No.

Date.

Data Reduction.

Statistical Models. $X_1 \cdots X_n \sim p(x;\theta)$.

nonparametric. parametric. usually assumed.

has a uniform form.

example: $P = \{$all distributions on $\mathbb{R}^2\}$.

$P = \{p$ as probability $| \int (p''(x))^2 dx < \infty\}$.

example: $P = \{p(x;\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-}$

statistics $T(X)$ reduce the data.

$\mathcal{T} = \{t \mid t = T(X),\ X \in S\}$.

$A_t = \{x \mid T(x) = t\}$.

Principles:

- Sufficiency Principle: keep the information vital for θ.
- Likelihood Principle:
- Equiv. Principle:

① SP.

Definition: if $T(X) = T(X_1 \cdots X_n)$, s.t. any inference about θ depends only on T.

T is a sufficient statistic for θ.

In other words: $\forall (x_1 \cdots x_n), (y_1 \cdots y_n)$

if $T(x) = T(y)$, then the inference are the same.

Mathematically:

Suppose: $X_1 \cdots X_n \sim p(x;\theta)$, then.

if $p(x_1, \cdots x_n \mid T(X) = t;\ \theta) = p(x_1 \cdots x_n \mid T(X) = t)$

p is sufficient.

Theory: $p(x;\theta)$ is the joint pdf/pmf of X.

$q(T(x);\theta)$ is the joint pdf/pmf of $T(x)$.

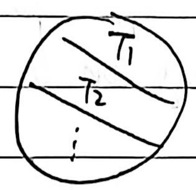
$T(x)$ is sufficient iff $\frac{p(x;\theta)}{q(T(x);\theta)}$ is independent from θ.

proof $p(x|T(X)=t;\theta) = P_\theta(X=x|T(X)=t)$

if $t \neq T(x)$, $P_\theta(X=x|T(X)=t)=0$.

else: $P_\theta(X=x|T(X)=t) = P_\theta(X=x|T(X)=T(x))$

$$= \frac{P_\theta(X=x, T(X)=T(x))}{P_\theta(T(X)=T(x))} = \frac{P_\theta(X=x)}{P_\theta(T(X)=T(x))} = \frac{p(x;\theta)}{q(T(x);\theta)}.$$

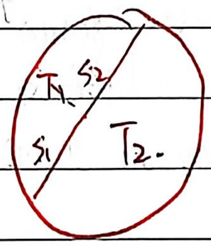
S^n:  $\rightarrow$ partion by T

A partition is sufficient if

$p(x|x\in T_i)$ does not rely on θ.

That is $p(x|x\in T_i) = p(x|x\in T_i, \theta)$.

再分仍是sufficient的.

如:  T_1内分布与θ无关.

故S_1, S_2内与θ无关.

常理,望 $p(\theta|X=x, T(x)=t) = p(\theta|T(x)=t)$

if consider θ as a random variable.

$$\Leftrightarrow \frac{P(\theta,X,T)}{P(X,T)} = \frac{P(\theta,T)}{P(T)}$$

$$\Leftrightarrow \frac{P(\theta,X,T)}{P(\theta,T)} = \frac{P(X,T)}{P(T)}.$$

$$\Leftrightarrow p(x_1 \cdots x_n|T;\theta) = p(x_1 \cdots x_n|T)$$

mark:

For Frequentist, θ is an unknown but exact number, for which formula like $P(\theta=2)$ is invalid.

While for Bayesian we consider θ as a random variable, following a distribution.

Theory:

$T(x)$ is sufficient for θ iff the joint pdf / pmf of X can be factored as:

$$p(x;\theta) = h(x)\cdot g(T(x);\theta)$$

含参函数，$p(x;\theta) = P_\theta(X=x)$

sample space $\to \mathbb{R}$

① suppose $T(x)$ is sufficient.

choose $g(T(x);\theta) = P_\theta(T(X)=T(x))$, $h(x) = P_\theta(X=x \mid T(X)=T(x))$

since $T(x)$ is sufficient, $h(x)$ does not rely on θ.

then $p(x;\theta) = P_\theta(X=x) = P_\theta(X=x, T(X)=T(x)) = \underbrace{P_\theta(T(X)=T(x))}_{g.} \cdot \underbrace{P_\theta(X=x \mid T(X)=T(x))}_{h.}$

$= h(x)\cdot g(T(x);\theta)$

② Let $q(T(x);\theta)$ be the pdf / pmf of $T(x)$ (suppose pmf).

We need to prove that $\frac{p(x;\theta)}{q(T(x);\theta)}$ does not rely on θ.

$$\frac{p(x;\theta)}{q(T(x);\theta)} = \frac{h(x)\, g(T(x);\theta)}{q(T(x);\theta)}.$$

$(q(T(x);\theta) = P_\theta(T(X)=T(x)).)$ when pmf

Define $A_{T(x)} = \{y \mid T(y) = T(x)\}$.

Then $q(T(x);\theta) = \sum_{y\in A_{T(x)}} p(y;\theta) = \sum_{y\in A_{T(x)}} h(y)\cdot g(T(y);\theta)$

$= g(T(x);\theta)\cdot \sum_{y\in A_{T(x)}} h(y)$ $\quad (\forall y\in A_{T(x)},\ T(y)=T(x))$.

$\therefore$ ratio $= \frac{h(x)}{\sum_{y\in A_{T(x)}} h(y)}$ does not rely on θ $\square$.

Minimal Sufficient Statistic.

$T(X)$ Minimal Sufficient if

① T Sufficient.

② If T_0 is sufficient, $\exists$ map g s.t. $T = g(T_0)$.

No.

Date.

Theory:

Let $p(x;\theta)$ be the pdf/pmf of X. ($p(x;\theta) = P_\theta(X=x)$ when pmf).

let $T(x)$ be a statistic.

$x \neq y$ are samples. key point: X, Y are random Variables. while x, y are random samples/values

Let $R(x,y;\theta) := \frac{p(y;\theta)}{p(x;\theta)}$.

[if $R(x,y;\theta)$ does not rely on $\theta \Leftrightarrow T(y)=T(x)$
Then T is a minimal sufficient statistic.] → when θ changes, $p(y)/p(x)$ does not change, which can be categorized in to one equivalence class!

proof: to simplify, assume that $p(x;\theta)>0$ for all x and θ

let $\mathcal{T} = \{t \mid t=T(x)\}$, $A_t := \{x \mid T(x)=t\}$.

for each A_t, choose a fixed $x_t \in A_t$ ∴ $\exists\, g$ as a map s.t. $x_t = g(t)$.

we have x and $x_{T(x)}$ in the same set. (They are allowed to be the same.)

∴ $T(x) = T(x_{T(x)})$

define $h(x) = \frac{p(x;\theta)}{p(x_{T(x)};\theta)}$ which does not rely on θ.

$$\therefore\ p(x;\theta) = h(x) \cdot \underbrace{p(x_{T(x)};\theta)}_{x_{T(x)} = g(T(x))}$$

$\Rightarrow T(\cdot)$ is a sufficient statistic.

Let T' is another sufficient statistic.

∴ exist h', g'

$p(x;\theta) = h'(x)\cdot g'(T'(x);\theta)$ mark: $g'(T'(x);\theta)$ is not a pdf or pmf. It's just a function with θ. while $p(x;\theta)$ in some degree should be the same......

if $T'(x) = T'(y)$

$\Rightarrow \frac{p(x;\theta)}{p(y;\theta)} = \frac{h'(x)}{h'(y)}$

$\Rightarrow T(x) = T(y)$ □.

No.

Date. . .

feature:

if T is a minimal statistic of X.

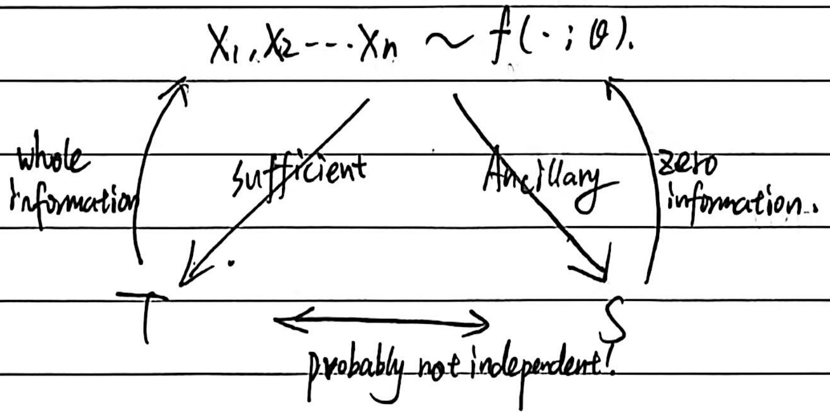
Then if T' is another, there exist a bijection f s.t.

$$T = f(T').$$

$\Rightarrow$. the two partition induced by T & T' are the same.

Ancillary Statistic 辅助统计量.

which is usually named as $S(x)$, whose distribution does not rely on θ at all. They're useless. $p(S(x);\theta) = p(S(x))$



Complete statistic.

T is complete

If $E_\theta(g(T)) = 0$ for all θ $\Rightarrow P_\theta(g(T)=0) = 1$.