MINIMAL SUFFICIENT STATISTICS

If a statistic is sufficient, then so is an “augmented” statistic \( S'(S,T) \). Since the goal is to summarize information concisely, we desire to work with “minimal” sufficient statistics.

**Def:** A statistic \( S = T(X) \) is minimal sufficient if, for any other sufficient statistic, \( T'(X), T(X) \) is a function of \( T'(X) \).

**Example:** Suppose \( T(X) = \sum_{i=1}^{n} X_i \), \( T'(X) = \left( \sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2 \right) \)

Clearly, \( T(X) = g_1(T'(X)) \), \( T'(X) \neq g_2(T(X)) \)

We desire to use minimal sufficient statistics whenever possible as the greatest reduction of the data.

**Example:** Assume \( X_i, i = 1,\ldots,n \) are iid Bernoulli. Let \( S = T(X) = \sum_{i=1}^{n} X_i \). Now suppose that \( V = g(S) \). By definition, \( V \) is a summary of \( S \) only if for two different values of \( S, S = \eta_1, S = \eta_2 \), \( g(\eta_1) = g(\eta_2) = v \). Let’s check it.

\[
P(S = T(x)|V = v) = \frac{P(S = s \cap V = v)}{P(V = v)}
\]

\[
\begin{align*}
&= \binom{n}{s} \theta^s (1-\theta)^{n-s} \\
&= \binom{n}{\eta_1} \theta^{\eta_1} (1-\theta)^{n-\eta_1} + \binom{n}{\eta_2} \theta^{\eta_2} (1-\theta)^{n-\eta_2}
\end{align*}
\]

But this is still a function of \( \theta \), so \( V \) is NOT sufficient. \( S \) is sufficient. \( S \) is also minimal.
Can we formalize this? Sometimes factorization of the “likelihood function”,

\[ L(\theta|x) = f(x|\theta) \]

gives us the minimal sufficient statistic directly. In other cases, we can exploit the relationship between ratios of likelihood functions and minimal sufficiency.

**Theorem:** Let \( f(x|\theta) \) be the pdf (pmf) of a sample \( X \). Suppose that a function \( T(X) \) exists such that for every two sample points (i.e. samples of observations) \( x, y \) the ratio \( \frac{f(x)}{f(y)} \) is a constant as a function of \( \theta \) iff \( T(X) = T(Y) \). Then \( T(X) \) is a minimal sufficient statistic.

**Proof:** See class handout.

Consider a statistic as dividing the sample space into classes called *equivalence classes*. Each class contains all observations \( X \) with the same value of \( S \). If \( S \) is minimal sufficient, then so is any \( S \) function of \( S \) (unique inverse). So minimal sufficiency is somehow related to the set of equivalence classes – but *not* to the particular labeling of the equivalence classes.

Consider the partition of the sample space:

\[
D(x) = \left\{ z; \frac{f(z|\theta)}{f(x|\theta)} = h(z,x) \forall \theta \in \Theta \right\}
\]

That is, where the ratios of likelihood functions are proportional.

If \( z \in D(x_1) \) and \( z \in D(x_2) \), then \( D(x_1) = D(x_2) \).

This gives insight into the requirements for minimal sufficiency.
By the Factorization Theorem:
\[ f(x|\theta) = g(T(x)|\theta) h(x) = g(x|\theta) h(x) \]

Now suppose that for some other set of data \( z \),
\[ f(z|\theta) = g(T(z)|\theta) h(z) \text{ (same family of pdfs)} \]

If \( T(x) = T(z) \), then
\[ g(T(x)) = g(T(z)) \text{ (since the function } g \text{ is the same)} \]
\[ f(x|\theta) = f(z|\theta) \frac{h(x)}{h(z)} \]
\[ \Rightarrow \frac{f(x|\theta)}{f(z|\theta)} = m(x,z) \]

But this implies that \( x \) and \( z \) are in the same equivalence class. Therefore, the partition defined by the ratio of the likelihoods includes that based on the statistic \( T \), so this partition is minimal sufficient. (This result is essentially Theorem 6.2.13, which is proved differently).

**Example:** Suppose that \( X_i \) are iid Poisson \((\lambda)\). We know that \( S_1 = T(X) = \sum_{i=1}^{n} X_i \) and \( S_2 = T'(X) = X_{(1)} < \ldots < X_{(n)} \) are both sufficient for \( \lambda \).

Consider \( S_1 \),
\[
\frac{f(x|\lambda)}{f(y|\lambda)} = \frac{\exp(-n\lambda) \lambda^{\Sigma_{x_i}}/\prod x_i!}{\exp(-n\lambda) \lambda^{\Sigma_{y_j}}/\prod y_j!} = \frac{\lambda^{\Sigma_{x_i}-\Sigma_{y_j}}/\prod x_i!/\prod y_j!}{\prod x_i!/\prod y_j!} = h(x,y) i,j = 1,\ldots,n
\]
So, this is a minimal sufficient statistic.

Now consider $S_2$.

$$f(x|\lambda) = \frac{\exp(-n\lambda) \lambda^{\sum_{i=1}^r x_i} / \prod_{i,j} x_{ij} !}{\exp(-n\lambda) \lambda^{\sum_{j=1}^m y_{j}'} / \prod_{j,k} y_{jk} !}$$

$S_1$ is a function of $S_2$, so $S_1$ is a “coarser” statistic. Therefore, $S_2$ cannot be minimum sufficient.

**Example:** Assume that you are performing life testing. Suppose that of $n$ components, $r$ die after $y_1, \ldots, y_r$ time periods and that $n-r$ are still alive after $y_1', \ldots, y_m'$. Assuming that the lives of $Y_1, \ldots, Y_n$ are $\sim$ iid 

$$f(y|\lambda) = \lambda \exp(-\lambda y_1)$$

then the joint pdf is,

$$f(y|\lambda) = \prod_{j=1}^r \lambda \exp(-\lambda y_j) \prod_{k=1}^m \exp(-\lambda y_{j}')$$

(The second factor is the probability of times to death exceeding $y_{j}', k = 1, \ldots, m$)

Then

$$f(y|\lambda) = \lambda^r \exp(-\lambda y^*),$$

where $y^* = \sum_{j=1}^r y_j + \sum_{k=1}^m y_{j}'$. What is a sufficient statistic $S$?

$$S = (R, Y^*)$$ (note that $R$ is the random variable of which $r$ is a particular observation).

For this problem, you have a sufficient statistic that is of dimension 2, while the dimension of the parameter vector is 1. Is this a minimal sufficient statistic? Yes, check the ratio.

Can we say anything special if we are dealing with the exponential family of distributions:

$$f(x|\theta) = h(x) c(\theta) \exp \left( \sum_{i=1}^k w_i(\theta) t_i(x) \right)$$
Recall that $\sum_{j=1}^{n} t_{j}(x_{j})$ is sufficient. Now look at the ratio of joint pdfs! Thus, if $x_{j}\simiid$ and a member of the exponential family, these statistics are also minimal sufficient.

**Example:** Suppose that $X \sim N(\mu, \sigma^{2})$ and that neither is known.

$$f(x) \quad f(y) = \exp\left(-\frac{1}{2\sigma^{2}}\left(\sum_{i=1}^{n} x_{i}^{2} - \sum_{i=1}^{n} y_{i}^{2} - 2\mu \left(\sum_{i=1}^{n} x_{i} - \sum_{i=1}^{n} y_{i}\right)\right)\right)$$

Now if $\sum_{i=1}^{n} x_{i} = \sum_{i=1}^{n} y_{i}$, the ratio is independent of $\mu, \sigma$. So,

$$\left(\sum_{i=1}^{n} x_{i}, \sum_{i=1}^{n} x_{i}^{2}\right)$$

is minimal sufficient for $\mu, \sigma$.

Note Example 6.2.14 in C&B. The densities are expressed in terms of $\bar{x}, s^{2}$. Not surprisingly, the ratio of the likelihood functions is still independent of the parameters, so these are minimal sufficient statistics as well. *Neither sufficient nor minimal sufficient statistics is necessarily unique for a family.*

**Example:** Assume $X_{i} \sim \text{Gamma}(\alpha, \beta)$.

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1} \exp(-x/\beta)}{\Gamma(\alpha) \beta^{\alpha}} = \exp\left\{\frac{-x}{\beta} + \alpha \ln\frac{1}{\beta} + (\alpha-1)\ln(x) - \ln(\Gamma(\alpha))\right\}$$

Assume that $\alpha$ is known:

$$w_{1}(\beta) = -\frac{1}{\beta}, \quad t(x) = x, \quad c(\beta) = \alpha \ln\frac{1}{\beta} - \ln(\Gamma(\alpha)), \quad h(x) = (\alpha-1)\ln x$$

Suppose that $\alpha$ were unknown……
Consider a generalization of the exponential family to $\theta = (\theta_i, i = 1, \ldots, q)$. The joint pdf is

$$f(x_j | \theta) = \exp\left\{ \sum_{m=1}^{k} a_m(\theta) b_{jm}(x_j) + c_j(\theta) + d_j(x_j) \right\}$$

$$f(x | \theta) = \exp\left\{ \sum_{m=1}^{k} a_m T_m(x) + c'(\theta) + d'(x) \right\}$$

$$T_m(x) = \sum_{j=1}^{n} b_{jm}(x_j)$$

For the general exponential family with multiple unknown parameters, the dimension of the parameter vector $q$ and the index $m$ of the sum over $k$ terms are not necessarily the same. (Recall the life testing example.)

If $k < q$, some nonlinear relationship exists between the parameters

- $k = q$, standard case
- $k > q$, not common, but can happen

**Example:** Assume that $\gamma = \frac{\sigma}{\mu}$, fixed. $X \sim N\left(\mu, \gamma^2 \mu^2\right)$

$$f(x) = \frac{1}{\gamma \mu \sqrt{2\pi}} \exp\left\{ -\frac{(x-\mu)^2}{2\gamma^2 \mu^2} \right\}$$

$$= \exp\left\{ -\frac{x^2}{2\gamma^2 \mu^2} + \frac{x}{\gamma^2 \mu} - \frac{1}{2} \ln \left(2\pi \gamma^2 \mu^2\right) \right\}$$

Here, $k=2, q=1$, where a minimum sufficient statistic is $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$

See also Example 6.2.15, a uniform distribution.