STATISTICAL INFERENCE WITHOUT PROBABILITY PREREQUISITES

H. Rouanet*, J.-M. Bernard, M.-C. Bert, V. Duquenne, J.-L. Guigues,
B. Lecoutre, M.-P. Lecoutre, B. Le Roux, M.-C. Weiss

Groupe Mathématiques et Psychologie
U.E.R. M.L.F.I., Université René Descartes
12, rue Cujas. 75005 Paris

Paper read by H. Rouanet at the International Congress for the Teaching of Statistics. Sheffield, August 12, 1982.

Yet consider the following situation:

- the *data* consist in *one group of observations;*
- in addition to the data, there is a known, finite *reference population* available;
- and the question is raised to compare the data to the reference population from the view point of a certain variable, such as the mean, or the relative frequency of a given character.

Notice that we do not assume any relationship between the group of observations and the population; in particular, we do *not* suppose that the group of observations is a "random sample" of the population. Clearly this occurs very commonly in concrete situations.

As a simple illustration, we will take here the case of a *dichotomous variable,* that is, we will suppose that k out of the n observations possess a given character of interest, so that the observed relative frequency of the character in the data is $f = \frac{k}{n}$. In addition, we suppose that in the reference population of N individuals, the corresponding relative frequency is a known value $\frac{K}{N} = \Phi_o$.

In the sequel of this paper, I will treat the following numerical example:

$$n = 5 \qquad k = 4 \quad (n-k = 1) \qquad f = \frac{4}{5} = .8$$

$$N = 20 \qquad K = 6 \quad (N-K = 14) \qquad \Phi_o = \frac{6}{20} = .3$$

In this example, the observed frequency, $f = .8$, appears to be much higher than the reference value, $\Phi_o = .3$. Therefore, people will naturally tend to raise the question wether .8 can be declared to be "significantly higher" than .3.

This is of course an intuitive way of stating the significance testing problem, and the notion of what is meant by "statistically significant" remains to be conceptualized in some new sense, since no probabilistic considerations have been introduced as in the conventional formulation.

Now as it turns out, a conceptualization of the notion of "statistical significance" is feasible, resorting to purely set theoretic considerations.

Set theoretic significance testing

In order to proceed, in the foregoing situation, to a set theoretic significance testing, it suffices to redefine the notion of a sample as follows: a *sample* of size n from a population of size N is simply defined as a n-element subset of the population; as a consequence, the number of all samples will be the combinatorial number $\binom{N}{n}$. For example, with N = 20 and n = 5, there are $\binom{20}{5} = 15504$ samples.

Then among the samples, we may count how many there are for which the relative frequency of the character of interest is greater than (or equal to) the observed relative frequency f, that is, we count those samples which are more extreme, on the positive side, than the data, with respect to the relative frequency of the character of interest .

Thus in our numerical example, we will look for the samples of size 5 (subsets of a population of 20 individuals), for which the frequency is greater than or equal to $\frac{4}{5}$. From purely combinatorial considerations, there are:

$\binom{6}{4} \times \binom{14}{1} = 15 \times 14 = 210$ samples for which the frequency is $\frac{4}{5}$; and

$\binom{6}{5} \times \binom{14}{0} = 6 \times 1 = 6$ samples for which the frequency is $\frac{5}{5}$.

Hence there are 210 + 6 = 216 samples for which the frequency is greater than, or equal to $\frac{4}{5}$.

From the number of samples we derive the *proportion* of samples for which the frequency is greater than or equal to $\frac{4}{5}$, that is $\bar{p} = \frac{216}{15504} = .014$. The value $\bar{p}$ will be called the *observed upper level* of the set theoretic significance test.

Now suppose we choose a proportion α that will be called a (prefixed) level of significance.

Then if $\bar{p}$ is less than α, we will say that, in a set theoretic sense, the observed relative frequency f is significantly higher than ϕ_o at the α level.

For instance, at the level $\alpha = .05$, the frequency $f = \frac{4}{5}$ is significantly higher than $\phi_o = .3$ (for a population of 20 individuals), since $\bar{p} = .014$ is less than .05; at the level $\alpha = .01$, this same frequency f is not significant, since .014 is greater than .01, etc.

Naturally, as you can see, the numerical calculations are just the same as in the conventional significance test, that is, in the present case, the classical hypergeometric significance test; whereas the justification and interpretation framework is purely set theoretic.

Even though theoretic significance testing seems paradoxical, at first, it amounts to a direct extension of one of the most familiar procedures of descriptive statistics, namely, evaluating a subject's performance vis à vis a reference population of scores, by taking the proportion of population scores that are higher than the score of the subject under examination; in this familiar procedure, the population is simply used as a reference, and no probabilistic judgment is implied. For $n = 1$, the set theoretic significance testing clearly reduces to this description procedure.

In this paper, I have sketched the set theoretic approach for significance testing, but it can be worked out as well for confidence limits. This approach appears to be both interesting intrinsically and attractive from a pedagogical viewpoint, since it can be taught immediately after a course on descriptive statistics and independly from a probability course.

As we all know, the acquisition of probability notions is always a long and laborious task, even with students with a good mathematical background. It is thus worthwhile to have at our disposal an approach of statistical inference which entirely avoids resorting to probabilistic considerations.

The set theoretic notions required for this approach are very elementary and can be restricted to the following ones; subset of a set, relation on a cartesian product of sets, mapping from one set to another.

On the other hand, the set theoretic approach appears to provide an excellent preparation to the usual probabilistic statistical inference. Once it is mastered, *random sampling situations* can be introduced in a natural way and are grasped more easily. When, in particular, *simple random* sampling is assumed, where all samples are given equal probabilities, the "proportion of samples for which the test statistic exceeds the observed value" becomes the familiar "probability that the test statistic exceeds the observed value under H_0". This equivalence property brings an insightful bridge between set theoretic and probabilistic statistical inference.

For more than two years now, we have tested on a large scale the set theoretic approach of statistical inference, both with undergraduate and graduate psychology students of our university. A programmed text-book that presents the normal distribution along these lines has been published(*), and more extensive publications are now under way.

(*) M.-P. Lecoutre & B. Lecoutre, Enseignement programmé sur l'utilisation d'une table de la distribution normale, Editions C.D.U.-S.E.D.E.S., Paris, 1979.