Statement of theorem from last time: Stat 562

2-19-19

①

Let $X_1, \dots, X_n \sim \text{iid } f(x|\theta)$

Let T have the following property:

$*\left[\text{the ratio } \dfrac{f(\vec{x}|\theta)}{f(\vec{y}|\theta)} \text{ is free of } \theta \text{ iff } T(\vec{x}) = T(\vec{y}).\right]$

Then T is a minimal sufficient statistic for θ.

<u>Proof</u>. Let $\tau = \{t : t = T(\vec{x}) \text{ for some } \vec{x}\}$

Let $A_t = \{\vec{x} : T(x) = t\}$

②

For each set A_t, choose one of its elements and call it $\vec{x}_t$

Now, for any $\vec{x}$, $\vec{x}$ will lie in A_t for some t. Specifically, $\vec{x} \in A_{T(\vec{x})}$

Also, $T(\vec{x}) = T(\vec{x}_{T(\vec{x})})$

Assume that T has the property $*$

Then $\dfrac{f(\vec{x}|\theta)}{f(\vec{x}_{T(\vec{x})}|\theta)}$ is free of θ.

③

Let $h(\vec{x}) =$ that ratio

So $f(\vec{x}|\theta) = \underbrace{h(\vec{x})}_{\text{free of } \theta} \underbrace{f(\vec{x}_{T(\vec{x})}|\theta)}_{\text{function of } t, \theta}$

$\Rightarrow T$ is a sufficient statistic for θ

Let T^* be any other sufficient statistic for θ. We need to show that T is a function of T^*.

④

Since T^* is sufficient,

$$f(\vec{x}|\theta) = \underbrace{h^*(\vec{x})}_{\text{free of } \theta} g^*(t^*|\theta)$$

Let $\vec{x}$ and $\vec{y}$ be any 2 points in the sample space where $T^*(\vec{x}) = T^*(\vec{y})$

Then $\dfrac{f(\vec{x}|\theta)}{f(\vec{y}|\theta)} = \dfrac{h^*(\vec{x})\, \cancel{g^*(t^*|\theta)}}{h^*(\vec{y})\, \cancel{g^*(t^*|\theta)}}$, free of θ

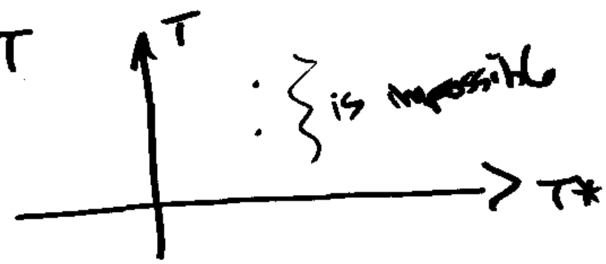
(5)

Since T was assumed to have the $*$ property, $T(\vec{x}) = T(\vec{y})$

That is, $T^*(\vec{x}) = T^*(\vec{y}) \Rightarrow T(\vec{x}) = T(\vec{y})$

Plot T^* vs T

(T^*, T)



$\therefore$ T is a function of T^*

$\therefore$ T is a minimal sufficient stat. for θ.

(6)

Defn. $S(\vec{X})$ is an ancillary statistic for θ if its probability distribution is free of θ.

Example: $X_1, \ldots, X_n \sim$ iid $N(\theta, \sigma^2)$ ← known

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

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$, which doesn't involve θ.

So S^2 is an ancillary stat. for θ.

7

Defn: Let $\{ f(t|\theta) : \theta \in \Omega \}$ be a family of pdfs or pmfs of a statistic T. The family is called complete if

$$E[g(T)] = 0 \ \forall \theta \Rightarrow P(g(T)=0) = 1$$

Also, in this case, T is called a complete statistic.

8

Example: Suppose that the statistic T has a binomial distribution $B(n,\theta)$.

That is, the family of distributions is

$$\left\{ \binom{n}{t} \theta^t (1-\theta)^{n-t} : 0 < \theta < 1 \right\}$$

Suppose that $E[g(T)] = 0 \ \forall \theta$

$$\sum_{t=0}^{n} g(t) \binom{n}{t} \theta^t (1-\theta)^{n-t} = 0 \quad \forall \theta$$

9

$$= (1-\theta)^n \sum_{t=0}^{n} g(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^t \quad \text{let } r = \frac{\theta}{1-\theta}$$

$$= \underbrace{(1-\theta)^n}_{\neq 0} \underbrace{\sum_{t=0}^{n} g(t) \binom{n}{t} r^t}_{\text{polynomial in } r} = 0 \quad \forall \quad 0 < r < \infty$$

From a theorem in calculus, all of the coefficient of the polynomial must be 0

That is, $g(t)\binom{n}{t} = 0 \quad \forall \quad t = 0, 1, \dots, n$

10

So $g(t) = 0 \quad \forall \quad t = 0, \dots, n$

Then $P(g(T) = 0) = 1$

Basu's Theorem

Let $X_1, \dots, X_n$ be iid $f(x|\theta)$, $\theta \in \Omega$.

Let T be a complete sufficient statistic for θ

and let S be an ancillary " " "

Then S and T are independent.

Proof. $h(s,t) = h(s|t)g(t)$ (11)

$$h(s) = \int_{-\infty}^{\infty} h(s|t)g(t)\,dt$$

Also $h(s) = h(s)\cdot 1$

$$= h(s)\int_{-\infty}^{\infty} g(t)\,dt$$

So $0 = \int_{-\infty}^{\infty} h(s|t)g(t)dt - h(s)\int_{-\infty}^{\infty} g(t)dt$

$$= \int_{-\infty}^{\infty}\left[h(s|t) - h(s)\right]g(t)dt$$

$0 = E\left[h(s|T) - h(s)\right]$, true $\forall \theta$ (12)

$\Rightarrow$ $h(s|t) - h(s) = 0$ with probability 1, since T was complete.

$\therefore$ $h(s|t) = h(s)$ a.s.

$\therefore$ S & T are independent

Theorem: Suppose that $f(x|\vec{\theta})$ is a member of the exponential family,

i.e. $f(x|\vec{\theta}) = h(x)c(\vec{\theta})e^{\sum_{j=1}^{k} w_j(\vec{\theta})t_j(x)}$

Then $T(\vec{x}) = (t_1(\vec{x}), t_2(\vec{x}), \ldots, t_k(\vec{x}))$

is complete if

$\{w_1(\vec{\theta}), \ldots, w_k(\vec{\theta})\}$ contains an

open ball in $\mathbb{R}^k$.

Proof: left for the 600-level course