled its called Emperical distribution function

Emperical distribution: Same calculation but put $\hat{F}_X(x)$

5.4.1 ex/ ← do it on ya own dawg.

Simple random sampling:

Ex, in a class, blindly picking #'s



To do it in R: sample(1:150, size=1)

for a random sample betwee 1-150

$\{A, B, C, D, E\}$

$N=5$ - draw 2 samples $n=2$

$P(A) = \frac{1}{5}$

$P(A \text{ being selected}) = \frac{1}{N} = \frac{1}{5}$

$P(A \text{ being selecte} \mid b \text{ is selected}) = \frac{1}{N-1} = \frac{1}{4} = 0.25$

Samples are not $\perp$ because one being in the sample changes prob of the other one.

After picking:
- with replacement — pick and put it back → samples are not $\perp$
- without replacement - pick # or name don't replace → samples are $\perp$

$\frac{1}{N}$ $\frac{1}{N-1}$ $N = 100{,}000$

0.2 0.25 $\underbrace{0.000060}_{\text{something like that}}1 \approx$ next numbers are closer.

large samples even though they are dependent because N is large change is insignificant, so we treat as $\perp$

Conditions

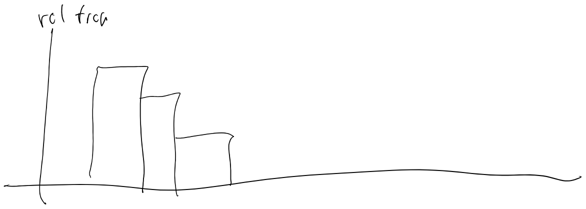
$N \to$ large
$n \to$ small relative to N

$\therefore \underbrace{\hat{F}_X(x)}_{\text{cdf calculated based on sample}} \to F_Y(x)$

Histogram

Ex height

$(h_1, h_2]$ $(h_2, h_3]$ $(h_3, h_4]$ . . . . .
5 5.6 5.6 6 6 6.5



relative frequency = proportion

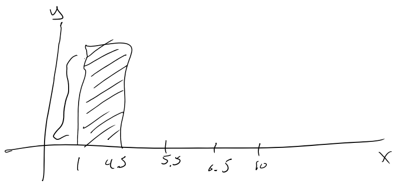
Density histogram

$h_X(x) = \dfrac{\text{proportion}}{\text{length}}$

5.4.5

a) (1.2 1.8 2.3 2.5 3.1 3.4 3.7 3.2 3.9 4.3 4.4 4.5 4.5]
(4.8 4.8] (5.6 5.8] (6.9 7.2 8.5]

$h_x(x) = \dfrac{\frac{13}{20}}{(4.5-1)}$ $(1, 4.5]$



$h_x(x) = \dfrac{\frac{13}{20}}{(4.5-1)}$, $(1, 4.5)$

$= \dfrac{13}{20}$

Question on Midterm on this.

$(4.5, 5.5]$

$h_x(x) = \dfrac{\frac{2}{20}}{(5.5-4.5)}$, $(4.5, 5.5]$

Ex

$$f_x(x) = \begin{cases} .4 & , x = 1 \\ .3 & , x = 2 \\ .2 & , x = 3 \\ .1 & , x = 4 \\ 0 & , o/w \end{cases}$$

from this list make the table

look for unique numbers

1, 2, 3, 4

count pop by total to get proportion

loop

Prof uploaded r code. look at the table.

Statistics

Descriptive — Inferential

* falls into two categories

Descriptive: describes any summary eg mean of sample
standard deviation
median

Inferential: moment you use those #'s to make a inference about the data.

Recall: $f_x(x)$ is the proportion of the population members whose X measurements equal x.

$\Rightarrow f_X(x) = P[X = x]$

$F_X(x)$ is the proportion of population members whose X measurements is less than or equal to x

Ex $\{1.2, 2.1, 0.4, 3.3, -2.1, 4.0, -0.3, 2.2, 1.5, 5.0\}$

Put in ascending order

$\{-2.1, -0.3, 0.4, 1.2, 1.5, 2.1, 2.2, 3.3, 4.0, 5.0\}$

Plug in each data point to get pdf

$f_X(x)$

$f_X(-3) = P[X = -3] = 0$

$f_X(-2.1) = P[X = -2.1] = \frac{1}{10}$

$f_X(-0.3) = P[X = -0.3] = \frac{1}{10}$

$\vdots$

$f_X(5) = P[X = 5] = \frac{1}{10}$

CDF same thing but everything up to the point

$F_X(-2.1) = P[X \leq -2.1] = \frac{1}{10}$

$F_X(-0.3) = P[X \leq 0.3] = \frac{2}{10} = \frac{1}{5}$

A natural estimate of $F_X(x)$ is given by $\hat{F}_X(x)$

$\hat{F}_X(x) = n^{-1} \sum_{i=1}^{n} I_{(-\infty, x]}(x_i)$ thats the formal definition for what we did above

This is also called emperical distribution function of x
↳ means sample.

Calculating Population Quantiles

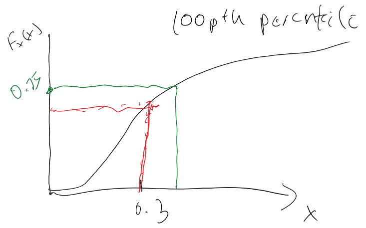
Given value calculate percentile.

p-quantile Note .75 quantile $\xrightarrow{\times 100}$ = 75th percentile

p-quantile Note .75 quantile = 75th percentile

100pth percentile $\frac{1}{100}$

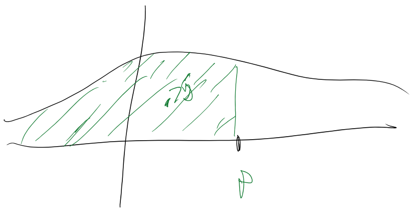
80th percentile = 0.8 quantile

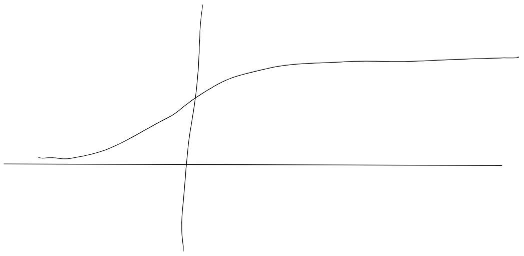


$F_x(0.3) = P[X \leq 0.3]$

Percentile is opposite

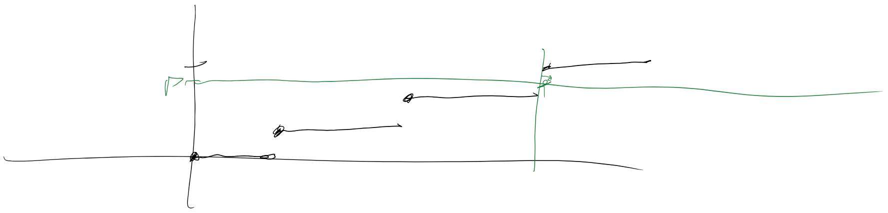
$F_x(x) = P[X \leq x] = 0.75$





$F(x_p) = p$

$x_p = ?$ if continuous cdf



$x_1 \; x_2 \; x_3 \; x_4 \; \ldots \; x_n$

$\{1.2, 2.1, 0.4, 3.3, -2.1, 4.0, -0.3, 2.2, 1.5, 5.0\}$

* if you don't have equal then go to the next one

Put in ascending order ← (x) bracket means ordered (n)

$x_{(1)} \; x_{(2)}$

$\{-2.1, -0.3, 0.4, 1.2, 1.5, 2.1, 2.2, 3.3, 4.0, 5.0\}$

$\frac{1}{10} \; \frac{1}{10} \; \frac{1}{10} \; \frac{1}{10} \; \frac{1}{10} \; \frac{1}{10} \; \frac{1}{10} = \frac{7}{10}$

70th percentile? 2.2

$P[X \leq 2.2] = 0.7$

$P = 0.75$? its a # between 2.2 - 3.3

2.2 (70% — 75% — 80%) 3.3

2.2 (10% of data) 3.3

& take $\frac{1}{2}$

2.2 (5% of data) = $2.2 + \frac{(3.3 - 2.2)}{2}$ ← interpolation

$\therefore \frac{i-1}{n} \leq p \leq \frac{i}{n}$

$\frac{7}{} \leq p \leq \frac{8}{}$

Continuous one always a solution, if discrete no solution sometimes!

$$x = x_{(i-1)} + (x_i - x_{i-1})\, n \left(p - \frac{i-1}{n}\right) \overset{\text{in this case}}{\Longrightarrow} 2.2 + (3.3-2.2)\,10\,(0.75-0.7)$$

$$= 2.2 + (1.1)(0.5)$$

$$= 2.75$$

25th percentile ⟵ midterm/final question

$$\frac{i-1}{n} < p \le \frac{i}{n}$$

$$0.2 < p \le 0.3 \quad , i = 3$$

$$x = x_{(i-1)} + (x_i - x_{i-1})\, n \left(p - \frac{i-1}{n}\right)$$

$$= -0.3 + (0.4+0.3)\,10\,(.25-.2)$$

$$= 0.05$$

* unconfirmed

if calculate percentile use this formula on midterm and final

25 th percentile = Q_1 = 1st quartile

75 th percentile = Q_3 = 3rd quartile

$P[x \le 1.5] = 0.5$ $\qquad$ $P(x \ge 1.5] = 0.6$

$\{1.2, 2.1, 0.4, 3.3, -2.1, 4.0, -0.3, 2.2, 1.5, 5.0\}$

\# Put in ascending order

$\{-2.1, -0.3, 0.4, 1.2, 1.5, 2.1, 2.2, 3.3, 4.0, 5.0\}$

median

In one definition $P[x \le 1.5] = 0.5$ is enough to call a median but in another definition $P[x \ge 1.5) = 0.6$ will not since it != 0.5.

Recall if n is odd → $\frac{n+1}{2}$ th term

even → $\dfrac{\frac{n}{2}\text{th} + \frac{n}{2}+1 \text{ term}}{2}$

$P(x \ge 1.8) = 0.5$ ✓

$P(x \le 1.8) = 0.5$ ✓

1.5	2.1
	1.8

Two definition of median, use easiest one.

1, 3, 5 median is $\frac{n+1}{2} = \frac{3+1}{2} = $ 2nd term

Interquartile range: (width of data)

$IQR = Q_3 - Q_1$

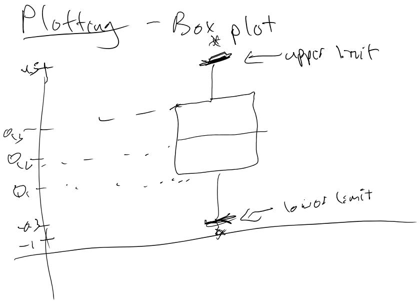
replacement of SD.

extreme outliers median doesn't suffer

Skew → median

Symmetric → mean.

Plotting - Box plot



lower limit = $Q_1 - 1.5 * IQR$

upper limit = $Q_3 + 1.5 * IQR$

Box plots also for continuous cases.

lower-limit = $Q_1 - 1.5 * (Q_3 - Q_1) = -1$ ← means line goes down to −1

stop at −0.3   * — outliers

upper limit = $Q_3 + 1.5 \times (Q_3 - Q_1)$

$= 4.5$

stops at 4.5

$\{-2.1 \,|\, \overset{-1}{} -0.3, 0.4, 1.2, 1.5, 2.1, 2.2, 3.3, 4.0 \,|\, \overset{4.5}{} 5.0\}$

4.5 is also an outlier.

whiskers = *

adjacent value: 4.5, −1

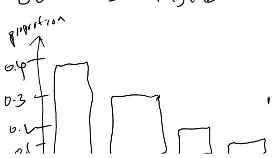
Ex/

(1) Cuc 0.42
(2) Van 0.28
(3) BS 0.22
(4) St 0.08

categorical variable has no order.

if categorical stuff above does not apply.

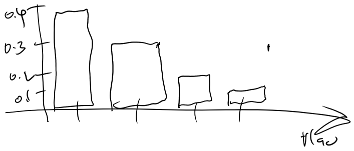
Do this instead



Steps

put shit in order

determine region



determine region

calculate percentile

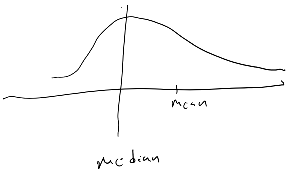
Recap

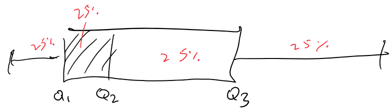
$\tilde{x}_p = x_{(i-1)} + (x_{(i)} - x_{(i-1)})\, n \left(p - \frac{i-1}{n}\right)$

odd: $\left(\frac{n+1}{2}\right)$th term

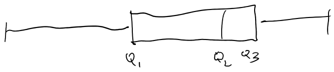
even: $\dfrac{\frac{n}{2}\text{th} + \left(\frac{n}{2}+1\right)\text{th}}{2}$

skewness:



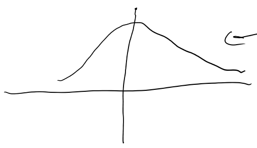


← right skewed distribution
↳ skew part is in the right



← left skewed distribution

if skewed use median and <u>not</u> the mean



← mean = median = mode in standard normal table.

5.2 $X \sim Exp(1)$

① Predict future value
② $P[x > 5]$
③ $P[X \leq \underline{a}] = 0.95$

} ← These are inferences about the future X. (observation)

5.3 inference is about distribution

In this chapter $Exp(\theta) \to$ estimate θ

chi
↓
$\psi(\theta) = \frac{1}{\theta}$ mean $= \frac{1}{\theta}$ $var(\theta) = \frac{1}{\theta^2}$

$\psi(\theta) = \frac{1}{\theta^2}$ median $= \int_0^m f_\theta(x)\, dx = 0.5$

$\Rightarrow F(m) = 0.5$

$\Rightarrow m = F^{-1}(0.5)$

$\boxed{\psi(\theta) = F_\theta^{-1}(0.5)}$ median

25^{th} percentile: $\psi(\theta) = F_\theta^{-1}(0.25)$

Inverse ex/ Exp(1)

$f_\theta(x) = e^{-x}$

$F_\theta(x) = \int_0^x e^{-x}\,dx = 1 - e^{-x}$

$$F_\theta(m) = 1 - e^{-m} = 0.5$$
$$e^{-m} = 0.5$$
$$-m = \ln(0.5)$$
$$\Rightarrow \quad m = -\ln(0.5)$$
$$\Rightarrow \quad F_\theta^{-1}(0.5)$$

$f_\theta(x) \longrightarrow \text{Normal}(\mu, \sigma^2)$

$\psi(\theta) \longleftarrow$ Parameter of interest

$\psi(\theta) = \mu$

$x_1, x_2 \ldots x_n \leftarrow$ can you use this to get mean?

yea $\rightarrow \bar{x}$ use sample mean

$T(s) = \frac{1}{n} \sum x_i$

$T(s)$ is an estimate of $\psi(\theta)$

Variance: $\psi(\theta) = \sigma^2$

$\underset{\overset{\shortparallel}{v}}{T(s)} = \frac{\sum (x_i - \bar{x})^2}{n-1}$ is an estimate of $\underset{\overset{\shortparallel}{\psi(\theta)}}{\sigma^2}$

$\bar{x} \longrightarrow E(x)$

$n \longrightarrow$

5.5.8

3rd moment $N(\mu_0, \sigma_0^2)$

mean not

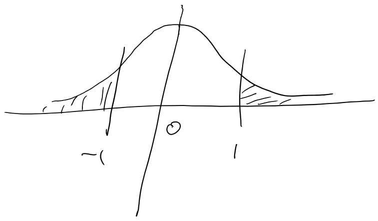
sigma not squared

$$\psi(\theta) = E(x^3) = E[x-\mu+\mu]^3$$

$$= E[(x-\mu)^3 + 3(x-\mu)^2\mu + 3(x-\mu)\mu^2 + \mu^3]$$

$$= E[(x-\mu)^3] + E[3(x-\mu)^2\mu] + E[3(x-\mu)\mu^2] + E[\mu^3]$$

$$= \underbrace{E[(x-\mu)^3]}_{\text{3rd central moment}} + 3\mu \underbrace{E[(x-\mu)^2]}_{\substack{var(x) \\ \shortparallel \\ \sigma^2}} + 3\mu^2 E[(x-\mu)] + \mu^3$$



$$\int_{-\infty}^{0} x f(x) + \int_{0}^{\infty} x f(x)$$

$$x \le 1 = x \ge 1$$

if the distribution is symmetric around (μ)

$$E[x-\mu]^1 = 0$$

$$E[x-\mu]^3 = 0$$

$$E[x-\mu]^5 = 0$$

$$\underbrace{E[(x-\mu)^3]}_{\substack{\text{3rd central moment} \\ \shortparallel \\ 0}} + 3\mu \underbrace{E[(x-\mu)^2]}_{\substack{var(x) \\ \shortparallel \\ \sigma^2}} + \underbrace{3\mu^2 E[(x-\mu)]}_{\substack{\shortparallel \\ 0}} + \mu^3$$

$$= 3\mu\sigma^2 + \mu^3$$

if $N(\mu, \sigma^2)$, whats θ? $\theta = \{\mu, \sigma^2\}$

means all list of parameter

$$G(\alpha, \beta) \Rightarrow \theta = \{\alpha, \beta\}$$

October 1, 2018 3:10 PM

Section: 5.2

- Predicting future value
- $X \sim (\lambda)$
  - → mean
  - → mode: predicting x
  - → $P[x>5]$, $P[x>5 \mid x>1]$
  - → $E[x \mid x>1]$

Section 5.3

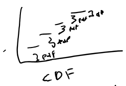
$S = x_1, x_2, x_3 \ldots x_n$ (Lets say you have a sample, what is $f_\theta(s)$ ← probability function of the sample)

only if ⫫: $f_\theta(s) = f_\theta(x_1) \times f_\theta(x_2) \ldots f_\theta(x_n)$

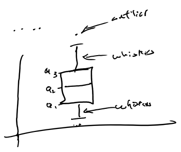
- reparameterization: $\theta^2 - \theta$, one to one then we can reparameterize

Section: 5.4

$\underbrace{F_x(x)}_{\text{population CDF}} \rightarrow f_x(x)$



- Once you know the small f, can calculate everything.
- $\hat{F}_x(x)$ ← empirical distribution function
- histogram
- $\hat{f}_x(x)$
  - mean
  - median $\hat{x}_p = \ldots$
- Q_1, Q_3
- IQR
- $Q - 1.5(IQR)$
- box plot



- $\theta = \mu, \sigma^2$

  $E(x^3) = \ldots$ (refer to last class notes)

6.1 Likelihood Function

$X \sim Bern(\theta)$ ⫫ $\theta \in [0,1]$

$S = \{x_1, x_2, \ldots x_n\}$

$[f_\theta(s) = f_\theta(x_1) \times f_\theta(x_2) \ldots f_\theta(x_n)]$

Toss coins 5 times:

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array}$$
$$[1, 0, 1, 1, 0] = S$$

$$\left.\begin{array}{l} f_\theta(s) = \theta(1-\theta)\theta\theta(1-\theta) \\ \underbrace{L(\theta|s)}_{\text{likelihood function of theta given the sample.}} = \theta^3(1-\theta)^2 \end{array}\right\} \text{likelihood of } \theta$$

* Not the probability of observing θ. It is the probability of observing the sample for a given true θ.

θ → fixed true value, $\theta \in \Omega$

θ_2 any other member from Ω

believe $f_{\theta_1}(s) > f_{\theta_2}(s)$

The more data the smaller the likelihood.

Ex/ 6.1.1

Suppose $S = \{1, 2, \ldots\}$ and that the statistical model is $\{P_\theta : \theta \in \{1,2\}\}$, where P_1 is the uniform distribution on the integers $\{1, \ldots, 10^3\}$ and P_2 is the uniform distribution on $\{1, \ldots, 10^6\}$

Dist 1 → f_{θ_1} → Unif $\{1, 2, 3, \ldots 1000\}$

→ Unif $\{1, 2, 3, \ldots \ldots 1000000\}$

Dist 1 $\to$ 001

Dist 2 $\to f_{\theta_2} \to$ Unif $\{1, 2, 3, \ldots \ldots 1000000\}$

$S = \{1001\}$ $\to$ comes from distribution 2. Since distribution 1 stops at 1000.

$S = \{10\}$ $\to$ # calculate likelihood for both and compare

$$\frac{L(\theta_1|S) = \frac{1}{1000}}{L(\theta_2|S) = \frac{1}{1000000}} \Rightarrow$$ sample 10 is a 1000 x more likely to come from θ_1

Example 6.1.4 $X \sim N(\theta, \sigma_0^2)$, σ_0^2 known?

1 unknown parameter θ, mean

$S = x_1 \cdot x_2 \ldots x_n$

$$f_\theta(s) = \prod_{i=1}^{n} \underbrace{(2\pi\sigma_0^2)^{-\frac{1}{2}}}_{\text{constant}} \exp\left[-\frac{1}{2\sigma_0^2}(x_i - \theta)^2\right]$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\theta)^2\right]$$

recognize this & expand it, Do it yourself.

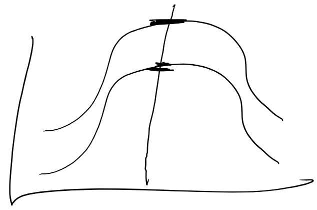
$$= \underbrace{(2\pi\sigma_0^2)^{-\frac{n}{2}}}_{\text{constant}} \exp\left[-\frac{n}{2\sigma_0^2}(\bar{x}-\theta)^2\right] \underbrace{\exp\left[-\frac{1}{2\sigma_0^2}(n-1)s^2\right]}_{\text{constant}}$$

sample variance

Since constant rewrite

$$L(\theta|S) \propto \exp\left[-\frac{n}{2\sigma_0^2}(\bar{x}-\theta)^2\right] \Rightarrow$$

proportional



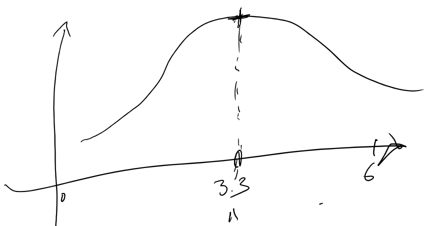
Same graph but just stretched as a result you can make inference about mean

$n = 25$, $\sigma_0^2 = 1$, $\bar{x} = 3.3$

$$L(\theta|S) \propto \exp\left[\frac{-25}{2}(3.3-\theta)^2\right]$$

recall

$y = 5x$
$y \propto x$ } proportional to x



$\hat{\theta} \to$ MLE

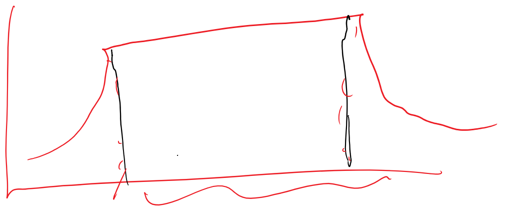
① Point of estimate of θ

$\underbrace{\hat{\theta}}_{\text{estimate}} \to \theta$

MLE $\to$ maximum likelihood estimator of θ

$L(\hat{\theta}|s) \geq L(\theta|s)$

Note you can have multiple MLE for ex/



Ex 6.2.1

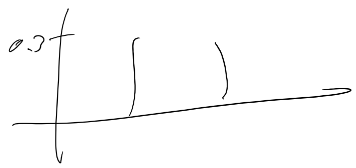
	$s=1$	$s=2$	$s=3$
f_1	0.3	0.4	0.3
f_2	0.1	0.7	0.2

$\{S=1\}$ ⇒ probably came from 1^{st} one

$\{S=2\}$ ⇒ probably came from 2^{nd} one

$\{S=3\}$ ⇒ probably came from 1^{st} one

	$s=1$	$s=2$	$s=3$
f_1	0.3	0.4	0.3
f_2	0	0.7	0.3



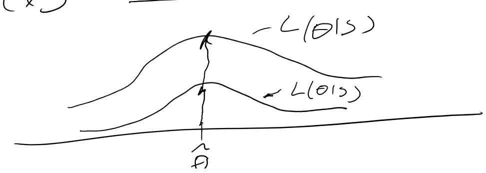
$\{S=3\}$ ⇒ MLE is probably θ_1 or θ_2

$$L(\theta \mid S) = \prod_{i=1}^{n} f_\theta(x_i)$$

$\Downarrow$ # taking the ln allows us to change to summation

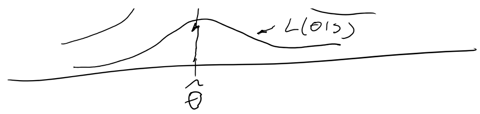
$$\ln(L(\theta \mid S))$$

$\ln(x) \xrightarrow{1-1 \text{ function}} x$



$$\ln(L(\theta|s)) = \sum_{i=1}^{n} \ln f_\theta(x_i)$$

Summa...



Now maximize

$$L(\theta|s) \propto \exp\left[\frac{-n}{2\sigma_0^2}(\bar{x}-\theta)^2\right]$$

proportional

1) $$\ell(\theta|s) = \frac{-n}{2\sigma_0^2}(\bar{x}-\theta)^2$$ ← ln likelihood of θ

Differentiate

2) Score function $= \dfrac{\partial \ell(\theta|s)}{\partial\theta}$

3) score equation $= \dfrac{\partial \ell(\theta|s)}{\partial\theta} = 0$

$\hat{\theta} = \boxed{\text{something}}$

To check if max or min, second derivative.

$$\left.\frac{\partial^2 \ell(\theta|s)}{\partial^2\theta}\right|_{\theta=\hat{\theta}} < 0$$

$$\ell(\theta|s) = \frac{-n}{2\sigma_0^2}(\bar{x}-\theta)^2$$

$$\frac{\partial \ell(\theta|s)}{\partial\theta} = -\frac{2n}{2\sigma_0^2}(\bar{x}-\theta)(-1)$$

$$= \frac{n}{\sigma_0^2}(\bar{x}-\theta)$$ ← score

$$\Rightarrow \frac{n}{\sigma_0^2}(\bar{x}-\theta) = 0$$

$$\theta = \bar{x}$$

check MLE

$$\text{score} = \frac{n}{\sigma_0^2}(\bar{x}-\theta)$$

$$\frac{\partial^2 \ell(\theta|s)}{\partial^2(\theta)} = -\frac{n}{\sigma_0^2}$$

n sample size

σ^2 positive

n, σ

$\delta^2(\theta)$

$\therefore \frac{-n}{\sigma_0^2} \leq 0$

$\therefore \bar{X}$ is the MLE of θ

Do it for exp distribution

Question midterm or final or both, learn to calculate MLE of everything distribution we know.

Invariance

$f_\theta(x)$ $\theta \to \hat{\theta}$ is MLE

$\underbrace{\psi(\theta)}$ function of θ

one to one

how do you find MLE ?

$\psi(\theta) \to \psi(\hat{\theta})$

6.2.2 If $(x_1, \dots, x_n)$ is a sample from $Ber(\theta)$ distribution where $\theta \in [0,1]$ is unknown, then determine the MLE of θ

$X \sim Ber(\theta) \longleftarrow$ mle ?

$S = x_1, x_2, \dots x_n$

write the distribution plug $x_1, x_2 \dots x_n$

Likelihood $L(\theta | s) = \theta^{x_1}(1-\theta)^{1-x_1} \cdot (\theta)^{x_1}(1-\theta)^{1-x_2} \cdot \dots \cdot (\theta)^{x_n}(1-\theta)^{1-x_n}$

$= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$

get ln likelihood

ln likelihood $\ell(\theta|s) = \ln(\theta^{\sum_{i=1} x_i}) + \ln(1-\theta) \cdots$

$$= \sum_{i=1}^{n} x_i \ln\theta + (n - \sum_{i=1}^{n} x_i)\ln(1-\theta)$$

get first derivative

$$\frac{\partial \ell(\theta|s)}{\partial\theta} = \sum_{i=1}^{n} x_i \cdot \frac{1}{\theta} - (n - \sum_{i=1}^{n} x_i)\cdot\frac{1}{1-\theta} \quad \Big\} \text{ score}$$

set it to zero to get score equation

$$\frac{\partial \ell(\theta|s)}{\partial\theta} = \sum_{i=1}^{n} x_i \cdot \frac{1}{\theta} - (n - \sum_{i=1}^{n} x_i)\cdot\frac{1}{1-\theta} = 0$$

$$\Rightarrow \frac{\sum x_i}{\theta} = \frac{n - \sum x_i}{1-\theta}$$

$$\Rightarrow \frac{1-\theta}{\theta} = \frac{n - \sum x_i}{\sum x_i}$$

$$\Rightarrow \frac{1}{\theta} - 1 = \frac{n}{\sum x_i} - 1$$

$$\Rightarrow \frac{1}{\theta} = \frac{n}{\sum x_i}$$

$$\sum x_i = n\theta$$

$$\hat{\theta} = \frac{\sum x_i}{n} = \bar{x}$$

check second derivative then you can say its MLE

Ex/ $\theta \rightarrow \bar{x}$ MLE
$\theta^2 \rightarrow \bar{x}^2$ $\Big\}$ since one to one on Ber(θ), [0,1]

$\Theta^2 \to \bar{X}^2$ ✓

6.2.5 Suppose $(x_1, \ldots, x_n)$

$X \sim \text{Uniform}[0, \theta]$

$f_\theta(x) = \frac{1}{\theta} I[x \in [0, \theta]]$

$x_1, x_2, \ldots, x_n$

$L(\theta | s) = \frac{1}{\theta} \cdot \frac{1}{\theta} \cdots \frac{1}{\theta}$

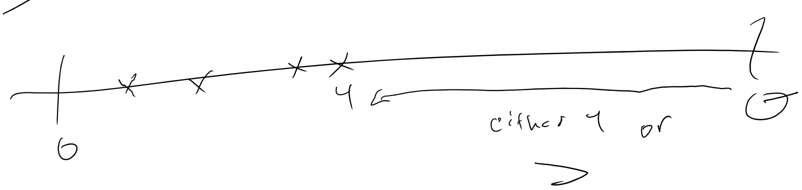
$= \frac{1}{\theta^n}$

Differentiation part doesn't work

$\frac{1}{\theta^n} \to$ this is maximized when bottom is minimized.

$\boxed{\therefore \text{ when } \theta \text{ is minimum } L(\theta) \text{ is maximum}}$

Ex assume 1, 1.2, 2.3, 4



minimum θ = max value in sample

$\hat{\theta} = \max(s)$

October 3, 2018 3:10 PM

Midterm: November 3^{rd}
1–4 pm
IC #130

Recap:

Likelihood
↓ take ln of likelihood to get MLE
MLE $\frac{\partial}{\partial \theta} = 0 \Rightarrow \hat{\theta}$
$\frac{\partial^2 \ell}{\partial \theta^2} < 0$

Ex $X \sim Bern(\theta)$

$s_1 = (1,0,1,1,0,0,1)$ $1 \Rightarrow H$
$s_2 = (0,1,0,1,0,1,1)$

$L(\theta \mid s_1) = \theta(1-\theta)\theta\theta(1-\theta)(1-\theta)\theta$
$= (1-\theta)^3\theta^4$

$L(\theta, s_2) = (1-\theta)\theta(1-\theta)\theta(1-\theta)\theta\theta$
$= (1-\theta)^3\theta^4$

Ex/ $n = 7$
\# of heads $= 4$

sol $\underbrace{\binom{7}{4}\theta^4(1-\theta)^3}_{L(\theta|s)}$

All 3 give you the same likelihood functions. because $\sum x_i = 4$, sum of random variables is 4.

Sufficient statistics

Important stuff

taking sum of sample
$T(s_1) = T(s_2)$
$\Rightarrow L(s_1) = L(s_2)$

May or may not have constant in the front. Constant doesn't matter when maximizing it.

T = sum of all your sample.
$f_\theta(s) = h(s) * \underbrace{g_\theta(T(s))}_{\theta,\ T(s)}$

$h(s)$: free of θ, only function of sample

$T(s)$: sufficient stat

Ex
$\underbrace{\binom{7}{4}}_{h(s)}\theta^4\underbrace{(1-\theta)^3}$

In general
$\underbrace{\theta^{\sum x}(1-\theta)^{n-\sum x}}_{g_\theta(\ \ )}$

Ex $X \sim Exp(\theta)$

$x_1, x_2, \ldots \ldots x_n$

Sol
$f_\theta(s) = \theta e^{-\theta x_1} \cdot \theta e^{-\theta x_2} \cdot \ldots \ldots \theta e^{-\theta x_n}$
$= \theta^n e^{-\theta \sum x_i}$
$\Rightarrow h(s) = 1$ because both terms are of θ

$= \underbrace{1}_{h(s)} \cdot \underbrace{\theta^n e^{-\theta \sum x_i}}_{g_\theta(\ )}$ ← sufficient statistic

Ex $X \sim Pois(\theta)$

$S = x_1, x_2, \ldots, x_n$

$$f_\theta(S) = \frac{e^{-\theta}\theta^{x_1}}{x_1!} * \frac{e^{-\theta}\theta^{x_2}}{x_2!} \times \cdots \frac{e^{-\theta}(\theta)^{x_n}}{x_n!}$$

likelihood function

$$= \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!}$$

$h(S) = \frac{1}{\prod_{i=1}^{n} x_i!}$

$g_\theta = (e^{-n\theta}\theta^{\sum x_i})$

$T(S) = \sum x_i$

if [2, 3, 4, 2, 1] is # of accidents on the 401

and follows $Po(\theta)$ then

$$\frac{e^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i} = e^{-5\theta}\theta^{12}$$

, don't need to know 5 different #, you just need the sum

take ln

$\frac{12}{5}$

, sum is sufficient stat of data

from likelihood you can get scenarios of x but not sample.

T(S) can be retain back from your likelihood is called a sufficient statistic.

sufficient means data reduction.

$T(s)$

↑

minimal sufficient stat means you can't reduce further

if $s_1 = [x_1, x_2, \dots x_n]$

$s_2 = [y_1, y_2, \dots y_n]$

$$\frac{\theta^n e^{-\theta \Sigma x}}{\theta^n e^{-\theta \Sigma y}} = e^{-\theta(\Sigma x - \Sigma y)} \Rightarrow \Sigma x = \Sigma y$$

if θ

$\Rightarrow$ minimal statistic