Foundations of Statistical Inference

Julien Berestycki

Department of Statistics
University of Oxford

MT 2016

Lecture 2 - Sufficiency, Factorization Theorem, Minimal sufficiency

Sufficient statistics

Let $X_1, \ldots, X_n$ be a random sample from $f(x;\theta)$.

Sufficient statistics

Let $X_1, \dots, X_n$ be a random sample from $f(x;\theta)$.

Definition (Sufficiency)

A statistic $T(X_1, \dots, X_n)$ is a function of the data that does not depend on unknown parameters.

Sufficient statistics

Let $X_1, \ldots, X_n$ be a random sample from $f(x; \theta)$.

Definition (Sufficiency)

A statistic $T(X_1, \ldots, X_n)$ is a function of the data that does not depend on unknown parameters.

A statistic $T(X_1, \ldots, X_n)$ is said to be sufficient for θ if the conditional distribution of $X_1, \ldots, X_n$, given T, does not depend on θ. That is,

$$f(x \mid t, \theta) = f(x \mid t)$$

Sufficient statistics

Let $X_1, \ldots, X_n$ be a random sample from $f(x;\theta)$.

Definition (Sufficiency)

A statistic $T(X_1, \ldots, X_n)$ is a function of the data that does not depend on unknown parameters.

A statistic $T(X_1, \ldots, X_n)$ is said to be sufficient for θ if the conditional distribution of $X_1, \ldots, X_n$, given T, does not depend on θ. That is,

$$f(x \mid t, \theta) = f(x \mid t)$$

Comment The definition says that a sufficient statistic T contains all the information there is in the sample about θ.

Definition (Sufficiency)

A statistic $T(X_1, \ldots, X_n)$ is a function of the data that does not depend on unknown parameters.

Definition (Sufficiency)

A statistic $T(X_1, \ldots, X_n)$ is a function of the data that does not depend on unknown parameters.

A statistic $T(X_1, \ldots, X_n)$ is said to be sufficient for θ if the conditional distribution of $X_1, \ldots, X_n$, given T, does not depend on θ. That is,

$$f(x \mid t, \theta) = f(x \mid t)$$

What does this even mean?

Definition (Sufficiency)

A statistic $T(X_1, \ldots, X_n)$ is a function of the data that does not depend on unknown parameters.

A statistic $T(X_1, \ldots, X_n)$ is said to be sufficient for θ if the conditional distribution of $X_1, \ldots, X_n$, given T, does not depend on θ. That is,

$$f(x \mid t, \theta) = f(x \mid t)$$

What does this even mean?
It means that for any function g the map

$$\theta \mapsto \mathbb{E}_\theta[g(X)|T = t]$$

is constant.

Example 7

n independent trials where the probability of success is p.

Example 7

n independent trials where the probability of success is p.
Let $X_1, \ldots, X_n$ be indicator variables which are 1 or 0 depending if the trial is a success or failure.

Example 7

n independent trials where the probability of success is p.
Let $X_1, \ldots, X_n$ be indicator variables which are 1 or 0 depending if the trial is a success or failure.
Let $T = \sum_{i=1}^{n} X_i$.

Example 7

n independent trials where the probability of success is p.
Let $X_1, \ldots, X_n$ be indicator variables which are 1 or 0 depending if the trial is a success or failure.
Let $T = \sum_{i=1}^n X_i$. The conditional distribution of $X_1, \ldots, X_n$ given $T = t$ is

$$g(x_1, \ldots, x_n \mid t, p) = \frac{f(x_1, \ldots, x_n, t \mid p)}{h(t \mid p)}$$

Example 7

n independent trials where the probability of success is p.
Let $X_1, \ldots, X_n$ be indicator variables which are 1 or 0 depending if the trial is a success or failure.
Let $T = \sum_{i=1}^{n} X_i$. The conditional distribution of $X_1, \ldots, X_n$ given $T = t$ is

$$g(x_1, \ldots, x_n \mid t, p) = \frac{f(x_1, \ldots, x_n, t \mid p)}{h(t \mid p)} \quad = \quad \frac{\prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}}{\binom{n}{t} p^t (1-p)^{n-t}}$$

Example 7

n independent trials where the probability of success is p.
Let $X_1, \ldots, X_n$ be indicator variables which are 1 or 0 depending if the trial is a success or failure.
Let $T = \sum_{i=1}^n X_i$. The conditional distribution of $X_1, \ldots, X_n$ given $T = t$ is

$$
\begin{aligned}
g(x_1, \ldots, x_n \mid t, p) = \frac{f(x_1, \ldots, x_n, t \mid p)}{h(t \mid p)} &= \frac{\prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} \\
&= \frac{p^t(1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}}
\end{aligned}
$$

Example 7

n independent trials where the probability of success is p.
Let $X_1, \ldots, X_n$ be indicator variables which are 1 or 0 depending if the trial is a success or failure.
Let $T = \sum_{i=1}^{n} X_i$. The conditional distribution of $X_1, \ldots, X_n$ given $T = t$ is

$$
\begin{aligned}
g(x_1, \ldots, x_n \mid t, p) = \frac{f(x_1, \ldots, x_n, t \mid p)}{h(t \mid p)} &= \frac{\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} \\
&= \frac{p^t (1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} \\
&= \binom{n}{t}^{-1},
\end{aligned}
$$

not depending on p, so T is sufficient for p.

Example 7

n independent trials where the probability of success is p.
Let $X_1, \ldots, X_n$ be indicator variables which are 1 or 0 depending if the trial is a success or failure.
Let $T = \sum_{i=1}^n X_i$. The conditional distribution of $X_1, \ldots, X_n$ given $T = t$ is

$$\begin{aligned} g(x_1, \ldots, x_n \mid t, p) = \frac{f(x_1, \ldots, x_n, t \mid p)}{h(t \mid p)} &= \frac{\prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} \\ &= \frac{p^t(1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} \\ &= \binom{n}{t}^{-1}, \end{aligned}$$

not depending on p, so T is sufficient for p.
Comment Makes sense, since no information in the order.

Theorem 1 : Factorization Criterion

Theorem

$T(X_1, \ldots, X_n)$ is a sufficient statistic for θ if and only if there exist two non-negative functions K_1, K_2 such that the likelihood function $L(\theta; x)$ can be written

$$L(\theta; x) = K_1[t(x_1, \ldots, x_n); \theta]K_2[x_1, \ldots, x_n] = K_1[t; \theta]K_2[x],$$

where K_1 depends only on the sample through T, and K_2 does not depend on θ.

Proof - for discrete random variables

1. Assume that T is sufficient, then the distribution of the sample is

$$L(\theta; x) = f(x|\theta) =$$

Proof - for discrete random variables

1. Assume that T is sufficient, then the distribution of the sample is

$$L(\theta; x) = f(x|\theta) = f(x, t|\theta) =$$

Proof - for discrete random variables

1. Assume that T is sufficient, then the distribution of the sample is

$$L(\theta; x) = f(x|\theta) = f(x, t|\theta) = f(x \mid t, \theta)h(t \mid \theta)$$

Proof - for discrete random variables

1. Assume that T is sufficient, then the distribution of the sample is

$$L(\theta; x) = f(x|\theta) = f(x, t|\theta) = f(x \mid t, \theta)h(t \mid \theta)$$

T is sufficient which implies

$$f(x \mid t, \theta) = f(x \mid t)$$

Proof - for discrete random variables

1. Assume that T is sufficient, then the distribution of the sample is

$$L(\theta; x) = f(x|\theta) = f(x, t|\theta) = f(x \mid t, \theta)h(t \mid \theta)$$

T is sufficient which implies

$$f(x \mid t, \theta) = f(x \mid t)$$

$h(t \mid \theta)$ depends on x through $t(x)$ only so

$$L(\theta; x) = f(x \mid t)h(t \mid \theta)$$

Proof - for discrete random variables

1. Assume that T is sufficient, then the distribution of the sample is

$$L(\theta; x) = f(x|\theta) = f(x, t|\theta) = f(x \mid t, \theta)h(t \mid \theta)$$

T is sufficient which implies

$$f(x \mid t, \theta) = f(x \mid t)$$

$h(t \mid \theta)$ depends on x through $t(x)$ only so

$$L(\theta; x) = f(x \mid t)h(t \mid \theta)$$

We set $L(\theta; x) = K_1 K_2$, where $K_1 \equiv h$, $K_2 \equiv f$.

2. Suppose $L(\theta; x) = f(x \mid \theta) = K_1[t; \theta]K_2[x]$.

2. Suppose $L(\theta; x) = f(x \mid \theta) = K_1[t;\theta]K_2[x]$.Then

$$h(t \mid \theta) = \sum_{\{x:T(x)=t\}} f(x, t \mid \theta) \quad = \quad \sum_{\{x:T(x)=t\}} L(\theta; x)$$

2. Suppose $L(\theta; x) = f(x \mid \theta) = K_1[t;\theta]K_2[x]$.Then

$$h(t \mid \theta) = \sum_{\{x:T(x)=t\}} f(x, t \mid \theta) = \sum_{\{x:T(x)=t\}} L(\theta; x) = K_1[t;\theta] \sum_{\{x:T(x)=t\}} K_2(x).$$

2. Suppose $L(\theta; x) = f(x \mid \theta) = K_1[t; \theta]K_2[x]$.Then

$$\begin{aligned} h(t \mid \theta) = \sum_{\{x:T(x)=t\}} f(x, t \mid \theta) &= \sum_{\{x:T(x)=t\}} L(\theta; x) \\ &= K_1[t; \theta] \sum_{\{x:T(x)=t\}} K_2(x). \end{aligned}$$

Thus

$$f(x \mid t, \theta) = \frac{f(x, t \mid \theta)}{h(t \mid \theta)} = \frac{L(\theta; x)}{h(t \mid \theta)} = \frac{K_2[x]}{\sum_{\{x:T(x)=t\}} K_2(x)},$$

not depending on θ. (K_1 cancels out in numerator and denominator.)

Minimal sufficiency

How much can we reduce the data without loosing information? Is there a minimal sufficient statistic?

Minimal sufficiency

How much can we reduce the data without loosing information? Is there a minimal sufficient statistic?

Example 7 (cont.) Consider $n = 3$ Bernoulli trials

1. $T_1(X) = (X_1, X_2, X_3)$ (the individual sequences of trials)

Minimal sufficiency

How much can we reduce the data without loosing information? Is there a minimal sufficient statistic?

Example 7 (cont.) Consider $n = 3$ Bernoulli trials

1. $T_1(X) = (X_1, X_2, X_3)$ (the individual sequences of trials)
2. $T_2(X) = (X_1, \sum_{i=1}^{3} X_i)$ (the 1st random variable and the total sum).

Minimal sufficiency

How much can we reduce the data without loosing information? Is there a minimal sufficient statistic?

Example 7 (cont.) Consider $n = 3$ Bernoulli trials

1. $T_1(X) = (X_1, X_2, X_3)$ (the individual sequences of trials)
2. $T_2(X) = (X_1, \sum_{i=1}^{3} X_i)$ (the 1st random variable and the total sum).
3. $T_3(X) = \sum_{i=1}^{3} X_i$ (the total sum)

Minimal sufficiency

How much can we reduce the data without loosing information? Is there a minimal sufficient statistic?

Example 7 (cont.) Consider $n = 3$ Bernoulli trials

1. $T_1(X) = (X_1, X_2, X_3)$ (the individual sequences of trials)
2. $T_2(X) = (X_1, \sum_{i=1}^{3} X_i)$ (the 1st random variable and the total sum).
3. $T_3(X) = \sum_{i=1}^{3} X_i$ (the total sum)
4. $T_4(X) = I(T_3(X) = 0)$ (I is indicator function; Exercise Prove T_4 not sufficient)

Minimal sufficiency

How much can we reduce the data without loosing information? Is there a minimal sufficient statistic?

Example 7 (cont.) Consider $n = 3$ Bernoulli trials

1. $T_1(X) = (X_1, X_2, X_3)$ (the individual sequences of trials)
2. $T_2(X) = (X_1, \sum_{i=1}^3 X_i)$ (the 1st random variable and the total sum).
3. $T_3(X) = \sum_{i=1}^3 X_i$ (the total sum)
4. $T_4(X) = I(T_3(X) = 0)$ (I is indicator function; Exercise Prove T_4 not sufficient)

Definition (Minimality)

A statistic is minimal sufficient if it can be expressed as a function of every other sufficient statistic.

Example 7 (cont.) : Minimal sufficiency

n Bernoulli trials with $T = \sum_{i=1}^{n} X_i$.

Example 7 (cont.) : Minimal sufficiency

n Bernoulli trials with $T = \sum_{i=1}^{n} X_i$. Suppose T above is not minimal sufficient but another statistic U is MS.

Example 7 (cont.) : Minimal sufficiency

n Bernoulli trials with $T = \sum_{i=1}^{n} X_i$. Suppose T above is not minimal sufficient but another statistic U is MS.Then U can be given as a function of T

Example 7 (cont.) : Minimal sufficiency

n Bernoulli trials with $T = \sum_{i=1}^{n} X_i$. Suppose T above is not minimal sufficient but another statistic U is MS.Then U can be given as a function of T(and not *vis versa* or T is MS) and there exist $t_1 \neq t_2$ values of T so that $U(t_1) = U(t_2)$

Example 7 (cont.) : Minimal sufficiency

n Bernoulli trials with $T = \sum_{i=1}^{n} X_i$. Suppose T above is not minimal sufficient but another statistic U is MS.Then U can be given as a function of T(and not *vis versa* or T is MS) and there exist $t_1 \neq t_2$ values of T so that $U(t_1) = U(t_2)$(ie $T \to U$ is many to one so $U \to T$ is not a function, and we assume for the moment no other t make $U(t) = U(t_1)$).

Example 7 (cont.) : Minimal sufficiency

n Bernoulli trials with $T = \sum_{i=1}^{n} X_i$. Suppose T above is not minimal sufficient but another statistic U is MS.Then U can be given as a function of T(and not *vis versa* or T is MS) and there exist $t_1 \neq t_2$ values of T so that $U(t_1) = U(t_2)$(ie $T \to U$ is many to one so $U \to T$ is not a function, and we assume for the moment no other t make $U(t) = U(t_1)$).The event $U = u$ is the event $T \in \{t_1, t_2\}$. Let $x_1, ...x_n$ contain t_1 successes.

Example 7 (cont.) : Minimal sufficiency

n Bernoulli trials with $T = \sum_{i=1}^{n} X_i$. Suppose T above is not minimal sufficient but another statistic U is MS.Then U can be given as a function of T(and not *vis versa* or T is MS) and there exist $t_1 \neq t_2$ values of T so that $U(t_1) = U(t_2)$(ie $T \to U$ is many to one so $U \to T$ is not a function, and we assume for the moment no other t make $U(t) = U(t_1)$).The event $U = u$ is the event $T \in \{t_1, t_2\}$. Let $x_1, ...x_n$ contain t_1 successes. Then

$$g(x_1, \ldots, x_n | u, p) \quad = \quad g(x_1, \ldots, x_n | t_1, p) P(t_1 | u, p)$$

Example 7 (cont.) : Minimal sufficiency

n Bernoulli trials with $T = \sum_{i=1}^{n} X_i$. Suppose T above is not minimal sufficient but another statistic U is MS.Then U can be given as a function of T(and not *vis versa* or T is MS) and there exist $t_1 \neq t_2$ values of T so that $U(t_1) = U(t_2)$(ie $T \to U$ is many to one so $U \to T$ is not a function, and we assume for the moment no other t make $U(t) = U(t_1)$).The event $U = u$ is the event $T \in \{t_1, t_2\}$. Let $x_1, ...x_n$ contain t_1 successes. Then

$$
\begin{aligned}
g(x_1, \ldots, x_n | u, p) &= g(x_1, \ldots, x_n | t_1, p) P(t_1 | u, p) \\
&= g(x_1, \ldots, x_n | t_1) P(T = t_1 | T \in \{t_1, t_2\}, p)
\end{aligned}
$$

Example 7 (cont.) : Minimal sufficiency

n Bernoulli trials with $T = \sum_{i=1}^{n} X_i$. Suppose T above is not minimal sufficient but another statistic U is MS.Then U can be given as a function of T(and not *vis versa* or T is MS) and there exist $t_1 \neq t_2$ values of T so that $U(t_1) = U(t_2)$(ie $T \to U$ is many to one so $U \to T$ is not a function, and we assume for the moment no other t make $U(t) = U(t_1)$).The event $U = u$ is the event $T \in \{t_1, t_2\}$. Let $x_1, ...x_n$ contain t_1 successes. Then

$$\begin{aligned} g(x_1, \ldots, x_n | u, p) &= g(x_1, \ldots, x_n | t_1, p) P(t_1 | u, p) \\ &= g(x_1, \ldots, x_n | t_1) P(T = t_1 | T \in \{t_1, t_2\}, p) \\ &= \frac{1}{\binom{n}{t_1}} \frac{\binom{n}{t_1} p^{t_1} (1-p)^{n-t_1}}{\binom{n}{t_1} p^{t_1} (1-p)^{n-t_1} + \binom{n}{t_2} p^{t_2} (1-p)^{n-t_2}} \end{aligned}$$

which depends on p, so U is not sufficient, a contradiction, and hence T must be MS (similar reasoning for multiple t_i).

Minimal sufficiency and partitions of the sample space

- Intuitively, a minimal sufficient statistic most efficiently captures all possible information about the parameter θ.

Minimal sufficiency and partitions of the sample space

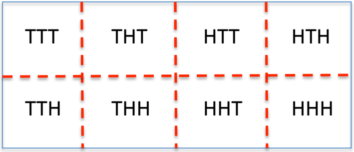
- Intuitively, a minimal sufficient statistic most efficiently captures all possible information about the parameter θ.
- Any statistic $T(X)$ partitions the sample space into subsets and in each subset $T(X)$ has constant value.

Minimal sufficiency and partitions of the sample space

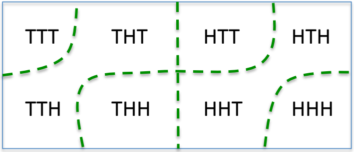
- Intuitively, a minimal sufficient statistic most efficiently captures all possible information about the parameter θ.
- Any statistic $T(X)$ partitions the sample space into subsets and in each subset $T(X)$ has constant value.
- Minimal sufficient statistics correspond to the coarsest possible partition of the sample space.

Minimal sufficiency and partitions of the sample space

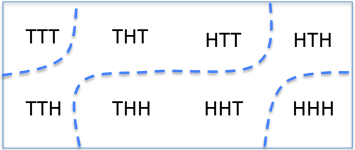
- Intuitively, a minimal sufficient statistic most efficiently captures all possible information about the parameter θ.
- Any statistic $T(X)$ partitions the sample space into subsets and in each subset $T(X)$ has constant value.
- Minimal sufficient statistics correspond to the coarsest possible partition of the sample space.
- In the example of $n = 3$ Bernoulli trials consider the following 4 statistics and the partitions they induce.



$$T_1(X) = (X_1, X_2, X_3)$$



$$T_2(X) = \left(X_1, \sum_{i=1}^{3} X_i \right)$$



$$T_3(X) = \sum_{i=1}^{3} X_i$$

$$T_4(X) = I\left(T_3(X) = 0\right)$$

Lemma 1 : Lehmann-Scheffé partitions

Theorem

Consider the partition of the sample space defined by putting x and y into the same class of the partition if and only if

$$L(\theta; y)/L(\theta; x) = f(y \mid \theta)/f(x \mid \theta) = m(x, y).$$

Then any statistic corresponding to this partition is minimal sufficient.

Lemma 1 : Lehmann-Scheffé partitions

Theorem

Consider the partition of the sample space defined by putting x and y into the same class of the partition if and only if

$$L(\theta; y)/L(\theta; x) = f(y \mid \theta)/f(x \mid \theta) = m(x, y).$$

Then any statistic corresponding to this partition is minimal sufficient.

Comment This Lemma tells us how to define partitions that correspond to minimal sufficient statistics. It says that ratios of likelihoods of two values x and y in the same partition (and hence same statistic value) should not depend on θ.

Proof (for discrete RVs)

1. Sufficiency.
Suppose T is such a statistic

$$g(x|t,\theta) = \frac{f(x \mid \theta)}{f(t \mid \theta)} = \frac{f(x \mid \theta)}{\sum_{y \in \tau} f(y \mid \theta)}, \quad \tau = \{y : T(y) = t\}$$

Proof (for discrete RVs)

1. Sufficiency.
Suppose T is such a statistic

$$g(x|t,\theta) = \frac{f(x \mid \theta)}{f(t \mid \theta)} = \frac{f(x \mid \theta)}{\sum_{y \in \tau} f(y \mid \theta)}, \quad \tau = \{y : T(y) = t\}$$
$$= \frac{f(x \mid \theta)}{\sum_{y \in \tau} f(x \mid \theta) m(x,y)}$$

Proof (for discrete RVs)

1. Sufficiency.
Suppose T is such a statistic

$$
\begin{aligned}
g(x|t,\theta) = \frac{f(x \mid \theta)}{f(t \mid \theta)} &= \frac{f(x \mid \theta)}{\sum_{y \in \tau} f(y \mid \theta)}, \quad \tau = \{y : T(y) = t\} \\
&= \frac{f(x \mid \theta)}{\sum_{y \in \tau} f(x \mid \theta) m(x,y)} \\
&= \left[\sum_{y \in \tau} m(x,y) \right]^{-1}
\end{aligned}
$$

which does not depend on θ. Hence the partition D is sufficient.

2. Minimal sufficiency.
Now suppose U is any other sufficient statistic and that $U(x) = U(y)$ for some pair of values (x, y).

2. Minimal sufficiency.
Now suppose U is any other sufficient statistic and that $U(x) = U(y)$ for some pair of values (x, y). If we can show that $U(x) = U(y)$ implies $T(x) = T(y)$, then the Lehmann-Scheffé partition induced by T includes the partition based on any other sufficient statistic.

2. Minimal sufficiency.
Now suppose U is any other sufficient statistic and that $U(x) = U(y)$ for some pair of values (x, y). If we can show that $U(x) = U(y)$ implies $T(x) = T(y)$, then the Lehmann-Scheffé partition induced by T includes the partition based on any other sufficient statistic.In other words, T is a function of every other sufficient statistic, and so must be minimal sufficient.

2. Minimal sufficiency.
Now suppose U is any other sufficient statistic and that $U(x) = U(y)$ for some pair of values (x, y). If we can show that $U(x) = U(y)$ implies $T(x) = T(y)$, then the Lehmann-Scheffé partition induced by T includes the partition based on any other sufficient statistic.In other words, T is a function of every other sufficient statistic, and so must be minimal sufficient.
Since U is sufficient we have

$$\frac{L(\theta; y)}{L(\theta; x)} = \frac{K_1[u(y); \theta]K_2[y]}{K_1[u(x); \theta]K_2[x]} =$$

2. Minimal sufficiency.

Now suppose U is any other sufficient statistic and that $U(x) = U(y)$ for some pair of values (x, y). If we can show that $U(x) = U(y)$ implies $T(x) = T(y)$, then the Lehmann-Scheffé partition induced by T includes the partition based on any other sufficient statistic.In other words, T is a function of every other sufficient statistic, and so must be minimal sufficient.

Since U is sufficient we have

$$\frac{L(\theta; y)}{L(\theta; x)} = \frac{K_1[u(y); \theta]K_2[y]}{K_1[u(x); \theta]K_2[x]} = \frac{K_2[y]}{K_2[x]}$$

which does not depend on θ.

2. Minimal sufficiency.
Now suppose U is any other sufficient statistic and that $U(x) = U(y)$ for some pair of values (x, y). If we can show that $U(x) = U(y)$ implies $T(x) = T(y)$, then the Lehmann-Scheffé partition induced by T includes the partition based on any other sufficient statistic.In other words, T is a function of every other sufficient statistic, and so must be minimal sufficient.
Since U is sufficient we have

$$\frac{L(\theta; y)}{L(\theta; x)} = \frac{K_1[u(y); \theta]K_2[y]}{K_1[u(x); \theta]K_2[x]} = \frac{K_2[y]}{K_2[x]}$$

which does not depend on θ. So the statistic U produces a partition at least as fine as that induced by T, and the result is proved.

Sufficiency in an exponential family

Theorem

For a sample $X_1, \ldots, X_n$ i.i.d. from a full-rank k-parameter exponential family it holds that

- *The statistic $T(\mathbf{x}) ;= \left(\sum_{i=1}^n B_1(x_i), \ldots, \sum_{i=1}^n B_k(x_i)\right)$ is minimal sufficient.*
- *The distribution of $T(\mathbf{x})$ belongs to a k-parameter exponential family.*

$$
\begin{aligned}
L(\theta; x) &= \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \exp\left\{\sum_{j=1}^k A_j(\theta) B_j(x_i) + C(x_i) + D(\theta)\right\} \\
&= \exp\left\{\sum_{j=1}^k A_j(\theta)\left(\sum_{i=1}^n B_j(x_i)\right) + nD(\theta) + \sum_{i=1}^n C(x_i)\right\}.
\end{aligned}
$$

Sufficiency in an exponential family

Suppose the family is in canonical form so $\phi_j = A_j(\theta)$, and let $t_j = \sum_{i=1}^n B_j(x_i)$, $C(x) = \sum_{i=1}^n C(x_i)$.

$$L(\theta; x) = \exp\left\{ \sum_{j=1}^k \phi_j t_j + nD(\theta) + C(x) \right\}.$$

By the factorization criterion $t_1, \ldots, t_k$ are sufficient statistics for $\phi_1, \ldots, \phi_k$. In fact, we do not need canonical form. If

$$L(\theta; x) = \exp\left\{ \sum_{j=1}^k A_j(\theta) t_j + nD(\theta) + C(x) \right\}$$

is a minimal k-dimensional linear exponential family then (by the regularity conditions above) $t_1, \ldots, t_k$ are minimal sufficient for $\theta_1, \ldots, \theta_k$. Minimal sufficiency is verified using Lemma 1.