THE BAYESIAN APPROACH TO THE REJECTION OF OUTLIERS

BRUNO DE FINETTI
UNIVERSITY OF ROME

1. Summary

The aim of this paper is twofold: first, to contribute to the subject in question; second, in so doing to illustrate some general features and techniques of the Bayesian (or neo-Bayesian) subjectivistic approach to problems of statistics.

What may be said about the "true value" X of a quantity after observations have yielded measurements $x_1, \cdots, x_n$? The variable X or "unknown parameter" X, as the usual terminology calls it, is in fact a *random number* in the subjectivistic probability theory. For in this theory, probability is nothing else than the expression of beliefs about unknown facts; for example, this may be the probability that X is greater than some given value x. We need then merely to show how our final probability distribution for X, after the observations, comes from the initial one and from the conditional distributions of the measurements $x_1, \cdots, x_n$ given $X = x$, according to Bayes' theorem. No estimation problem per se is acknowledged to exist from such a viewpoint. X has a final distribution, and if such an expression as "estimated value" is used at all, it should be conceived of as a measure of location of the final distribution, suitably chosen for some practical purpose.

Everything is embodied in this over-all formulation. A rational answer to the question of how and why to attach less or no "confidence" to some "outlying" observations can arise from nothing else.

To fill the gap between such an abstract (or, as some might perhaps say, "philosophical") formulation, and a realistic detailed analysis of practical situations, we need only consider some set of more or less "reasonable" and interesting hypotheses about our opinions concerning the process of taking observations.

Three cases, all concerning an "error distribution," will be studied: (a) the errors are independent; (b) the errors are exchangeable; (c) the errors are partially exchangeable.

(a) *Independence* means "independence with known error distribution"; if the distribution is not normal, the combination of the observations is no longer so simple, and particular problems arise for "outlying" observations. This case has been considered by Poincaré [5] and others. More recently, in a paper of mine [1] at the 1957 Meeting of the Societá Italiana di Statistica, the case of mixtures of normal centered distributions has been particularly stressed.

(b) *Exchangeability* translates "independence with unknown error distribu-