ON THE CORRECT USE OF BAYES' FORMULA

By R. v. Mises

Harvard University

The problem that we try to solve by using Bayes' formula consists in making an inference from an observed statistical value upon the unknown value of a parameter, and in examining the chance of this inference being correct. One may call this the principle problem of practical statistics or the estimation problem, or, as the author put it in German (Rueckschluss-Wahrscheinlichkeit) problem of inference probability; at any rate we encounter this kind of problem in various forms in almost every branch of statistical investigation. It will be convenient to base the following discussion on a concrete question in quite specified form which will allow us to see clearer the points that are to be stressed in this paper.

1. The problem. In examining the quality of water supplies with respect to the number of bacterias of a certain kind they contain, a definite procedure is usually adopted. One takes $n = 5$ samples out of the water, each sample of exactly 10 ccm. Then by a certain biological test one finds out whether or not each sample contains at least one bacteria of the kind under consideration. The number x (zero to five) of positive tests is the observed value from which an inference is drawn upon the probability θ for a sample containing at least one bacteria. It is assumed that this θ is connected with the average number λ of bacterias per 10 ccm by

$$\theta = 1 - e^{-\lambda}; \qquad \theta = \theta_1 = 0.63 \qquad \text{for } \lambda = 1 \tag{1}$$

according to Poisson's law. A particular question which we want to answer is this: What is the chance of being right, if we conclude from the observed fact $x = 0$, (in other cases from $x = 1$) that θ lies between 0 and $\theta_1 = 0.63$ (or λ between 0 and 1)?

For a given θ the probability of getting x positive tests out of n tests is according to Bernoulli's formula

$$p(x \mid \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}. \tag{2}$$

The chance of having a θ-value between 0 and θ_1 when x positive tests are observed is according to Bayes' formula

$$P_x(\theta_1) = \frac{\int_0^{\theta_1} p(x \mid \theta)\, dP(\theta)}{\int_0^{1} p(x \mid \theta)\, dP(\theta)} \tag{3}$$