36-705 11/07/2020

Lecture Notes 5: Statistical Inference - review

(A1) - Having difficulty with example 1

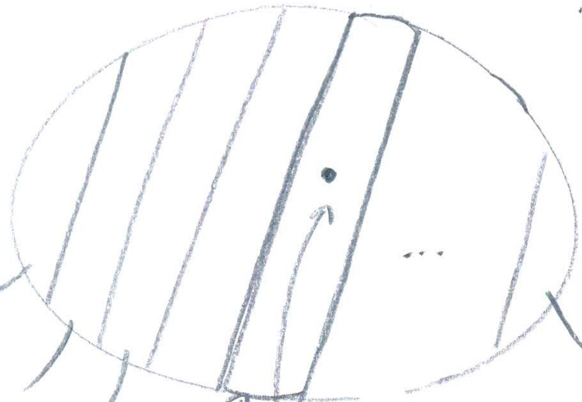
- i.e. deriving:-

$$\frac{P_\theta(X_1=x_1,\dots,X_n=x_n,T=t)}{P_\theta(T=t)} = \begin{cases} 0 & \text{if } T(x_1,\dots,x_n)\neq t \\ P_\theta(X_1=x_1,\dots,X_n=x_n) & \text{if } T(x_1,\dots,x_n)=t \end{cases}$$

(*) Helps to anchor this example with example data

- sample space

- each point $\rightarrow x^n = (X_1=x_1,\dots,X_n=x_n)$
- $T=\sum_{i=1}^n X_i$ creates partition
- statistic



$T=1$

$T=2$

$T=3$

...

$t=$

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

- event B is a point

$A=\{T=\sum_{i=1}^n X_i=t\}$

- event A is a region of sample space (all points where the sum of the r.v.s is a number, t)

- Note $(A\cap B)$ i.e. $\{X_1=x_1,\dots,X_n=x_n,T=t\}$ is a point

- so $P(A\cap B)=P(X_1=x_1,\dots,X_n=x_n,T=t)$ is 0 when $T\neq t$
- when $T=t$, $P(A\cap B)=P(X_1=x_1,\dots,X_n=x_n,T=t)=P(X_1=x_1,\dots,X_n=x_n)=P(B)$

(use logic of probability as 'proportion' of sample space) in diagram.

- this yields $P_\theta(X_1=x_1,\dots,X_n=x_n,T=t) = \begin{cases} 0 & \text{if } T(x_1,\dots,x_n)\neq t \\ P_\theta(X_1=x_1,\dots,X_n=x_n) & \text{if } T(x_1,\dots,x_n)=t \end{cases}$

(A1) check derivation:-

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

$X \sim Po(\lambda)$

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

$T = \sum_{i=1}^n X_i \quad X_i \sim Po(\lambda)$

$\Rightarrow T \sim Po(n\theta)$

$$\Rightarrow p_\theta(x_1,\dots,x_n|T=t) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n(x_i!)} \cdot \frac{t!}{e^{-n\theta}(n\theta)^t}$$

$$= \frac{\theta^t t!}{(n\theta)^t\prod_{i=1}^n(x_i!)} = \frac{t!}{n^t\prod_{i=1}^n(x_i!)} \neq f(\theta)$$

- As required - Hence $T = \sum_{i=1}^n X_i$ is a sufficient statistic for θ

3.2. Sufficient Partitions

1. A partition $B_1, \dots B_K$ is sufficient if $f(x|X \in B)$ does not depend on θ
2. A statistic T induces a partition. For each t $\{x: T(x)=t\}$ is one element of the partition.
   T is sufficient iff the partition is sufficient.
3. Two statistics can generate the same partition.
4. If we split any element B_i of a sufficient partition into smaller pieces, we get another sufficient partition.

(A3) - Note:- $P(X_1=0, X_2=0, X_3=1 | T=1) = \dfrac{P(X_1=0, X_2=0, X_3=1)}{P(T=1)}$

$$= \frac{(1-\theta)(1-\theta)\theta}{3(1-\theta)(1-\theta)\theta} = \frac{1}{3}$$

Theorem 4: Factorisation

- If $p(x_1, \dots, x_n; \theta) = h(x_1, \dots, x_n) g(t; \theta)$ then T is **sufficient**

(Q). Not entirely clear to me the restrictions on functional forms.

- Googling/Wiki-ing:-

$h(x_1, \dots, x_n)$ - does NOT depend on θ

$g(t; \theta) = g(t(x_1, \dots, x_n), \theta)$ - depends on θ, depends on $(x_1, \dots, x_n)$ (i.e. data) only through $t(x_1, \dots, x_n)$

Examples

$X_1, \dots, X_n \sim Po(\theta)$

$$p(x^n; \theta) = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n (x_i)!} = \frac{1}{\prod_{i=1}^n (x_i)!} \cdot e^{-n\theta} \theta^{\sum_{i=1}^n x_i}$$

$$= \underbrace{\frac{1}{\prod_{i=1}^n (x_i)!}}_{= h(x_1, \dots, x_n)} \cdot \underbrace{e^{-n\theta} \theta^{T(x_1, \dots, x_n)}}_{g(T(x_1, \dots, x_n), \theta)}$$

Trick:- (has to be verified)

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

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

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

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

$= \sum_{i=1}^{n} \{(x_i-\bar{x})^2 + (\bar{x}-\mu)^2\} \underbrace{- 2n\bar{x}^2 + 2n\mu\bar{x} - 2\mu n\bar{x} + 2n\bar{x}^2}_{=0 \text{ 'crossterms cancel'}}$ as $\sum x_i = n\bar{x}$

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

(*) Note: $\sum_{i=1}^{n} \{(x_i-\bar{x})^2 + (\bar{x}-\mu)^2\} = \sum_{i=1}^{n} (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$

Example 6 (b) (μ, σ^2) unknown

- X, x used interchangeably (i.e. r.v., realisation)
- just be aware

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

$T = (\bar{X}, S^2) \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

$$p(X^n; \mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{\frac{-\sum_{i=1}^{n}(X_i-\bar{X})^2}{2\sigma^2}\right\} \exp\left\{\frac{-n(\bar{X}-\mu)^2}{2\sigma^2}\right\}$$

As $\sum_{i=1}^{n} (X_i - \bar{X})^2 = (n-1)S^2$

$$p(X^n; \mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{\frac{-(n-1)S^2}{2\sigma^2}\right\} \exp\left\{\frac{-n(\bar{X}-\mu)^2}{2\sigma^2}\right\}$$

? confused as to a particular nuance.
- does $h(X^n)$ have to be directly a function of the data $(X_1, \dots, X_n)$?
- not entirely sure which element constitutes $h(X^n)$ and $g(T(X^n); \mu, \sigma^2)$

O/S 1 (see end of review)

3.4 Minimal suff stats

(*) some expo on distinction between sufficient statistic and minimal sufficient statistics :-

- suppose U is sufficient (i.e. $p(x|u)$ does not depend on θ)
- suppose $T = H(U)$ is also sufficient
- T provides a 'greater reduction' than U unless H is a 1-1 transform; in which case, T and U are equivalent.

(W): Did not understand Example 9 on review on why U is not minimal.
- Note U is not minimal because it defines a sub-partition

$n=73$
$n=73$

$n=91$

- i.e. pay attention to diagram

Theorem 10

- provides a way of checking if a statistic is MSS

(*)(*): -

i) R does not depend on $\theta \Rightarrow \theta^0 = \sum_{i=1}^{n}(y_i - x_i) = 0$ $(\longrightarrow T(x^n) = T(y^n))$

$\Rightarrow \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i$

$\Rightarrow T(y^n) = T(x^n)$ as our candidate MSS was $T(y^n) = \sum_{i=1}^{n} y_i$

ii) $T(x^n) = T(y^n) \longrightarrow R$ does not depend on θ.

Example 12

Cauchy distn: $p(x;\theta) = \dfrac{1}{\pi(1+(x-\theta)^2)}$

Then
$$R(x^n, y^n; \theta) = \frac{p(y^n;\theta)}{p(x^n;\theta)} = \frac{\prod_{i=1}^{n} \frac{1}{\pi(1+(y_i-\theta)^2)}}{\prod_{i=1}^{n} \frac{1}{\pi(1+(x_i-\theta)^2)}}$$

$$= \frac{\prod_{i=1}^{n} \{1+(x_i-\theta)^2\}}{\prod_{i=1}^{n} \{1+(y_i-\theta)^2\}}$$

(?) 0/S2

(*):- <u>example 6</u>

$$p(x^n;\mu,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{\frac{-(n-1)S^2}{2\sigma^2}\right\} \exp\left\{\frac{-n(\bar{X}-\mu)^2}{2\sigma^2}\right\}$$

- Here we <u>have</u>:-

$$h(x_1,\dots,x_n) = 1$$

$$g(\bar{X},S^2,\mu,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)S^2 + n(\bar{X}-\mu)^2\right]\right)$$

- Factorisation theorem applies to joint PMF/PDF ; distinct from showing condit. distn of (suff) data on statistic does not depend on θ.