Sufficient Statistics

The Factorization Theorem

Minimal Sufficient Statistics
By data reduction we mean using a statistic $T(X)$, instead of the entire sample $X = (X_1, \ldots, X_n)$, in order to make inferences about an unknown parameter $\theta$.

When $T(X)$ is used for inference, two different observed samples $x$ and $y$ that satisfy $T(x) = T(y)$ lead to the same inference.

Let $\mathcal{X}$ be the sample space and $\mathcal{T} = \{t : t = T(x) \text{ for some } x \in \mathcal{X}\}$. The statistic $T(x)$ partitions the sample space into sets

$$A_t = \{x : T(x) = t\}, \text{ for } t \in \mathcal{T}.$$ 

Thus, reporting $T(x) = t$ is equivalent to reporting $x \in A_t$. 
A statistic $T(X)$ is **sufficient** for $\theta$ if inferences about $\theta$ depend on $X$ only through $T(X)$. (Informal definition.)

A statistic $T(X)$ is **sufficient** for $\theta$ if the conditional distribution of $X$ given $T(X)$ does not depend on $\theta$. (Formal definition.)

**Remark:** The need to work with conditional distributions restricts all our proofs to the discrete case:

$$P_\theta(X = x | T(X) = t) = \begin{cases} 0 & \text{if } T(x) \neq t \\ \frac{P_\theta(x, T(X) = t)}{P_\theta(T(X) = t)} & \text{if } T(x) = t. \end{cases}$$

In the continuous case, $P_\theta(X = x, T(X) = t)/P_\theta(T(X) = t)$ would be replaced by the corresponding expression with densities. However, the density of $(X', T(X))'$ does not exist because the dimension of the subspace it ranges is smaller than its dimension.
As an example, in the normal case \((X', \bar{X})'\) does not have a density. However, results will also be stated for the continuous case. □

**Example**

An interesting consequence of the definition of sufficiency is that by observing only \(T(X)\) we can generate a random vector \(Y\) which has the same distribution as \(X\).

**Proof.** Since \(T(X)\) is sufficient, \(P(X = x| T(X) = T(x))\) is known (since it does not depend on the unknown \(\theta\)). Generate an observation \(Y\) from the the conditional distribution \(P(X = x| T(X) = T(x))\). Then

\[
P_\theta(Y = x) = P(Y = x| T(X) = T(x)) P_\theta(T(X) = T(x))
\]

\[
= P(X = x| T(X) = T(x)) P_\theta(T(X) = T(x))
\]

\[
= P_\theta(Y = x).
\]
Example (Sufficient statistic for a density \( f \))

Let \( X_1, \ldots, X_n \) be iid with density \( f \). Show that the order statistics \( X_{(1)}, \ldots, X_{(n)} \) are sufficient for \( \theta = f \).

**Proof.** Set \( T(\mathbf{X}) = (X_{(1)}, \ldots, X_{(n)}) \) and note that \( A_{T(\mathbf{x})} \) consists of the \( n! \) vectors \( \mathbf{y} = (y_1, \ldots, y_n) \) such that 
\[
(y_{(1)}, \ldots, y_{(n)}) = (x_{(1)}, \ldots, x_{(n)}).
\]
For example, if \( n = 3 \) and \( \mathbf{x} = (x_1, x_2, x_3) = (1, 3, 2) \), then \( T(\mathbf{x}) = (1, 2, 3) \) and \( A_{T(\mathbf{x})} \) consists of the 3! permutations of \( \{1, 2, 3\} \). Since the observations are iid, when \( T(\mathbf{x}) \) is given the \( n! \) members of \( A_{T(\mathbf{x})} \) are equally likely. Thus,
\[
P(\mathbf{X} = \mathbf{x} \mid T(\mathbf{X}) = t) = \frac{1}{n!} I(\mathbf{x} \in A_t).
\]
Thus, \( T(\mathbf{X}) = (X_{(1)}, \ldots, X_{(n)}) \) is sufficient for \( f \).
Theorem

$T(X)$ is sufficient for $\theta$ iff the ratio $p(x|\theta)/q(T(x)|\theta)$ is independent of $\theta$, where $p(x|\theta)$ and $q(t|\theta)$ are the PMFs, or PDFs, of $X$ and $T(X)$, respectively.

Proof. It follows from

$$P_\theta(X = x|T(X) = T(x)) = \frac{P_\theta(X = x, T(X) = T(x))}{P_\theta(T(X) = T(x))}$$

$$= \frac{P_\theta(X = x)}{P_\theta(T(X) = T(x))}$$

$$= \frac{p(x|\theta)}{q(T(x)|\theta)}$$
Example (Sufficient statistic for the Bernoulli $p$)

$X_1, \ldots, X_n$ are iid Bernoulli($p$). Show that $T(X)$ is sufficient for $p$.

**Solution.** The ratio of pmf’s is

$$\frac{p(x|\theta)}{q(T(x)|\theta)} = \frac{\prod p^{x_i}(1 - p)^{1-x_i}}{(\frac{n}{T(x)}) p^{T(x)}(1 - p)^{n-T(x)}}$$

$$= \frac{1}{(\frac{n}{T(x)})}$$

Thus, $T(X)$ is sufficient for $p$. (Working directly with the definition is not straightforward.)
Example (Sufficient statistic for the Normal $\mu$)

Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$, where $\sigma^2$ is known. Show that $T(X) = \bar{X}$ is sufficient.

Solution: The pdf of $X$ is

$$f_X(x|\mu) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \right\}$$

$$= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right] \right\}$$

Since $f_{\bar{X}}(\bar{x}|\mu) = (2\pi\sigma^2)^{-1/2} e^{-\frac{1}{2\sigma^2} n(\bar{x}-\mu)^2}$ the ratio $f_X(x|\mu)/f_{\bar{X}}(\bar{x}|\mu)$ does not depend on $\mu$, which shows that $\bar{X}$ is sufficient for $\mu$. 
Theorem (Factorization Theorem)

\( T(\mathbf{X}) \) is sufficient for \( \theta \) iff there exist functions \( g(t|\theta) \) and \( h(\mathbf{x}) \) such that the joint pdf of pmf, \( f(\mathbf{x}|\theta) \), of \( \mathbf{X} \) can be written as

\[
f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}).
\]

Proof. "⇒"

\[
f(\mathbf{x}|\theta) = P_{\theta}(\mathbf{X} = \mathbf{x}) = P_{\theta}(\mathbf{X} = \mathbf{x}, T(\mathbf{X}) = T(\mathbf{x}))
= P_{\theta}(T(\mathbf{X}) = T(\mathbf{x}))P(\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = T(\mathbf{x}))
= g(T(\mathbf{x})|\theta)h(\mathbf{x}).
\]

"⇐" Then the pmf, \( q(t|\theta) \), of \( T(\mathbf{X}) \) can be written as

\[
q(t|\theta) = \sum_{\mathbf{y} \in A_{T(\mathbf{x})}} f(\mathbf{y}|\theta) = g(T(\mathbf{x})|\theta) \sum_{\mathbf{y} \in A_{T(\mathbf{x})}} h(\mathbf{y})
\]
Thus,

\[ \frac{f(x|\theta)}{q(T(x)|\theta)} = \frac{h(x)}{\sum_{y \in A_{T(x)}} h(y)} \]

does not depend on \( \theta \) which implies that \( T(X) \) is sufficient for \( \theta \).
Example (Sufficient statistic for the Normal $\mu$)

Let again $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$, where $\sigma^2$ is known. Show that $T(X) = \bar{X}$ is sufficient.

Solution: We saw that the pdf of $X$ is

$$f_X(x | \mu) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right] \right\}$$

This can be written in the form of the Factorization Theorem with

$$g(T(x) | \theta) = \exp \left\{ -\frac{1}{2\sigma^2} n(\bar{x} - \mu)^2 \right\} \quad \text{and} \quad h(x) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum (x_i - \bar{x})^2 \right\}$$
Example (Sufficient statistic for the discrete uniform $1, \ldots, \theta$)

Let $X_1, \ldots, X_n$ be iid from the discrete uniform on $\{1, 2, \ldots, \theta\}$, where $\theta$ is a positive integer. Show that $T(x) = \max\{x_1, \ldots, x_n\}$ is sufficient for $\theta$.

Solution. Here

$$f(x|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I(x_i \in \{1, 2, \ldots, \theta\})$$

$$= \frac{1}{\theta^n} I(T(x) \in \{1, 2, \ldots, \theta\}) \prod_{i=1}^{n} I(x_i \in \mathcal{N})$$

$$= g(T(x)|\theta)h(x).$$
Example (Sufficient statistic for the Normal $\theta = (\mu, \sigma^2)$)

Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$, where both $\mu$ and $\sigma^2$ are unknown. Show that $T(X) = (\bar{X}, S^2)$ is sufficient for $\theta = (\mu, \sigma^2)$.

**Solution:** We saw that the pdf of $X$ is

$$f_X(x|\theta) = (2\pi \sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right] \right\}$$

$$= (2\pi \sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (n - 1)S^2 + n(\bar{x} - \mu)^2 \right] \right\}$$

This can be written in the form of the Factorization Theorem with

$$g(T(x)|\theta) = f_X(x|\theta) \quad \text{and} \quad h(x) = 1.$$
Theorem

Let $X_1, \ldots, X_n$ be iid from a pdf of pmf $f(x|\theta)$ that belongs in an exponential family given by

$$f(x|\theta) = h(x) \exp \left( \sum_{i=1}^{k} w_i(\theta) t_i(x) \right),$$

where $\theta = (\theta_1, \ldots, \theta_d)$, $d \leq k$. Then,

$$T(X) = \left( \sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j) \right)$$

is sufficient for $\theta$. 
Sufficient statistics are not unique. If \( T(X) \) is sufficient and \( T^*(X) \) is any other statistic such that \( T(X) = g_1(T^*(X)) \), for some function \( g_1 \), then \( T^*(X) \) is also sufficient: Write

\[
f(x|\theta) = g(T(x)|\theta)h(x) = g(g_1(T^*(x))|\theta)h(x) = g^*(T^*(x)|\theta)h(x)
\]

In particular, if \( T(X) \) is sufficient then so is

\[
T^*(X) = (T(X), T_1(X)),
\]

where \( T_1(X) \) is any other statistic.

If \( T(X) = g_1(T^*(X)) \) then the partition of \( \mathcal{X} \) defined by \( T(x) \) is coarser than that defined by \( T^*(x) \).
If \( T(x) \) and \( T^*(x) \) are both sufficient then, because we are interested in data reduction, we prefer the one which induces the coarser partition of \( \mathcal{X} \). In general, we prefer what is called *minimal sufficient* statistic.

**Definition**

\( T(X) \) is minimal sufficient if it is sufficient and for any other sufficient statistic \( T^*(X) \), \( T(X) \) is a function of \( T^*(X) \).

The Factorization theorem is only meant to identify a sufficient statistic, not necessarily the minimal sufficient statistic. For example, in the normal case with \( \sigma \) known we can do the factorization of the density as in the case where \( \sigma \) is unknown, and correctly conclude that \((\bar{X}, S^2)\) is sufficient for \( \mu \).
Theorem

If $T(X)$ has the property that the ratio

$$\frac{f(x|\theta)}{f(y|\theta)}$$

does not depend on $\theta$ iff $T(x) = T(y)$, then $T(X)$ is minimal sufficient for $\theta$.

Remark: Minimal sufficient statistics are still not unique: If $T(X)$ is minimal sufficient and $g$ is 1-1 then $T^*(X) = g(T(X))$ is also minimal sufficient. However, the partitions of $\mathcal{X}$ induced by $T(X)$ and $T^*(X)$ are the same. It should also be kept in mind that

a) if the partitions induced by $T(X)$ and $T^*(X)$ are the same, then $T(X) = g(T^*(X))$ for some 1-1 function $g$, and

b) if the partition induced by $T^*(X)$ is finer than that induced by $T(X)$ then $T^*(X) = g(T(X))$. 
Proof of Theorem:

Let \( T(X) \) satisfy the condition of the theorem. We will first show that it is sufficient and then that it is minimal sufficient. Let \( x_t \) denote an element of \( A_t \). With this notation we have

\[
T(x_t) = t \quad \text{and} \quad T(x_{T(x)}) = T(x).
\]

From the assumption we have

\[
f(x|\theta)/f(x_{T(x)}|\theta) = h(x),
\]

some function that does not depend on \( \theta \). Thus,

\[
f(x|\theta) = f(x_{T(x)}|\theta)h(x),
\]

so by the Factorization theorem \( T(X) \) is sufficient.
Proof of Theorem Continued:

Let $T'(X)$ be another sufficient statistic. By the Factorization theorem

$$
\frac{f(x|\theta)}{f(y|\theta)} = \frac{h'(x)g'(T'(x)|\theta)}{h'(y)g'(T'(y)|\theta)}
$$

Thus, $T'(x) = T'(y)$ implies that $f(x|\theta)/f(y|\theta)$ does not depend on $\theta$ and thus, from the assumption, $T(x) = T(y)$. It follows that the partition of $\mathcal{X}$ induced by $T'(X)$ is finer than that induced by $T(X)$ which implies that $T(X)$ is minimal sufficient.
Example

Show that \((\bar{X}, S^2)\) is minimal sufficient for the normal \(\theta = (\mu, \sigma^2)\).

Solution.

\[
\frac{f(x|\theta)}{f(y|\theta)} = \frac{\exp \left[ - \left( n(\bar{x} - \mu)^2 + (n - 1)s_x^2 \right) / 2\sigma^2 \right]}{\exp \left[ - \left( n(\bar{y} - \mu)^2 + (n - 1)s_y^2 \right) / 2\sigma^2 \right]}
\]

This is constant with respect to \(\theta\) iff \(\bar{x} = \bar{y}\) and \(s_x = s_y\).
Example

Let \( X_1, \ldots, X_n \) be iid from the uniform on \([\theta, \theta + 1]\) distribution. Find a minimal sufficient statistic for \( \theta \).

**Solution.** Here

\[
f(x|\theta) = \prod_i I(\theta \leq x_i \leq \theta + 1) = I(\theta \leq \min_i \{x_i\}) I(\max_i \{x_i\} \leq \theta + 1).
\]

Thus,

\[
\frac{f(x|\theta)}{f(y|\theta)} = \begin{cases} 
1 & \text{if } \begin{cases} 
\min_i \{x_i\} = \min_i \{y_i\} \\
\max_i \{x_i\} = \max_i \{y_i\}
\end{cases} \\
h(x, y, \theta) & \text{otherwise}
\end{cases}
\]

where \( h(x, y, \theta \) can be 0 or \( \infty \) depending on the value of \( \theta \). It follows that \((\min_i \{x_i\}, \max_i \{x_i\})\) is minimal sufficient.