Maximum Likelihood Estimation on Generalized Sample Spaces: An Alternative Resolution of Simpson's Paradox

Matthias P. Kläy and David J. Foulis

Received January 22, 1990

We propose an alternative resolution of Simpson's paradox in multiple classification experiments, using a different maximum likelihood estimator. In the center of our analysis is a formal representation of free choice and randomization that is based on the notion of incompatible measurements.

We first introduce a representation of incompatible measurements as a collection of sets of outcomes. This leads to a natural generalization of Kolmogoroff's axioms of probability. We then discuss the existence and uniqueness of the maximum likelihood estimator for a probability weight on such a generalized sample space, given absolute frequency data.

As a first example, we discuss an estimation problem with censored data that classically admits only biased ad hoc estimators.

Next, we derive an explicit solution of the maximum likelihood estimation problem for a large class of experiments that arise from various kinds of compositions of sample spaces. We identify the (categorical) direct sum of sample spaces as a representation of "free choice," and the (categorical) direct product as a representation of "randomization."

Finally, we apply the foregoing discussion to the case of multiple classification experiments in order to show that there is no Simpson's paradox if the difference between free choice and randomization is recognized in the structure of the experiment.

A comparison between our new estimator and the "usual" calculation can be summarized as follows: Pooling the data over one classification factor in the "usual" way in fact destroys or ignores the information contained in it, whereas our proposed maximum likelihood estimator is a proper marginal over this factor that "averages out" the information contained in it. The estimators agree with each other in the case of proportional sample sizes.

1 Work supported by the Swiss National Science Foundation.
2 Sandoz AG, CH-4002 Basel, Switzerland.
3 Department of Mathematics, University of Massachusetts, Amherst, Massachusetts 01003.
1. GENERALIZED SAMPLE SPACES

Consider the case where a person suffers from a certain illness for which there are two treatments, say $A$ and $B$, that cannot both be applied to one and the same patient, e.g., because of the intolerable side effects of the combination of medicaments $A$ and $B$. Assume that there are just two outcomes of a treatment, say “survival” and “death” of the patient. It is immediately clear that the treatments are incompatible in the sense that once we have applied one treatment to a patient and recorded its outcome, it is impossible in principle and in fact to determine a meaningful outcome for the other treatment. Even if the patient survives and is healed, such a determination is impossible since a basic premise for the measurement, namely the presence of the illness, is missing.

This kind of incompatibility of measurements is very common in the empirical sciences. It arises typically in cases of destructive testing, such as the determination of the breaking strength of a wooden board in the direction of its width and its breadth, or the determination of different component elements in an alloy. We find a famous instance of incompatibility of measurements in physics, where it is impossible to measure simultaneously the position and the momentum of one and the same elementary particle. “Not simultaneously measurable” does not mean that the measurements cannot be made at the same time. It means, as above, that the execution of one measurement precludes the determination of a meaningful outcome for the second measurement even in principle.

According to Kolmogoroff, an experiment is axiomatically represented by its sample space which consists of all possible outcomes of the experiment, and the collection of probability measures on the sample space. In the presence of incompatible measurements, there is not just one “grand canonical measurement” that produces all outcomes of the experiment, but there are several measurements, each capable of producing only a subset of the outcomes of an experiment. For example, denoting by $a$ the survival and by $\bar{a}$ the death of the patient under treatment $A$, and, likewise, by $b$ the survival, respectively by $\bar{b}$ the death, of a patient under treatment $B$, the collection of all possible outcomes is $X = \{a, \bar{a}, b, \bar{b}\}$, and there are the incompatible measurements $A = \{a, \bar{a}\}$ and $B = \{b, \bar{b}\}$. Thus, the complete description of the sample space consists not only of the outcome set $X$ alone, but also of the collection of measurements $\mathcal{A} = \{A, B\}$ as well.

Although this may look at first sight like a trivial reformulation of the simple two-sample experiment, we emphasize that the explicit inclusion of the structure of measurements in the sample space is a true generalization of Kolmogoroff’s axiomatic theory of probability. Its impact will become clear in more complicated experiments, described in the examples below.