STK4011 and STK9011
Autumn 2016

Sufficient statistics

Covers sections 6.1, 6.2.1 and 6.2.2

Ørnulf Borgan
Department of Mathematics
University of Oslo

Data reduction

Assume that we have random variables $X_1, X_2, ..., X_n$ with joint pmf or pdf $f(x_1,..., x_n \mid \theta)$

We write $\mathbf{X}$ for the vector of the random variables $X_1, X_2, ..., X_n$ and $\mathbf{x}$ to denote their values $x_1, x_2, ..., x_n$

A statistic $T(\mathbf{X}) = T(X_1, X_2, ..., X_n)$, which may be a vector, defines a form of data reduction or data summary

$T(\mathbf{X})$ defines a partition of the sample space (the space $\mathcal{X}$ of possible values of $\mathbf{X}$)

More specifically, let

$$\mathcal{T} = \{t : t = T(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathcal{X}\}$$

be the space of possible values for the statistic $T(\mathbf{X})$

Then $T(\mathbf{x})$ partitions the sample space into sets A_t, $t \in \mathcal{T}$ of the form $A_t = \{\mathbf{x} \in \mathcal{X} : T(\mathbf{x}) = t\}$

We will make inference (e.g. estimate or test a hypothesis) about the unknown population parameter θ (which may be a vector)

Note that an experimenter who base the inference on the observed value $T(\mathbf{x})$ of the statistic, rather than the entire sample $\mathbf{x}$, will treat the samples $\mathbf{x}$ and $\mathbf{y}$ equally if $T(\mathbf{x}) = T(\mathbf{y})$

Sufficient statistics

A sufficient statistic $T(\mathbf{X})$ is a statistic that (in a certain sense) contains all the information about θ contained in the sample $\mathbf{X} = (X_1, X_2, ..., X_n)$

The sufficiency principle states that if $T(\mathbf{X})$ is a sufficient statistic, then any inference about θ should depend on the sample $\mathbf{X}$ only through the value $T(\mathbf{X})$

If we follow the sufficiency principle, we may disregard the sample $\mathbf{X}$ once we know $T(\mathbf{X})$

Note that this assumes that the model is correctly specified, so it does not apply to model checking

Definition 6.2.1

A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ

If $T(\mathbf{X})$ has a continuous distribution, then $P_\theta(T(\mathbf{X})=t)=0$ for all values of t, and the standard definition of a conditional probability does not apply

Therefore we will do our calculations for the discrete case and only point out that analogous results are true for the continuous case (but require more sophisticated arguments)

Theorem 6.2.2

If $p(\mathbf{x}\mid\theta)$ is the joint pmf or pdf of $\mathbf{X}$ and $q(t\mid\theta)$ is the pmf or pdf of $T(\mathbf{X})$, then $T(\mathbf{X})$ is a sufficient statistic for θ if, for every $\mathbf{x}$ in the sample space, the ratio $p(\mathbf{x}\mid\theta)/q(T(\mathbf{x})\mid\theta)$ does not depend on θ

Proof:

Note that $P_\theta(\mathbf{X}=\mathbf{x}\mid T(\mathbf{X})=t)=0$ when $T(\mathbf{x})\neq t$

So we need to show that $P_\theta(\mathbf{X}=\mathbf{x}\mid T(\mathbf{X})=T(\mathbf{x}))$ does not depend on θ. Now

$$P_\theta(\mathbf{X}=\mathbf{x}\mid T(\mathbf{X})=T(\mathbf{x}))=\frac{P_\theta(\mathbf{X}=\mathbf{x}\text{ and }T(\mathbf{X})=T(\mathbf{x}))}{P_\theta(T(\mathbf{X})=T(\mathbf{x}))}$$

$$=\frac{P_\theta(\mathbf{X}=\mathbf{x})}{P_\theta(T(\mathbf{X})=T(\mathbf{x}))}=\frac{p(\mathbf{x}\mid\theta)}{q(T(\mathbf{x})\mid\theta)}$$

which does not depend on θ

Example 6.2.3 – binomial sufficient statistic

Let $X_1, X_2, \ldots, X_n$ be iid Bernoulli random variables with success probability $\theta\in(0,1)$

Then $T(\mathbf{X})=X_1+\ldots+X_n$ is a sufficient statistic for θ

Example 6.2.4 – normal sufficient statistic

Let $X_1, X_2, \ldots, X_n$ be iid $n(\mu,\sigma^2)$ random variables, where σ^2 is known

Then $T(\mathbf{X})=\bar{X}=(X_1+\ldots+X_n)/n$ is a sufficient statistic for μ

Example 6.2.5 – sufficient order statistic

Let $X_1, X_2, \ldots, X_n$ be iid with pdf $f(x)$, where f is not given a parametric specification, so here f itself is the unknown (infinite dimensional) parameter

Then the order statistics $T(\mathbf{X})=(X_{(1)}, \ldots, X_{(n)})$ are sufficient statistics for f

In order to use Theorem 6.2.2, we have to guess that a statistic $T(\mathbf{X})$ is sufficient and to find the pmf or pdf of $T(\mathbf{X})$

The following theorem avoids this:

Theorem 6.2.6 (factorization theorem)
Let $f(\mathbf{x}\,|\,\theta)$ denote the joint pmf or pdf of $\mathbf{X}$.
A statistic $T(\mathbf{X})$ is sufficient for θ if and only if there exists functions $g(t\,|\,\theta)$ and $h(\mathbf{x})$ such that

$$f(\mathbf{x}\,|\,\theta) = g(T(\mathbf{x})\,|\,\theta)\,h(\mathbf{x})$$

for all parameter values θ and all $\mathbf{x} \in \mathcal{X}$

Example – binomial sufficient statistic

Let $X_1, X_2, \ldots, X_n$ be iid Bernoulli random variables with success probability $\theta \in (0,1)$

Then the joint pmf of $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ may be written (for $x_i \in \{0,1\}$)

$$f(\mathbf{x}\,|\,\theta) = P_\theta(X_1 = x_1, \ldots, X_n = x_n)$$

$$= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Thus $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is sufficient

Example 6.2.8 – uniform sufficient statistic

Let $X_1, X_2, \ldots, X_n$ be iid random variables with pmf

$$f(x\,|\,\theta) = \begin{cases} 1/\theta & x = 1, 2, \ldots, \theta \\ 0 & \text{otherwise} \end{cases}$$

The joint pmf of $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ may be written

$$f(\mathbf{x}\,|\,\theta) = \begin{cases} 1/\theta^n & \text{if } x_i \in \{1, 2, \ldots, \theta\} \text{ for } i = 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

The restriction

$$x_i \in \{1, 2, \ldots, \theta\} \text{ for } i = 1, \ldots, n$$

may be expressed as

$$x_i \in \{1, 2, \ldots\} \text{ for } i = 1, \ldots, n \text{ and } \max_i x_i \le \theta$$

In order to use the factorization theorem we define $T(\mathbf{x}) = \max_i x_i$ and introduce

$$h(\mathbf{x}) = \begin{cases} 1 & \text{if } x_i \in \{1, 2, \ldots\} \text{ for } i = 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

and

$$g(t\,|\,\theta) = \begin{cases} 1/\theta^n & \text{for } t \le \theta \\ 0 & \text{otherwise} \end{cases}$$

Then

$$f(\mathbf{x}\,|\,\theta) = h(\mathbf{x})\,g(T(\mathbf{x})\,|\,\theta)$$

so $T(\mathbf{X}) = \max_i X_i$ is sufficient

The factorization theorem also holds for vector-valued statistics $T(\mathbf{X}) = (T_1(\mathbf{X}), ..., T_r(\mathbf{X}))$ and vector-valued parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_s)$

Example 6.2.9 – normal sufficient statistic

Let $X_1, X_2, ..., X_n$ be iid $n(\mu, \sigma^2)$ random variables, where both parameters are unknown

Using the factorization theorem one may show that $T(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X})) = (\bar{X}, S^2)$ is a sufficient statistic for (μ, σ^2)

For a given situation, there are in general more than one sufficient statistic

E.g. for a sample from $n(\mu, \sigma^2)$ the following are some sufficient statistics:

- $T(\mathbf{X}) = (X_1, ..., X_n)$
- $T(\mathbf{X}) = (X_{(1)}, ..., X_{(n)})$
- $T(\mathbf{X}) = (\bar{X}, S^2)$
- $T(\mathbf{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2 \right)$

The last to sufficient statistics on the previous slide is an example of the following general result:

Let $T(\mathbf{X})$ be a sufficient statistic and define $T^*(\mathbf{x}) = r(T(\mathbf{x}))$, where $r(t);\ t \in \mathcal{T}$; is a one-to-one function. Then $T^*(\mathbf{X})$ is a sufficient statistic

The result is an easy consequence of the factorization theorem

Minimal sufficient statistics

Definition 6.2.11
A sufficient statistic $T(\mathbf{X})$ is called a minimal sufficient statistic if, for any other sufficient statistic $T'(\mathbf{X})$, $T(\mathbf{X})$ is a function of $T'(\mathbf{X})$

If $T(\mathbf{x})$ partitions the sample space into sets $A_t\ , t \in \mathcal{T}$ defined by $A_t = \{\mathbf{x} \in \mathcal{X} : T(\mathbf{x}) = t\}$ and $T'(\mathbf{x})$ similarly partitions the sample space into sets $B_{t'}\ , t' \in \mathcal{T}'$, then the definition states that every $B_{t'}$ is a subset of some A_t

The partition associated with a minimal sufficient statistic is the coarsest possible partition for a sufficient statistic

Note that any one-to-one transformation of a minimal sufficient statistic is itself a minimal sufficient statistics (and the two give the same partition of the sample space)

It is impractical to use Definition 6.2.11 to find a minimal sufficient statistic

The following theorem is easier to use:

Theorem 6.2.13

Let $f(\mathbf{x}\,|\,\theta)$ denote the joint pmf or pdf of $\mathbf{X}$. Suppose there exist a function $T(\mathbf{x})$ such that for every $\mathbf{x},\mathbf{y} \in \mathcal{X}$, the ratio $f(\mathbf{x}\,|\,\theta)/f(\mathbf{y}\,|\,\theta)$ does not depend on θ if and only if $T(\mathbf{x}) = T(\mathbf{y})$.
Then $T(\mathbf{X})$ is a minimal sufficient statistic for θ

Example 6.2.14 – normal minimal sufficient statistic

Let $X_1, X_2, \ldots, X_n$ be iid $n(\mu,\sigma^2)$ random variables, where both parameters are uknown

Using Theorem 6.2.13 one may show that $(\bar{X}, S^2)$ is a minimal sufficient statistic for (μ,σ^2)

Example 6.2.15 – uniform minimal sufficient statistic

Let $X_1, X_2, \ldots, X_n$ be iid and $\text{uniform}(\theta, \theta+1)$

Using Theorem 6.2.13 one may show that $(X_{(1)}, X_{(n)})$ is a minimal sufficient statistic for θ