Sufficiency principle

If $T(x)$ is a suffucient statistic for θ, then any inference about θ should depend on the sample on the sample X only through the calue $T(x)$

Definition for sufficient statistic

$T(x)$ is a sufficient statistic for θ if the coditional distribution of sample Xn given the value of $T(x)$ does not depend on θ.

- Intuition under the definition: Given a data sample, if the only information we want to know is just θ. Then a person who knowns all the data(X) have the same amount of knowledge of θ as a person who only knowns $T(x)$.
- Example: 1. Sample sum is a ss for θ of $B(n, \theta)$.

Why order statistics are sufficient to an unknown pdf f

The joint distribution off all order statistics $X_{(1)}, \dots, X_{(n)}$ is

$$f_o(y_1, \dots, y_n) = n! f(y_1) \times \dots \times f(y_n)$$

for $y_1 \le \dots \le y_n$.

Thus, the joint distribution of $X_1, \dots, X_n$, given $X_{(1)}, \dots, X_{(n)}$ does not depend of the density f! We have

$$Pr(X_1 = x_1, \dots, X_n = x_n | X_{(1)} = y_1, \dots, X_{(n)} = y_n) = \frac{1}{n!}$$

when the multisets of the x_i's and of the y_i's are equal.

The intuition is that loosing the drawing order doesn't matter, for the X_i's are independent.

PS Re-reading this answer long after, I tend to think that the conclusion is clear in itself: given $y_1 < \dots < y_n$, $Pr(X_1 = x_1, \dots, X_n = x_n | X_{(1)} = y_1, \dots, X_{(n)} = y_n) = \frac{1}{n!}$ means that all re-orderings of the y_i's are equally probable, whatever the density f is. In fact I don't see how to prove the starting result I used (the joint distribution of order statistics) without proving this at the same time... This just comes from the fact that all the points $x = (x_1, \dots, x_n)$ obtained from a permutation of the y_i have the same density $\prod_i f(x_i) = \prod_i f(y_i)$.

Factorization Theorem

Let $f(x|\theta)$ be the pdf of sample X, $T(x)$ is a ss for θ iff there exists $g(T(x)|\theta), h(x)$ such that

$$f(x|\theta) = g(T(x)|\theta)h(x)$$

- Intuition: ss means that θ exert influence on the populatiuon X only through $T(X)$, while this theorem implies that the joint pdf of a population provides a channel for us to measure this connection. For a population, we can split its joint pdf into two parts where the first one characterize the relationship of θ and $T(X)$, and the other part is totally independent of θ. This is a mathematical characteristic of how θ exert influence on the populatiuon X only through $T(X)$.
- If θ is a vector, so does $T(x)$.
- A one-to-one mapping of a ss is also a ss.
- Minimal ss: $T(x)$ is a minimal ss iff for every two sample points $x, y, \frac{f(x|\theta)}{f(y|\theta)}$ is a const iff $T(x) = T(y)$.(A minimal ss is a statistic that has achieved the maximal amount of data reduction possible while still retaining all the informatin about the parameter θ).

Ancillary stats definition

Does not depend on the parameter θ is called an as.

- A location family: the range of a sample is as
- A scale family: $\frac{X_1}{X_2}, \ldots, \frac{X_{n-1}}{X_n}$ is as.
- as is not indep of ss. See Eg 6.2.20

An intuitive and interesting way to see completeness.

I will try to add to the other answer. First, completeness is a technical condition which mainly is justified by the theorems that use it. So let us start with some related concepts and theorems where they occur.

Let $X = (X_1, X_2, \ldots, X_n)$ represent a vector of iid data, which we model as having a distribution $f(x; \theta), \theta \in \Theta$ where the parameter θ governing the data is unknown. $T = T(X)$ is **sufficient** if the conditional distribution of $X \mid T$ does not depend on the parameter θ. $V = V(X)$ is **ancillary** if the distribution of V does not depend on θ (within the family $f(x; \theta)$). $U = U(X)$ is an **unbiased estimator of zero** if its expectation is zero, irrespective of θ. $S = S(X)$ is a **complete statistic** if any unbiased estimator of zero based on S is identically zero, that is, if $E\, g(S) = 0 (\text{for all } \theta)$ then $g(S) = 0$ a.e. (for all θ).

Now, suppose you have two different unbiased estimators of θ based on the sufficient statistic T, $g_1(T), g_2(T)$. That is, in symbols

$$\mathbb{E}\, g_1(T) = \theta,$$
$$\mathbb{E}\, g_2(T) = \theta$$

and $\mathbb{P}(g_1(T) \neq g_2(T)) > 0$ (for all θ). Then $g_1(T) - g_2(T)$ is an unbiased estimator of zero, which is not identically zero, proving that T is not complete. So, completeness of an sufficient statistic T gives us that there does exist only one unique unbiased estimator of θ based on T. That is already very close to the Lehmann–Scheffé theorem.

Let us look at some examples. Suppose $X_1, \ldots, X_n$ now are iid uniform on the interval $(\theta, \theta + 1)$. We can show that ($X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ is the order statistics) the pair $(X_{(1)}, X_{(n)})$ is sufficient, but it is not complete, because the difference $X_{(n)} - X_{(1)}$ is ancillary, we can compute its expectation, let it be c (which is a function of n only), and then $X_{(n)} - X_{(1)} - c$ will be an unbiased estimator of zero which is not identically zero. So our sufficient statistic, in this case, is not complete and sufficient. And we can see what that means: there exist functions of the sufficient statistic which are not informative about θ (in the context of the model). This cannot happen with a complete sufficient statistic; it is in a sense maximally informative, in that no functions of it are uninformative. On the other hand, if there is some function of the minimally sufficient statistic that has expectation zero, that could be seen as a *noise term*, disturbance/noise terms in models have expectation zero. So we could say that non-complete sufficient statistics *do contain some noise*.

Look again at the range $R = X_{(n)} - X_{(1)}$ in this example. Since its distribution does not depend on θ, it doesn't **by itself alone** contain any information about θ. But, together with the sufficient statistic, it does! How? Look at the case where $R = 1$ is observed.Then, in the context of our (known to be true) model, we have perfect knowledge of θ! Namely, we can say with certainty that $\theta = X_{(1)}$. You can check that any other value for θ then leads to either $X_{(1)}$ or $X_{(n)}$ being an impossible observation, under the assumed model. On the other hand, if we observe $R = 0.1$, then the range of possible values for θ is rather large (exercise ...).

In this sense, the ancillary statistic R does contain some information about the precision with which we can estimate θ based on this data and model. In this example, and others, the ancillary statistic R "takes over the role of the sample size". Usually, confidence intervals and such needs the sample size n, but in this example, we can make a *conditional confidence interval* this is computed using only R, not n (exercise.) This was an idea of Fisher, that inference should be conditional on some ancillary statistic.

Now, Basu's theorem: If T is complete sufficient, then it is independent of any ancillary statistic. That is, inference based on a complete sufficient statistic is simpler, in that we do not need to consider conditional inference. Conditioning on a statistic which is independent of T does not change anything, of course.

Then, a last example to give some more intuition. Change our uniform distribution example to a uniform distribution on the interval (θ_1, θ_2) (with $\theta_1 < \theta_2$). In this case the statistic $(X_{(1)}, X_{(n)})$ **is** complete and sufficient. What changed? We can see that completeness is really a property of the **model**. In the former case, we had a restricted parameter space. This restriction destroyed completeness by introducing relationships on the order statistics. By removing this restriction we got completeness! So, in a sense, lack of completeness means that the parameter space is not big enough, and by enlarging it we can hope to restore completeness (and thus, easier inference).

Some other examples where lack of completeness is caused by restrictions on the parameter space,

- see my answer to: What kind of information is Fisher information?
- Let $X_1, \dots, X_n$ be iid $\text{Cauchy}(\theta, \sigma)$ (a location-scale model). Then the order statistics in sufficient but not complete. But now enlarge this model to a fully nonparametric model, still iid but from some completely unspecified distribution F. Then the order statistics is sufficient and complete.
- For exponential families with canonical parameter space (that is, as large as possible) the minimal sufficient statistic is also complete. But in many cases, introducing restrictions on the parameter space, as with *curved exponential families*, destroys completeness.

A very relevant paper is An Interpretation of Completeness and Basu's Theorem.