STA3000 Note on sufficiency

Definition A statistic $S = s(X)$ is sufficient for θ (in the family of models $\{f(x;\theta); \theta \in \Theta\}$) if and only if $f(x|s)$ is free of θ.
Factorization Theorem A statistic S is sufficient for θ (in the family of distributions $\{f(x;\theta), \theta \in \Theta\}$), if and only if there exist functions g and h such that

$$f(x;\theta) = g(s(x);\theta)h(x) \qquad \forall \quad \theta.$$

This theorem can be restated as saying that the likelihood function (which we've defined as an equivalence class) only depends on x through $s(x)$.

The usual interpretation of sufficiency (see, e.g. CH Ch. 2) is that a statistician observing X should come to the same conclusion about θ as a statistician observing only S, because the second statistician could recover the joint density for X from the marginal density for S by multiplying by a fixed density (that of X given S). That is, an observation from the conditional density of X given S can be generated without any knowledge of θ. Thus it could be generated by a random number generator. It is argued (and generally accepted) that this additional randomness cannot provide any information about the unknown parameter θ. This discussion assumes that the model is correct, of course. In fact, we shall see that the conditional density of X given S can be used for checking the model.

A sufficient statistic partitions the sample space $\mathcal{X}$ into subspaces. Two elements of $\mathcal{X}$ are in the same partition if they have the same value of $s(x)$. Typically the 'size' of the partition is substantially smaller that the 'size' of the sample space. For example, if X is a random variable on R^n, $S(X) = \sum X_i$ takes values on R. The biggest reduction in dimension is obtained with the coarsest partition. The statistic inducing the coarsest possible partition (if it exists) is called the minimal sufficient statistic.
Definition. A statistic $S = s(X)$ is a minimal sufficient statistic if it is a function of every other sufficient statistic.

This definition is obviously difficult to work with, but can be avoided by using the following theorem:
Theorem. The likelihood statistic is minimal sufficient.
Proof. The likelihood statistic is the partition of the sample space that puts two values of x into the same partition if they have the same likelihood function (i.e. their likelihood functions are in the same equivalence class). Thus x and x' are in the same partition if

$$\begin{aligned} L(\theta;x) &= L(\theta;x') \\ c(x)f(x;\theta) &= c(x')f(x';\theta) \\ f(x;\theta) &= k(x',x)f(x';\theta). \end{aligned}$$

If this holds, we write $s(x) = s(x')$ to indicate that $s(\cdot)$ is the function that indexes this partition.

Suppose $T = t(X)$ is some other sufficient statistic. By the factorization theorem

$$\begin{aligned} f(x;\theta) &= g(t(x);\theta)h(x) \\ f(x';\theta) &= g(t(x');\theta)h(x') \end{aligned}$$

If $t(x) = t(x')$ then

$$\begin{aligned} f(x;\theta) &= g(t(x');\theta)h(x) \\ &= f(x';\theta)\{h(x)/h(x')\} \\ &= f(x';\theta)k(x,x') \end{aligned}$$

so $s(x) = s(x')$, which shows that $s(\cdot)$ is a function of $t(\cdot)$.

This result is called the Lehmann-Scheffé theorem, and the proof given here is taken from CH, Ch.2. The difficulty in making it properly precise is that the conditional density of X, given S, is only defined up to sets of measure zero.