2.2. Statistics, Sufficiency and Completeness

11.03.2020

(A) What is a statistic?

- data set is realisation of a random vector X from unknown popl. P

Def.: Let T be measurable fct. of X. $T(X)$ is called a statistic if it is a known value whenever X is known. ↗ T known fct.

↗ "known" means here sth. different to "known" of a population P on space $(\Omega, \mathcal{F}, P)$, i.e. $P(A)$ known for all $A \in \mathcal{F}$

→ ~~Diff~~ statistic is a characteristic of a sample like mean, variance, min, max … ↗ given when X (sample) given.

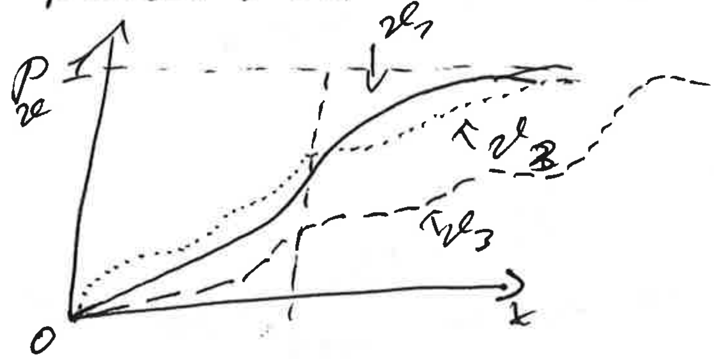
↗ Difference between a parameter $\vartheta \in \mathbb{R}$ and a statistic $T(X)$?

(1) Parameter ϑ related to entire population, i.e. ϑ enters the parametric family P_ϑ as variable and is principally unknown. ~~(parametric family / set on the sample)~~

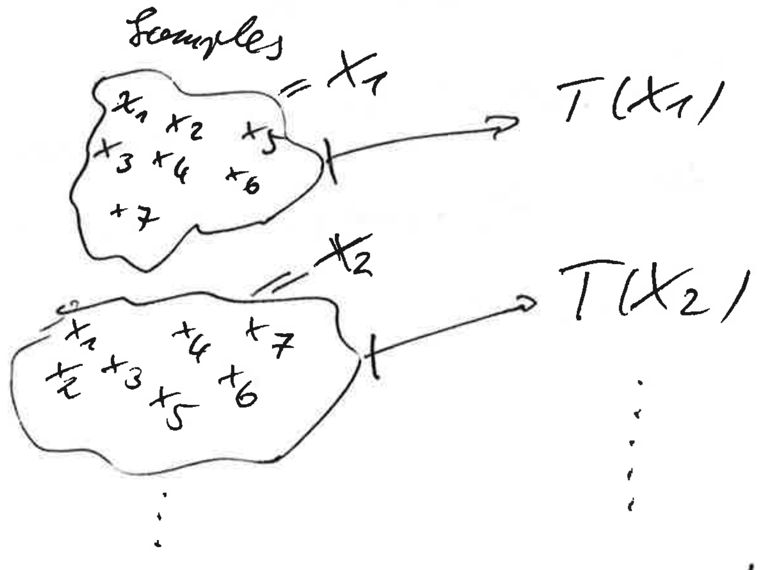
(2) A statistic $T(X)$ is a measurable map ~~of~~ X, e.g. $T: X \to \mathbb{R}$. X is a finite sample $(X_1, X_2, \dots, X_n)$ and $T(X)$ is a quantity which is computed by the given sample/data.

(3) In pictures:

parameter $\vartheta \in \mathbb{R}$



statistic $T(X)$



- Goal is still is getting information about population of X.

→ information about P within statistic $T(X)$ is given in $\sigma(T(X))$ (and if S any other statistic with $\sigma(S(X)) = \sigma(T(X))$, then T is mb. map of S and vice versa.)

→ Important: $\sigma(T(X)) = \sigma(X)$, iff T is one-to-one. Otherwise, $\sigma(T(X)) \subsetneq \sigma(X)$ and hence, a statistic can be seen as a reduction of the σ-field.

Observe also that ~~any~~ $T(X)$ is a random element.

Question: $T(X)$, in general, reduces $\sigma(X)$. How much information is lost? ⟶ Sufficiency

(B) How to characterize information of X and $T(X)$?

Def.: Let X be a sample from an unknown population $P \in \mathcal{P}$, where $\mathcal{P}$ is a family of populations. A statistic $T(X)$ is said to be sufficient for $P \in \mathcal{P}$ (or for $\vartheta \in \Theta$, when $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$ is parametric) iff the cond. distr. of X given T is known (does not depend on P or ϑ).

→ $T(X)$ contains some information concerning unknown pop. P as original data X do.

→ Sufficiency depends on given family $\mathcal{P}$! If T is sufficient for $P \in \mathcal{P}$, then T is sufficient for $P \in \mathcal{P}_0 \subset \mathcal{P}$, but not necessarily for $P \in \mathcal{P}_1 \supset \mathcal{P}$.

Example: Let $\{X_i\}_{i=1}^{n}$ be iid and $X = (X_1, X_2, \dots, X_n)$.

X_i are from PDF (w.r.t. counting m.) $f_\vartheta(z) = \vartheta^z (1-\vartheta)^z \mathbb{1}_{\{0,1\}}(z)$

for $z \in \mathbb{R}$, $\vartheta \in (0,1)$.

→ Realisations x of X are sequence of ones and zeros. → $(0,1,1,0,\dots)$, $(1,0,1,1,\dots)$

Choose statistic $T(X) = \sum_{i=1}^{n} X_i$. Is T sufficient?

(1) Thinking: Unknown quantity is ϑ in f_ϑ. Regarding ϑ, only the total number of ones (ϑ is prob. of ~~ones~~ ones in x) is of importance. Information about the detailed arrangement is not providing more information about ϑ and hence redundant.

→ $T(X)$ should be sufficient for ϑ.

(2) Computation: $P(X=x \mid T=t) = \dfrac{P(X=x, T=t)}{P(T=t)}$, where

$$P(T=t) = \binom{n}{t} \vartheta^t (1-\vartheta)^{n-t} \mathbb{1}_{\{0,1,\dots,n\}}(t).$$

Denote the ith component of x by x_i.

→ If $t \neq \sum_{i=1}^{n} x_i$, then ~~$P(X=x \mid T=t) = 0$~~ $P(X=x, T=t) = 0$

(since impossible!).

→ If $t = \sum_{i=1}^{n} x_i$, then

$$P(X=x, T=t) = \prod_{i=1}^{n} P(X_i = x_i)$$

$$= \prod_{j=1}^{t} P(X_j = 1) \cdot \prod_{i=t+1}^{n} P(X_i = 0)$$

$$= \vartheta^t \cdot (1-\vartheta)^{n-t} \cdot \prod_{i=1}^{n} \mathbb{1}_{\{0,1\}}(x_i)$$

~~Define~~ set $B_t := \{(x_1,\dots,x_n) \mid x_i \in \{0,1\}, t = \sum_{i=1}^n x_i\}$

We find:

$$P(X=x \mid T=t) = \frac{P(X=x, T=t)}{P(T=t)}$$

$$= \frac{\vartheta^t (1-\vartheta)^t \prod_{i=1}^n 1_{\{0,1\}}(x_i)}{\binom{n}{t} \vartheta^t (1-\vartheta^t) \cdot 1_{\{0,1,\dots,n\}}(t)}$$

$$= \underline{1/\binom{n}{t} \, 1_{B_t}(x)},$$

independent on ϑ and P_ϑ !

$\Rightarrow T(X)$ sufficient for $\vartheta \in (0,1)$.

<u>Observe</u>: Def. of sufficiency theoretically nice but practically unhandy. One has to guess the sufficient statistics $T(X)$.

$\rightarrow$ factorisation theorem helps finding such $T(X)$.

<u>Fact. theorem</u>: Suppose that X is a sample from $P \in \mathcal{P}$ and $\mathcal{P}$ is a family of prob. measures on $(\mathbb{R}^n, \mathcal{B}^n)$ dominated by a σ-finite measure ν.

Then $T(X)$ is sufficient ~~for~~ for $P \in \mathcal{P}$ $\Leftrightarrow$ there are nonnegative Borel functions ~~h (which~~

- h (which does not depend on P) on $(\mathbb{R}^n, \mathcal{B}^n)$
- g_P (which depends on P) on the range of T

such that $$\frac{dP}{d\nu}(x) = g_P(T(x)) h(x).$$

Remember: $P \in \mathcal{P}$ dominated by ν, if $\nu(A) = 0 \Rightarrow P(A) = 0$ for all A measurable. Notation $P \ll \nu$, "P dominated by ν".

Examples :

(1) Let $\mathcal{P}$ be an exponential family (given by (2.4) in book.) and $X(\omega) = \omega$. We can apply factorisation theorem (FT) with $g_\vartheta(t) = \exp\{\eta(\vartheta)^T t - \xi(\vartheta)\}$
$\Rightarrow T(X) = \sum_{i=1}^n X_i$ sufficient for $\vartheta \in (0,1)$

(2) Truncation fam.

Let $\phi(x)$ positive Borel on $(\mathbb{R}, \mathcal{B})$ s.t. $\int_a^b \phi(x)dx < \infty$ for $-\infty < a < b < \infty$.
Let $\vartheta = (a,b)$, $\Theta := \{(a,b) \in \mathbb{R}^2 / a < b\}$ and

$$f_\vartheta(x) = c(\vartheta)\phi(x) 1_{(a,b)}(x), \text{ where } c(\vartheta) = \left(\int_a^b \phi(x)dx\right)^{-1}$$

Then, $\{f_\vartheta / \vartheta \in \Theta\}$ truncated family, is parametric and dom. by Lebesgue.
Let $X_1 \dots X_n$ be iid random variables having PDF f_ϑ. Then, the joint PDF of $X = (X_1, \dots X_n)$ is

$$\prod_{i=1}^n f_\vartheta(x_i) = c(\vartheta)^n 1_{(a,\infty)}(x_{(1)}) 1_{(-\infty,b)}(x_{(n)}) \cdot \prod_{i=1}^n \phi(x_i),$$

where $x_{(i)}$ is the i-th smallest value of $\{x_1, \dots, x_n\}$.

Choose $T(X) = (X_{(1)}, X_{(n)})$ and

- $g_\vartheta(t_1, t_2) = c(\vartheta)^n 1_{(a,\infty)}(t_1) 1_{(-\infty,b)}(t_2)$
- $h(x) = \prod_{i=1}^n \phi(x_i)$,

then T is sufficient for $\vartheta \in \Theta$ by the FT.

(3) (Order statistics)

Let $X = (X_1, \dots, X_n)$ and $X_1, \dots, X_n$ be iid random variables having a distribution $P \in \mathcal{P}$, where $\mathcal{P}$ is the family of distr. on $\mathbb{R}$ having Lebesgue PDF. $X_{(1)}, \dots, X_{(n)}$ as in ex. 2.9.

Joint PDF of X is

$$f(x_1) \cdots f(x_n) = f(x_{(1)}) \cdots f(x_{(n)}).$$

Hence, $T(X) = (X_{(1)}, \dots, X_{(n)})$ is suff. for $P \in \mathcal{P}$.

__Observe__: For a given family $\mathcal{P}$ many sufficient statistics can be found. Let T be sufficient stat. and S another statistic, and ψ measurable, s.t. $T = \psi(S)$, then S is also sufficient.

In general $\sigma(T) \subseteq \sigma(S)$ and hence T provides further reduction.

$\Rightarrow$ Is there a sufficient statistic that provides maximal reduction of data?

(plus other nice properties?)

__Def.__: Let T be a sufficient statistic for $P \in \mathcal{P}$. T is called a __minimal suff. statistic__ if and only if, for any other statistic S sufficient for $P \in \mathcal{P}$, there is a measurable function ψ such that $T = \psi(S)$ a.s. $\mathcal{P}$.