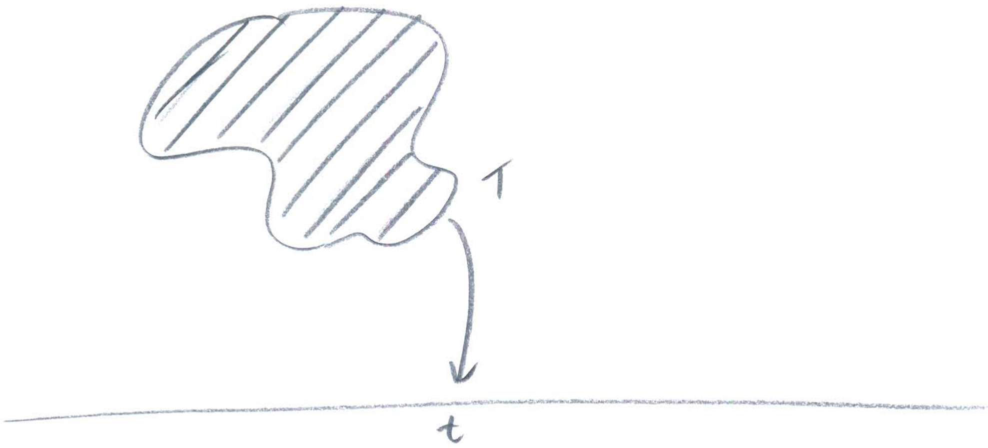
36-705 13/05/2020

L8 - convergence

YouTube 21/09/16

- T is a sufficient statistic $T(X_1, \dots, X_n)$

if $p(x_1, \dots, x_n | T = t, \theta)$ does not depend on θ ✓



LW: Think of a statistic as a mapping, but also as a partition

e.g. $T(X_1, \dots, X_n) = \bar{X}_n$

- lines represent all points in sample space with same value $\bar{X}_n$
- whenever you have a statistic that defines a partition of the space, and you need to talk about the statistic or the partition induced by the statistic

LW: clearer to talk about statistics in terms of partitions

- 2 examples: calculation, conceptually revealing one

Notation: $X^n = (X_1, \dots, X_n)$ $x^n = (x_1, \dots, x_n)$

Example 1

- $X_1, \dots, X_n \sim Po(\theta)$ - show T is sufficient (using def.)

- let $T = \sum_{i=1}^n X_i$

- $P_\theta(X_1, \dots, X_n | t) = \dfrac{P_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, T = t)}{P_\theta(T = t)}$

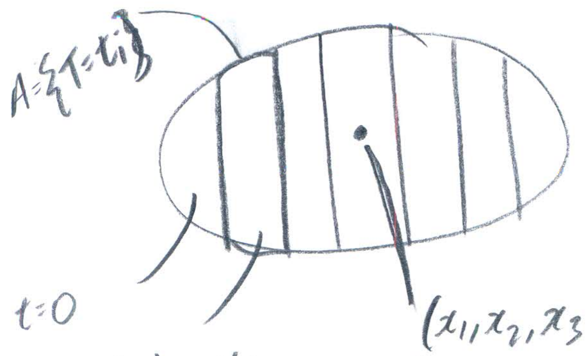
e.g. $(1, 3, 4)$

$p(X_1 = 1, X_2 = 3, X_3 = 4 | T = 8)$

$T(X_1, X_2, X_3) = t$

$t = 8?$ $t = 17?$

LW: claim is that θs drop out



- sum fn creates partition int.
- Have to ensure statistic value agrees. / outside/diff part.
- If $t=0$ for example, intersection of t and (x_1, x_2, x_3) is 0.

(*) - Think intersection of event A and B

- So $\dfrac{P_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, T = t)}{P_\theta(T = t)} = \begin{cases} 0 & \text{if } T(x_1, \dots, x_n) \neq t \\ P_\theta(X_1 = x_1, \dots, X_n = x_n) & \text{if } T(x_1, \dots, x_n) = t \end{cases}$ ✓ (A1b)

- So $p_\theta(x_1, x_2, \dots, x_n | T = t) = \dfrac{p_\theta(x_1, \dots, x_n)}{p_\theta(T = t)} \overset{\text{IID}}{=} \dfrac{\prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!}}{\frac{e^{-n\theta}(n\theta)^t}{t!}}$

- Note $T = \sum_{i=1}^n X_i \sim Po(n\theta)$
- (W). Property proved with MGFs
- "probability of getting sequence $X^n = x^n$ given $T = \sum_{i=1}^n X_i = t$"

(W) (A1): check deriv

Yielding $p_\theta(x_1, \dots, x_n | T = t) = \dfrac{t!}{(\prod_{i=1}^n x_i)!\, n^t}$

- which does not depend on θ.

LW: compute $P(X_1 = 1, X_2 = 3, X_3 = 4) \longrightarrow$ require θ (using Poisson)
compute $P(X_1 = 1, X_2 = 3, X_3 = 4)$ given sum $T = 8 \longrightarrow$ don't require θ

(*) once I have the sum T (statistic), I have all info and can compute probability without knowing θ.

(*) There are other sufficient statistics e.g. data
- There are other sufficient statistics $\longrightarrow$ see notes

LW: something missing

3.2 Sufficient partitions

example 2

$X_1, X_2, X_3 \sim Bern(\theta)$

- statistic e.g. $T = \sum_i X_i$ induces a partition

$(X_1$	X_2	$X_3)$	t	$p(x\|t)$
0	0	0	0	1
0	0	1	1	1/3
0	1	0	1	1/3
1	0	0	1	1/3
0	1	1	2	1/3
1	0	1	2	1/3
1	1	0	2	1/3
1	1	1	3	1

- Any other statistic that induces the same partition is an <u>equivalent statistic</u>
- The statistics can be thought of as an <u>equivalence class</u>
- E.g. use average $T = \frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow$ same <u>partition induced</u>

(*) values of statistic in a sense irrelevant, only partition matters
- Use partition / statistic semantically <u>interchangeably</u>

(*) consider $p(x|t)$ OT

e.g. $P(X_1 = 0, X_2 = 0, X_3 = 1 | T = 1) = \frac{(1-\theta)(1-\theta)\theta}{3(1-\theta)(1-\theta)\theta} = \frac{1}{3}$ A3

(*) all of conditional distri do not have θ

Q: Should the partitions be countable (in general NO; <u>finite discete</u> case)

h A2: Review <u>key definitions</u>

Q: Any rules of thumb for gauging sufficiency? → will get there.

example 3

- Not a sufficient statistic example
- define statistic as outcome of 1^{st} flip of coin
- intuitively; info being binned, will not expect it to be sufficient
- verify formally (via partitions)
  - compute $p(x|t)$; and check if it depends on θ.

lw: Theorem to help you assess sufficiency

3.3 Factorisation Theorem

Theorem 4

- If $p(x_1, x_2, \ldots, x_n; \theta) = h(x_1, \ldots, x_n)\, g(t; \theta)$

function of $x_1, \ldots$

this function allowed to depends on θ only though t.

a realisation of statistic T

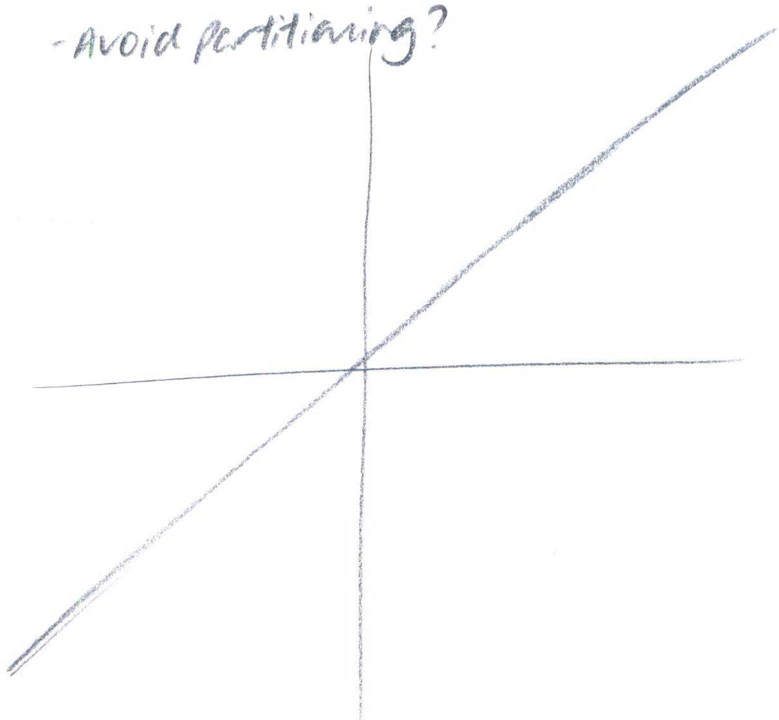
$\Rightarrow T$ is sufficient

- Avoid partitioning?

example 6

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$

- suppose σ^2 known; only consider sufficiency for μ



$p(x_1, \ldots, x_n; \mu)$

$= \prod_{i=1}^{n} p(x_i; \mu)$

$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x_i-\mu)^2}{2\sigma^2}\right\}$

$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right\}$

$| T = \sum_{i=1}^{n} x_i = t$

(w) (93): review trick

key trick here for stats
- add and subtract $\bar{x}$
→ cross term 0

$= \underbrace{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{\frac{-\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2}\right\}}_{h(x_1, \ldots, x_n)} \underbrace{\exp\left\{\frac{-n(\bar{x}-\mu)^2}{2\sigma^2}\right\}}_{g(t; \mu)}$

· Therefore, $\bar{X}_n$ is sufficient for μ

W: But $\bar{X}_n$ also appears in $g(t;\mu)$?

W A4 - review query

- conclude - sat.

- Anything induces some partition e.g. sum is also sufficient

(Neyman factoris.)

W A5: To check:-

(and minimal)

b) If (μ, σ^2) is unknown; $T=(\bar{X}_n, S^2)$ is sufficient, so is $T=(\sum_1 X_i, \sum_1 X_i^2)$

LW: Note that 2 dim parameter yields 2d sufficient statistic (dimensionality does not always hold)

Sufficient statistics for $N(\mu, \sigma^2)$, σ^2 known (examples)

i. $\sum_1 X_i$

ii) $\frac{1}{n}\sum_{i=1}^{n} X_i$

iii) $17\bar{X}_n$

v) $(\bar{X}_n, X_3, X_8)$

vi) $(X_1, \dots, X_n)$

vi) $(\bar{X}_n, S_n^2)$

· In a certain sense, LHS 'better' than RHS

· Note that you could in principle, and in practice, compute LHS given RHS, but not the other way around.

· Sufficient statistics on LHS are a function of RHS

· In terms of partitions; we can formalise:-

3.4 Minimal sufficient statistics

· T is minimally sufficient if:-

i) T is sufficient

ii) If U is any other sufficient statistic then $T=g(U)$ for some function g

"Does not contain redundant information"

· E.g. $\bar{X}_n$ and $X_1, \dots, X_n$ (w)(16): review transforms of each other (one-line)

$\bar{X}_n = g(X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$

$(X_1, \dots, X_n) \overset{?}{=} g(\bar{X})$ "cannot get $(X_1, \dots, X_n)$ via a function of $\bar{X}_n$"

↝ not possible

(*) so T generates the coarsest sufficient partition (OT)

· can see sufficient statistics (not minimal ones) as defining sub-partitions in partitions induced by the minimal sufficient statistic

example 9

(*) U is sufficient → $p(x|u)$ does not depend on θ

{ BUT not minimal; as I can take partition induced by T and subdivide it into a finer partition

↳ - equivalent to saying $T = g(U)$

- UW: cannot make partitions induced by T coarser
  - Trying to make it coarse would make $p(x|t)$ depend on θ
  - (W): not entirely sure how this works.

- UW (*) You can make a minim. sufficient statistic narrower by further subdividing its partitions, to yield a subdivided partition and a new sufficient statistic U.
  - But you cannot make it coarser without creating a dependency of $p(x|t)$ on θ

(Q): what is $g(\cdot)$ on the diagram?
- mapping from $u \to t$

(Q): How many minimal sufficient statistics, are they unique?
- NO

3.5 How to find a MSS

Theorem 10

- Define

$$R(x^n, y^n; \theta) = \frac{p(y^n; \theta)}{p(x^n; \theta)}$$

- Suppose T has the following property:-

$R(x^n, y^n; \theta)$ does not depend on $\theta \iff T(x^n) = T(y^n)$

- Then T is a MSS.

U - Suppose you had a guess of a sufficient statistic (e.g. mean), and you want to know if it's minimal sufficient

(*) Form R, check implications both ways:-

i) If R does not depend on θ then $T(x^n) = T(y^n)$

ii) If $T(x^n) = T(y^n)$ then R does not depend on θ

Example 11

$Y_1, \dots, Y_n \sim \text{IID } Po(\theta)$

$$\frac{p(y_1, \dots, y_n; \theta)}{p(x_1, \dots, x_n; \theta)} = \frac{p(y^n; \theta)}{p(x^n; \theta)} = \frac{\dfrac{e^{-n\theta} \theta^{\Sigma y_i}}{\prod_{i=1}^n (y_i)!}}{\dfrac{e^{-n\theta} \theta^{\Sigma x_i}}{\prod_{i=1}^n (x_i)!}} = \frac{\theta^{\sum_{i=1}^n y_i - x_i}}{\prod_{i=1}^n (y_i!) / \prod_{i=1}^n (x_i!)}$$

(*) Check conditions

- e.g. candidate MSS - $T(Y^n) = \sum_{i=1}^n Y_i$

- i) Does $T(x^n) = T(y^n)$ follow from R not depending on θ.

- In order for R not to depend on θ, we require $\theta^0 \Rightarrow \sum_{i=1}^n (y_i - x_i) = 0$

- So this condition checks as R not dep. $\theta \Rightarrow T(y^n) = T(x^n)$

(*) (ii) Does setting $T(x^n) = T(y^n) \Rightarrow$ ~~imply~~ imply that R does not depend on parameter θ?

- Yes (shown from above)

- (*) $T(Y^n) = \sum_{i=1}^{n} Y_i$ is MSS

$\boxtimes$

- sample av. is also MSS as a corollary

$\bar{Y}_n$

(*) sufficiency of statistic T means you have all info you need to compute the likelihood function; not that it contains all the information in the data.