

# 1. Introduction

There are at least four reasons for being interested in the estimation of probabilities:

1. For betting, including insurance decisions and for all other decisions and predictions.
2. For discrimination between hypotheses, when these hypotheses are not simple statistical hypotheses. (When the hypotheses are “simple,” the probabilities of relevant events have uncontroversial values.) This second reason can be understood in the light of examples mentioned later.
3. For deciding whether an event is surprising. If it is surprising, the correctness of the observations should be checked.
4. For the improved understanding of the philosophy of probability and statistics—and of the nature of probability judgments.

The basic theme in the following pages is the estimation of probabilities from *effectively small samples*. By “effectively small,” we mean that the sample is small for the purpose of estimating some probability, though the sample might be absolutely large. The problem is not quite the same as that of estimating small probabilities, in which part of the technique is to enlarge the sample. (See, for example, Haldane<sup>47, 48</sup> and Pearson.<sup>90</sup>) We shall suppose for the most part that the sample already available is to be used by itself for the estimation. Circumstances are always changing, but humans have the facility of estimating the probabilities of many events that have never previously occurred. Their

sample is their past experience, and they must often make predictions without the benefit of an increased sample. Suppose, for example, that we wish to estimate the probability that there will be a storm here tomorrow. It might be suggested that the probability could be taken as the proportion of days in February on which there are storms, or of those days on which there are storms when there has been no storm for 24 hours and the temperature at noon was 53 degrees and at 2:30 p.m. was 51.7 degrees. The totality of information is such that, if it were all used, the proportion would be of the indeterminate form  $0/0$ ; this would be true even if we ignored all the information that we judged to be virtually irrelevant or, at any rate, if this were done in any obvious manner. This difficulty would remain even if we could give a precise definition of a storm. The difficulty is well known to actuaries.

Nevertheless, for the purpose of making decisions, we do manage to make approximate estimates of probabilities. How this is done is an interesting problem in psychology and in neurophysiology. It might, for example, be conjectured that neural circuits automatically use a maximum-entropy estimation (see Chapter 9). The problem of estimating probabilities of events that have never occurred is philosophically interesting and is, in my opinion, likely to be important for the design of ultraintelligent machines. The unconscious goal of the scientific philosopher is the automation of science.

A completely general solution of the problem of estimating probabilities seems entirely out of reach, and this book therefore deals only with a number of simple problems: binomial and multinomial estimation, multinomial sampling when the number of categories is very large (the "sampling of species"), estimation of small probabilities in a large pure contingency table, and in a multidimensional population contingency table.

The estimation of probabilities leads directly to the discussion of other matters, such as tests for no association in a contingency table when the sample is not necessarily large, estimates of the "coverage" of a sample of species or words, and the inconsistency of maximum-likelihood estimation with a Bayesian philosophy.

Each distinct problem of probability estimation is liable to need somewhat different methods, but there are several unifying themes in this monograph. In particular, we rely mainly on a modern Bayesian approach. Some of the methods involve the notion of a Type III distribution (in a sense described in Chapter 2, not in Karl Pearson's sense!). Sometimes the initial distributions have parameters in them, and can be regarded as "semi-initial" after the parameters are estimated. The method mentioned in Chapter 8 is Bayesian but uncontroversial. There is some discussion of the principle of maximum entropy for

generating null hypotheses, and of invariance theories. Included is the use of the device of imaginary results, which involves Bayes' theorem in reverse, of methods that seem to be *good enough*, and of methods that *improve* on previous methods; since final answers do not seem to be attainable even in the simplest of problems. My own view, following Keynes and Koopman, is that judgments of probability inequalities are possible but not judgments of exact probabilities; therefore a Bayesian should have upper and lower betting probabilities. In order to avoid complications, this viewpoint is de-emphasized in this monograph. We stress at this point, however, that a convenient approximate method of using this philosophy is to make use of two or more initial distributions, in effect, as several alternative models. This will give rise to a number of final betting probabilities, the largest and smallest of which would be the upper and lower betting probabilities.

In all honesty, the Bayesian will wish to ascribe different weights to different initial (or Type II) distributions. These weights will again typically lie in intervals of values, but, for simplicity, they can be taken to be precise numbers. They then act as a Type III distribution, but the weighted average of the set of Type II distributions, weighted with the Type III distribution, is in effect a single Type II distribution. Thus the Bayesian can use the Type III distribution merely as an aid to his own judgment, and is not compelled to state the Type III distribution aloud. But to do so helps to explain what otherwise might seem to be a peculiar and unnatural-looking Type II distribution. The main example of this method in this book is in connection with initial Type II distributions that are of the symmetrical Dirichlet form.

An alternative to the use of a Type III distribution is to plot a graph or make a table of the Type II likelihoods. This is an example of a compromise between Bayesian and non-Bayesian methods, since one of the techniques used by the non-Bayesian is to graph the ordinary (Type I) likelihoods rather than making a weighted combination of them as in the usual Bayesian method.

If a unique initial distribution (a "credibility" distribution) could be generally agreed, for any class of estimation problems, then there would be no need to make use of distributions of Type III. Such agreement has not yet been reached, and I do not believe that it ever will be. (Compare Good,<sup>27</sup> p. 48.)

Although this book is largely a survey, most of it has not previously been published; it has seemed to me to be more valuable to state old results and prove new ones rather than to describe in detail what is already conveniently available.

The reader might find it useful to read the Summary (Chapter 10) before proceeding to Chapter 2.