Comments on sufficiency (30 march 2020)

Last week we found that in the class of unbiased regular estimators there is, sometimes, one with the least variance, which we call *efficient.* Furthermore, this smallest variance can be computed in many cases *even if we do not know, or even guess, which estimator might attain it.* That was the Cramér-Rao lower bound theorem.

We now explore another notion, not of optimality in any sense, but sort of a condition *sine qua non* for optimality. Loosely speaking, a sufficient statistic extracts as much information as the sample can supply about the value of a parameter (slide **The concept of sufficiency (I)**). Certainly, if our estimator is not obtaining as much information as possible from the sample, it can hardly be optimal: it is in this sense that we say sufficiency is a prerequisite for optimality.

In keeping with our usual method, we try first to motivate the concept with a simple example, then formalize it. These notes are meant to be read in conjunction with the slides prepared for presential lectures.

Sufficiency: the intuitive idea

Consider the following simplified setting. You are drilling for oil. Sometimes you find gas before you hit a deeper layer of oil, sometimes you don't. On the basis of extensive experience, you are able to set up the following Table, giving the probabilities of finding and not finding oil in a particular type of underground depending on whether you find gas first:

Table 1: Probability of finding oil in a perforation, depending on whether or not gas is found previously

Upper level	Deeper level	
Gas	Oil (θ_1)	Not oil (θ_2)
Found	0.70	0.30
Not found	0.70	0.30

When working on a new search at a mid depth you strike gas. Would that finding enhance your belief that oil lies below?

The superficial (and incorrect) answer is "yes". When you strike gas there is a 0.70 chance that deeper there will be oil. But when you look closely, you realize that when you do not strike gas, you also have a 0.70 chance of finding oil below. It seems clear, then, that the finding of gas does not add or detract to your chances of finding oil below.

We can make a general rule out of this simple observation: if the probability of "something" (here, existence of oil) does not depend of the value of "something else" (here, the observation of gas or no gas), then "something else" is not informative about "something".

Sufficiency: the formal concept

Consider now the following situation: we want to learn about the value of a parameter ("something"). We take a sample $\vec{X} = (X_1, \ldots, X_n)$ of observations whose density $f_X(x; \theta)$ depends on θ. Looking at the values of of the observations, we can learn about θ. Clear enough.

Now, suppose we compute a summary statistic of the sample, say $S(\vec{X})$. This summary statistic could be something like the mean, the median, the sample variance, whatever.

There are in general many samples that can produce a given value of $S(\vec{X})$. It makes sense to speak of the conditional density of $\vec{X} = (X_1, \ldots, X_n)$ given $S(\vec{X})$. This conditional density,

$$f_X(\vec{X}|S; \theta) = \frac{f_X(\vec{X}; \theta)}{f_X(\vec{S}; \theta)}$$

will in general depend on the same θ as the original density of X. However, in some cases (slide **The concept of sufficiency (II)**), this is not the case: $f_X(\vec{X}|S)$ does not depend on θ.

What can we conclude in these cases? Given S, the density of the particular values observed does not depend on θ, hence *it is not informative about* θ according to the rule we stated before.

Where is the information contained in the sample? In $S(\vec{X})$. It can be nowhere else. In these cases, we say that $S(\vec{X})$ is *sufficient for* θ. Once we know the value of $S(\vec{X})$, there is nothing to be gained from looking at the values $X_1, \ldots, X_n$.

Knowing that a statistic is sufficient, does not guarantee that it is optimal in any sense for the estimation of a parameter; but we should expect not to find an optimal estimator which is not a sufficient statistic. Therefore, it makes perfectly good sense to limit our search of estimators to the set of sufficient statistics.

There are a couple of examples of sufficient statistics in slides **The concept of sufficiency (III)** and **The concept of sufficiency (IV)**.

Characterizing sufficiency

Notice that to use the definition you need to know or guess which statistic might be sufficient and compute its density to replace in the ratio

$$\frac{f_X(\vec{X};\theta)}{f_X(\vec{S};\theta)}$$

This is very tedious if we want to check a large number of candidates to sufficiency.

The factorization theorem is a very simple result (which we do not prove, however) which has the advantage of being "constructive". We only have to write the likelihood and factor it in two terms, one containing and another one not containing the parameter. Whichever function of the data goes with the parameter, is a sufficient statistic (slides **The factorization theorem - I** and **The factorization theorem - II**). Some examples are shown. One interesting detail (slide **The factorization theorem - III**) is that the MLE has a sort of "built in" sufficiency: whenever a sufficient statistic exists, the MLE is a function of it.

Finally, although sufficient statistics are nice in that they allow us to compress the sample with no loss of information and narrow the search for good estimators, they not always exist in usable form. The slide **Some ill behaved distributions** re-visits the example of the Cauchy distribution with a location parameter θ, for which no useful form of sufficiency exists (try to apply the factorization theorem and see what happens).