ON THE INADMISSIBILITY OF SOME STANDARD ESTIMATES IN THE PRESENCE OF PRIOR INFORMATION[1]

By M. Skibinsky[2] and L. Cote

University of California, Berkeley and Purdue University

0. Introduction. A common formulation of statistical decision problems involves a sample space and a class of probability distributions on this space indexed by a parameter θ. The loss consequent upon action to resolve the problem is regarded as a function of θ. Thus in a sense, to choose a proper action, we must "know about θ". Ordinarily, the parameter is not considered to be random. When this view prevails, the decision rule may be chosen so as to keep the expected loss from assuming its worst possible value. In many physical problems, however, certain extreme values of the parameter, though not disallowed completely, are held by the experimenter to be rather unlikely. Because such prior information is rarely used to formulate the decision procedure (it may be judged too crude), experimenters often find they must modify its recommendations to suit their judgement. Some statisticians may accept this with good grace, but the existence of such disparity with its attendant misunderstandings provides a powerful incentive to reformulate the standard procedures so as to allow for a more efficient and exacting use of prior information. For an interesting general discussion of basic theories involved here, see De Finetti (1951).

In the first example which follows, we take into account prior information about the probability of success in a sequence of independent Bernoulli trials. We are to estimate this probability from an observation on the number X of successes in n such trials. The conventional estimate is of course X/n. We will assume below that the probability of success in our trials is the value of a random variable Θ. This value is what we would like to "know about". If nothing is known about the distribution of Θ, we can do no better than the usual formulation. We will assume, however, that Θ has a distribution which belongs to a subclass of the distributions on [0, 1] roughly conforming to the type of prior information that an experimenter might have; e.g., that values of Θ near 0 or 1 are unlikely. If we now regard the binomial distribution to be conditional on Θ, the members of this subclass generate a family of joint distributions for X and Θ. With this as background we may view our problem as a special case of conventional prediction theory. In the following paragraphs, we present a generalized maximum likelihood principle as applied to this example and investigate a class of predictors which it suggests. Under appropriate conditions to be discussed below each of these has a uniformly smaller mean square error than the conventional estimate.

Received July 17, 1962.

[1] Prepared with the partial support of the National Science Foundation, Grant G-14648 (for M. Skibinsky).

[2] Now at the University of Minnesota.