Stochastic Processes

Estimation Theory
Basic concepts

Hamid R. Rabiee

Overview

- Reading Assignment
  - Chapter 6 of C.B. book.
- Further Resources
  - MIT Open Course Ware

Outline

- Basic Definitions
  - Sample, Parameter and Parametric distribution, Statistics
- Sufficient Statistics
  - How to find an SS?
- Minimal Sufficient Statistics
  - How to find an MSS?

Basic Definitions

- let $x_1, x_2, \dots, x_n$ be a Random Sample from X.

  $x_i \sim f(x|\theta)$, and $\mathrm{x_i'}$s are independent.

  $\underline{X} = (x_1, x_2, \dots, x_n)$

- θ: A parameter that describes the distribution, for example θ may be the mean value in a particular distribution.

Statistic

- Any function of the random samples $\underline{X}$ is a statistic: $\begin{cases} variance \\ mean \\ \max value \\ \min value \end{cases}$

  $T: \chi \rightarrow \mathbb{R}, \quad \chi$ *is the* *sample space,* *i.e. set of all* $\underline{X}$

  $t = T(\underline{X})$

  $\tau = \{t : t = T(\underline{X}) \text{ for some } \underline{X} \in \chi \}$

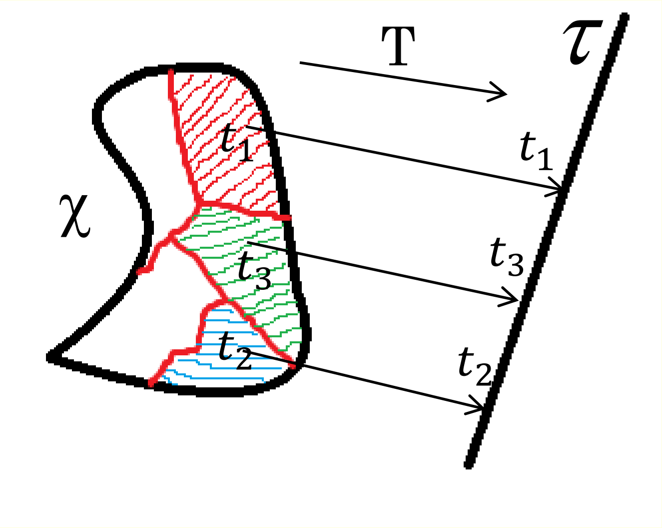
  - Data reduction
  - Partitioning the sample space

  T partitions χ into sets A_t $t \in \tau$.

  $A_t = \{ \underline{X} \in \chi \mid t = T(\underline{X}) \}$

  $T(\underline{X}) = t \equiv \underline{X} \in A_t$

  Example: $T(\underline{X}) = x_1 + x_2 + \cdots + x_n$



Sufficient Statistics

- A *sufficient statistic for a parameter* θ is a statistic, $T(\underline{X})$ that captures all the information about θ contained in the samples.
- Sufficiency Principle:
  - If $T(\underline{X})$ is a sufficient statistic for θ then any inference about θ should depend on the sample $\underline{X}$ only through $T(\underline{X})$.

Sufficient Statistics(Cont'd)

- Definition:
  - If $p(\underline{X} \mid \theta)$ is the joint pdf or pmf of $\underline{X}$ and $q(t|\theta)$ is the pdf or pmf of $T(\underline{X})$, then $T(\underline{X})$ is a sufficient statistic for θ, if for every $\underline{X} \in \chi$ the ratio $\frac{p(\underline{X})}{q(T(\underline{X})|\theta)}$ is constant as a function of θ.

Sufficient Statistics(Cont'd)

- Example 1:
- Let $x_1, \ldots, x_n$ be i.i.d. Bernoulli(θ), $0 < \theta < 1$ is $T(X) = x_1 + \ldots + x_n$ a sufficient statistic?
- Yes. But how?

$$\frac{1}{\binom{n}{\sum x_i}}$$ is independent of θ.

Sufficient Statistics(Cont'd)

- Example 2:
- Let $x_1,\ldots,x_n$ be i.i.d. $N(\mu,\sigma^2)$, σ^2 is known. Is $\bar{x}=(x_1+\ldots+x_n)/n$ a sufficient statistic for μ ?

- Left as Exercise for YOU!

How to find an SS for θ

- Factorization Theorem:
  - Let $f(\underline{X}|\theta)$ denote the joint pdf or pmf of a sample $\underline{X}$, $T(\underline{X})$ is sufficient statistic for θ iff there exists functions $g(t|\theta)$ and $h(\underline{X})$ such that:
    $\forall \underline{X} \in \chi \qquad f(\underline{X}|\theta) = g(T(\underline{X})|\theta)h(\underline{X})$
- So, to find $T(\underline{X})$ factorize $f(\underline{X}|\theta)$ into two parts, $g(T(\underline{X})|\theta)$, which depends on θ, and $h(\underline{X})$ which is independent of θ.

How to find an SS for θ(Cont'd)

- Example 1(continued):
  - Find SS for a Bernoulli distribution

$$f(\underline{X}|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum x_i}(1-\theta)^{1-\sum x_i} =$$

$$g(\sum x_i|\theta)h(\underline{X})\ where: \begin{cases} g(\sum x_i|\theta) = \theta^{\sum x_i}(1-\theta)^{1-\sum x_i} \\ h(\underline{X}) = 1 \end{cases}$$

So: $T(\underline{X}) = \sum x_i$ is a SS for θ.

How to find an SS for θ(Cont'd)

- Example 2:
  - Find SS for a discrete uniform distribution on $\{1, 2, \dots, \theta\}$ [Hint: Use Indicator function]

How to find an SS for θ(Cont'd)

- Example 2:
  - Find SS for a discrete uniform distribution on $\{1, 2, \ldots, \theta\}$ [Hint: Use Indicator function]

$$T(\underline{X}) = \max x_i \qquad i=1,2,\ \ldots,\ n$$

Sufficient Statistics (cont'd)

- Sometimes θ is a vector of parameters.

  In such cases, T(X̲) is usually also vector valued.
- Example: $x_1, \ldots, x_n$ iid $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$

Sufficient Statistics (cont'd)

- Exponential class of distributions:
  - **Theorem**: Let $x_1, \ldots, x_n$ be iid from

$$f(x \mid \theta) = h(x)c(\theta)\exp\{\sum_{i=1}^{k} w_i(\theta)t_i(x)\}$$

then

$$T(x) = \left(\sum_{j=1}^{n} t_1(x_j), \sum_{j=1}^{n} t_2(x_j), \ldots, \sum_{j=1}^{n} t_k(x_j) \right)$$

is a sufficient statistic for θ.

Minimal Sufficient Statistics

- There may be many Sufficient Statistics for a parameter θ. For example $T(\underline{X}) = \underline{X}$ is always an SS.

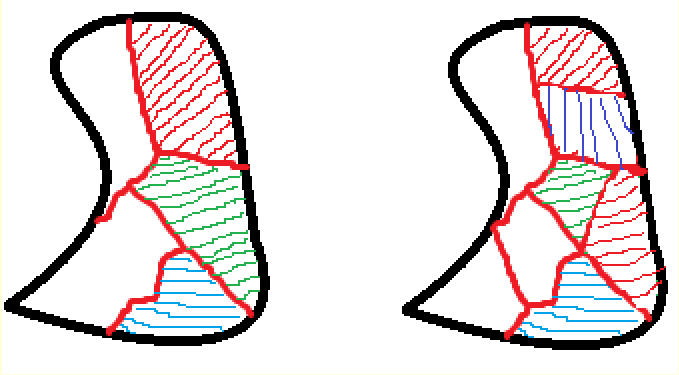
  i.e. $f(\underline{X}|\theta) = f(\underline{X}|\theta)h(\underline{X}), \quad where \quad h(\underline{X}) = 1$

- Also any one-to-one function of an SS is an SS.
- Which SS is the best?

Minimal Sufficient Statistics(Cont'd)

- Goal: Data reduction while preserving info. about θ.
- A *sufficient statistic* $T(\underline{X})$ is called a minimal sufficient statistic, if for any other SS $T'(\underline{X})$, $T(\underline{X})$ is a function of $T'(\underline{X})$.
- So MSS $\equiv$ Maximum data reduction
- MSS gives the coarsest partitioning

MSS SS but not MSS



Minimal Sufficient Statistics(Cont'd)

- Example 4:

  $x_1, x_2, \ldots, x_n \sim \mathrm{N}(\mu, \sigma^2), \sigma^2$ *is known,* Are i.i.d. samples

- Factorization Theorem: $\begin{cases} \overline{X} & \text{is an } SS. \\ (\overline{X}, s^2) & \text{is also an } SS. \end{cases}$

- Clearly, $\overline{X}$ achieves higher data reduction and is thus better.

- If σ^2 where unknown, then $\overline{X}$ is not an SS. And $(\overline{X}, \mathrm{s}^2)$ contains more info about (μ, σ^2).

How to find an MSS?

- Theorem [Lehmann, Sheffe 1950]:
  - Let $f(\underline{X}|\theta)$ be the pdf or pmf of a sample $\underline{X}$. Suppose $T(\underline{X})$ exists such that: $\forall \underline{X}, \underline{Y} \in \chi$, $\frac{f(\underline{X}|\theta)}{f(\underline{Y}|\theta)}$ is constant as a function of θ iff $T(\underline{X}) = T(\underline{Y})$. Then $T(\underline{X})$ is a Minimal Sufficient Statistic.
  - If $[T(\underline{X}) = T(\underline{Y}) \rightarrow \frac{f(\underline{X}|\theta)}{f(\underline{Y}|\theta)}]$ is a constant, then $T(\underline{X})$ is an SS.

How to find an MSS?

- Example 5:

$$x_1, x_2, \dots, x_n \sim U(\theta, \theta + 1)$$

Find an MSS for X.

- does the dimension of the MSS equal the dimension of the parameter?

How to find an MSS?

- Example 5:

$$x_1, x_2, \dots, x_n \sim U(\theta, \theta+1)$$

  Find an MSS for X.

- does the dimension of the MSS equal the dimension of the parameter?

$T(X) = (\min x_i, \max x_i)$ is an MSS

IS it unique??

So any one-to-one function of an MSS is also MSS.