Lecture 20: Minimal sufficiency

There are many sufficient statistics for a given family $\mathcal{P}$.
In fact, X (the whole data set) is sufficient.
If T is a sufficient statistic and $T = \psi(S)$, where ψ is measurable and S is another statistic, then S is sufficient.
This is obvious from Theorem 2.2 if the population has a p.d.f., but it can be proved directly from Definition 2.4 (Exercise 25).
For instance, if $X_1, ..., X_n$ are iid with $P(X_i = 1) = \theta$ and $P(X_i = 0) = 1 - \theta$, then $(\sum_{i=1}^m X_i, \sum_{i=m+1}^n X_i)$ is sufficient for θ, where m is any fixed integer between 1 and n.
If T is sufficient and $T = \psi(S)$ with a measurable ψ that is not one-to-one, then $\sigma(T) \subset \sigma(S)$ and T is more useful than S, since T provides a further reduction of the data (or σ-field) without loss of information.
Is there a sufficient statistic that provides "maximal" reduction of the data?

If a statement holds except for outcomes in an event A satisfying $P(A) = 0$ for all $P \in \mathcal{P}$, then we say that the statement holds a.s. $\mathcal{P}$.

Definition 2.5 (Minimal sufficiency). Let T be a sufficient statistic for $P \in \mathcal{P}$. T is called a *minimal sufficient* 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. $\mathcal{P}$.

If both T and S are minimal sufficient statistics, then by definition there is a one-to-one measurable function ψ such that $T = \psi(S)$ a.s. $\mathcal{P}$.
Hence, the minimal sufficient statistic is unique in the sense that two statistics that are one-to-one measurable functions of each other can be treated as one statistic.

Example 2.13. Let $X_1, ..., X_n$ be i.i.d. random variables from P_θ, the uniform distribution $U(\theta, \theta+1)$, $\theta \in \mathcal{R}$. Suppose that $n > 1$. The joint Lebesgue p.d.f. of $(X_1, ..., X_n)$ is

$$f_\theta(x) = \prod_{i=1}^n I_{(\theta,\theta+1)}(x_i) = I_{(x_{(n)}-1, x_{(1)})}(\theta), \quad x = (x_1, ..., x_n) \in \mathcal{R}^n,$$

where $x_{(i)}$ denotes the ith smallest value of $x_1, ..., x_n$. By Theorem 2.2, $T = (X_{(1)}, X_{(n)})$ is sufficient for θ. Note that

$$x_{(1)} = \sup\{\theta : f_\theta(x) > 0\} \quad \text{and} \quad x_{(n)} = 1 + \inf\{\theta : f_\theta(x) > 0\}.$$

If $S(X)$ is a statistic sufficient for θ, then by Theorem 2.2, there are Borel functions h and g_θ such that $f_\theta(x) = g_\theta(S(x))h(x)$. For x with $h(x) > 0$,

$$x_{(1)} = \sup\{\theta : g_\theta(S(x)) > 0\} \quad \text{and} \quad x_{(n)} = 1 + \inf\{\theta : g_\theta(S(x)) > 0\}.$$

Hence, there is a measurable function ψ such that $T(x) = \psi(S(x))$ when $h(x) > 0$. Since $h > 0$ a.s. $\mathcal{P}$, we conclude that T is minimal sufficient.

Minimal sufficient statistics exist under weak assumptions, e.g., $\mathcal{P}$ contains distributions on $\mathcal{R}^k$ dominated by a σ-finite measure (Bahadur, 1957).

Useful tools for finding minimal sufficient statistics.

Theorem 2.3. Let $\mathcal{P}$ be a family of distributions on $\mathcal{R}^k$.
(i) Suppose that $\mathcal{P}_0 \subset \mathcal{P}$ and a.s. $\mathcal{P}_0$ implies a.s. $\mathcal{P}$. If T is sufficient for $P \in \mathcal{P}$ and minimal sufficient for $P \in \mathcal{P}_0$, then T is minimal sufficient for $P \in \mathcal{P}$.
(ii) Suppose that $\mathcal{P}$ contains p.d.f.'s $f_0, f_1, f_2, ...$, w.r.t. a σ-finite measure. Let $f_\infty(x) = \sum_{i=0}^\infty c_i f_i(x)$, where $c_i > 0$ for all i and $\sum_{i=0}^\infty c_i = 1$, and let $T_i(X) = f_i(x)/f_\infty(x)$ when $f_\infty(x) > 0$, $i = 0, 1, 2, ...$ Then $T(X) = (T_0, T_1, T_2, ...)$ is minimal sufficient for $P \in \mathcal{P}$. Furthermore, if $\{x : f_i(x) > 0\} \subset \{x : f_0(x) > 0\}$ for all i, then we may replace f_∞ by f_0, in which case $T(X) = (T_1, T_2, ...)$ is minimal sufficient for $P \in \mathcal{P}$.
(iii) Suppose that $\mathcal{P}$ contains p.d.f.'s f_P w.r.t. a σ-finite measure and that there exists a sufficient statistic $T(X)$ such that, for any possible values x and y of X, $f_P(x) = f_P(y)\phi(x, y)$ for all P implies $T(x) = T(y)$, where ϕ is a measurable function. Then $T(X)$ is minimal sufficient for $P \in \mathcal{P}$.
Proof. (i) If S is sufficient for $P \in \mathcal{P}$, then it is also sufficient for $P \in \mathcal{P}_0$ and, therefore, $T = \psi(S)$ a.s. $\mathcal{P}_0$ holds for a measurable function ψ. The result follows from the assumption that a.s. $\mathcal{P}_0$ implies a.s. $\mathcal{P}$.
(ii) Note that $f_\infty > 0$ a.s. $\mathcal{P}$. Let $g_i(T) = T_i$, $i = 0, 1, 2, ...$ Then $f_i(x) = g_i(T(x))f_\infty(x)$ a.s. $\mathcal{P}$. By Theorem 2.2, T is sufficient for $P \in \mathcal{P}$. Suppose that $S(X)$ is another sufficient statistic. By Theorem 2.2, there are Borel functions h and $\tilde{g}_i$ such that $f_i(x) = \tilde{g}_i(S(x))h(x)$, $i = 0, 1, 2, ...$ Then $T_i(x) = \tilde{g}_i(S(x))/\sum_{j=0}^\infty c_j \tilde{g}_j(S(x))$ for x's satisfying $f_\infty(x) > 0$. By Definition 2.5, T is minimal sufficient for $P \in \mathcal{P}$. The proof for the case where f_∞ is replaced by f_0 is the same.
(iii) From Bahadur (1957), there exists a minimal sufficient statistic $S(X)$. The result follows if we can show that $T(X) = \psi(S(X))$ a.s. $\mathcal{P}$ for a measurable function ψ. By Theorem 2.2, there are Borel functions g_P and h such that $f_P(x) = g_P(S(x))h(x)$ for all P. Let $A = \{x : h(x) = 0\}$. Then $P(A) = 0$ for all P. For x and y such that $S(x) = S(y)$, $x \notin A$ and $y \notin A$,

$$\begin{aligned} f_P(x) &= g_P(S(x))h(x) \\ &= g_P(S(y))h(x)h(y)/h(y) \\ &= f_P(y)h(x)/h(y) \end{aligned}$$

for all P. Hence $T(x) = T(y)$. This shows that there is a function ψ such that $T(x) = \psi(S(x))$ except for $x \in A$. It remains to show that ψ is measurable. Since S is minimal sufficient, $g(T(X)) = S(X)$ a.s. $\mathcal{P}$ for a measurable function g. Hence g is one-to-one and $\psi = g^{-1}$. The measurability of ψ follows from Theorem 3.9 in Parthasarathy (1967).

Example 2.14. Let $\mathcal{P} = \{f_\theta : \theta \in \Theta\}$ be an exponential family with p.d.f.'s

$$f_\theta(x) = \exp\{[\eta(\theta)]^\tau T(x) - \xi(\theta)\}h(x)$$

Suppose that there exists $\Theta_0 = \{\theta_0, \theta_1, ..., \theta_p\} \subset \Theta$ such that the vectors $\eta_i = \eta(\theta_i) - \eta(\theta_0)$, $i = 1, ..., p$, are linearly independent in $\mathcal{R}^p$. (This is true if the family is of full rank.) We have shown that $T(X)$ is sufficient for $\theta \in \Theta$. We now show that T is in fact minimal sufficient for $\theta \in \Theta$. Let $\mathcal{P}_0 = \{f_\theta : \theta \in \Theta_0\}$. Note that the set $\{x : f_\theta(x) > 0\}$ does not depend on θ. It follows from Theorem 2.3(ii) with $f_\infty = f_{\theta_0}$ that

$$S(X) = \left(\exp\{\eta_1^\tau T(x) - \xi_1\}, ..., \exp\{\eta_p^\tau T(x) - \xi_p\}\right)$$

is minimal sufficient for $\theta \in \Theta_0$, where $\xi_i = \xi(\theta_i) - \xi(\theta_0)$. Since η_i's are linearly independent, there is a one-to-one measurable function ψ such that $T(X) = \psi(S(X))$ a.s. $\mathcal{P}_0$. Hence, T is minimal sufficient for $\theta \in \Theta_0$. It is easy to see that a.s. $\mathcal{P}_0$ implies a.s. $\mathcal{P}$. Thus, by Theorem 2.3(i), T is minimal sufficient for $\theta \in \Theta$.

The results in Examples 2.13 and 2.14 can also be proved by using Theorem 2.3(iii).

The sufficiency (and minimal sufficiency) depends on the postulated family $\mathcal{P}$ of populations (statistical models).
It may not be a useful concept if the proposed statistical model is wrong or at least one has some doubts about the correctness of the proposed model.
From the examples in this section and some exercises in §2.6, one can find that for a wide variety of models, statistics such as the sample mean $\bar{X}$, the sample variance S^2, $(X_{(1)}, X_{(n)})$ in Example 2.11, and the order statistics in Example 2.9 are sufficient.
Thus, using these statistics for data reduction and summarization does not lose any information when the true model is one of those models but we do not know exactly which model is correct.

A minimal statistic is not always the "simplest sufficient statistic".
For example, if $\bar{X}$ is minimal sufficent, then so is $(\bar{X}, \exp\{\bar{X}\})$.