Summary

Distributional inference: the limits of reason

Science advances by combining rational arguments and empirical information. In fields like philosophy and pure mathematics, emphasis is laid on the rational arguments, whilst in the applied sciences the collection and interpretation of data are the field of interest. In mathematical statistics one tries to combine these aspects. The primary goal is to make statistical inferences about something unknown. Such inferences can be of help in further discussion, e.g. in selecting a decision. The methods should not depend on 'the intentions that might be furthered by utilizing the knowledge inferred'[1]. When the available data are too limited, then different procedures may yield different inferences. The statistician should refrain from providing a specific inference in case the differences are 'too large'. When such an inference can be given, this inference should be accompanied by a statement about the uncertainty of the inference. This could be done by providing a distributional inference, or by providing the results of different approaches.

An example is as follows. Ornithologist G.Th. de Roos is observing a population of Ruddy Turnstones (*Arenaria Interpres*) on the Frisian island Vlieland. Some of these birds are ringed, however the ring-number is not always observable, e.g. because another bird is blocking the view. After how many days of observing is it safe to assume that all ringed birds in the population have been observed at least once? This question can be answered by constructing a distributional inference about the number of present, yet unseen, ringed birds, including a probability statement about the hypothesis that all ringed birds have been seen. Of course, the results depend in some way on the probabilistic assumptions one makes, and on the statistical principles one follows.

The first part of this thesis consists of 'finger exercises' illustrating that information about the unknown can only be of value if the mechanism generating the information is (sufficiently well) known. In probability theory, information is incorporated by conditioning to it. This generates difficulties in statistical practice, because unknown aspects are involved in the joint distribution of the random variables X and Y that are behind the observations x and the unknown y. Firstly, this is extensively exemplified by a die-rolling game. From the information 'the number of eyes is even' one cannot conclude automatically: 'the probability that a six has been thrown, equals one third'. The way in which the source of information operates, should be incorporated in

[1] R.A. Fisher, *Statistical Methods and Scientific Inference*, third edition, Macmillan, New York, 1973, p. 107