C. v. Eeden.

On distribution free ...

ON DISTRIBUTIONFREE BIO-ASSAY

CONSTANCE VAN EEDEN
Mathematical Centre, Amsterdam, The Netherlands

In bio-assay the following problem is considered: a stimulus (e.g. a drug or a vitamin) is applied to a subject, resulting in a response produced by the subject. In this paper we only consider the quantal response, i.e. the case where the subject gives or does not give a response. Then for each subject there will be a level of intensity of the stimulus above which the response occurs and below which it does not occur; this level is called the *tolerance* of the subject and will be denoted by λ. So if d is the dose of the stimulus the subject will give a response if $d \geqq \lambda$ and will not give a response if $d < \lambda$.
On a population of subjects the tolerance will be a random variable, which is denoted by underlining its symbol ($\underline{\lambda}$); the distribution-function of $\underline{\lambda}$ is denoted by $F(\lambda)$. Then if d is the dose of the stimulus $F(d) = \mathrm{P}[\underline{\lambda} \leqq d]$ is the probability of a response at dose d.

One of the problems in bio-assay is to give an estimate of $F(\lambda)$. The observations consist of the results of applying the stimulus once to each of a number of subjects at several doses. A second problem is to compare the distributionfunctions $F_1(\lambda)$ and $F_2(\lambda)$ of two different stimuli.

These problems of estimating and comparing distributionfunctions have been solved for the parametric case, e.g. when $\underline{\lambda}$ has a normal distribution; the methods given for this situation are called probit analysis and are e.g. described by D. J. FINNEY (1947).

The application of this method, however, requires in many cases laborious computational work. Moreover difficulties arise if, for one or more doses, the number of observations is very small or if none of the subjects or all subjects give a response. Finally the assumption of normality may not be fulfilled.

We will now consider the case where no assumptions are made on the form of $F(\lambda)$. Let observations be available at k different doses $d_1, d_2, \ldots, d_k$, satisfying $d_1 < d_2 < \ldots < d_k$. The number of ob-

servations at dose d_i is denoted by n_i and the number of subjects giving a response by a_i. We suppose all observations to be independent. Let further p_i denote the probability of a response at dose d_i then $p_i = P[\underline{\lambda} \leqq d_i] = F(d_i)$. Now $F(\lambda)$ is a distributionfunction, so $F(\lambda)$ is a monotone nondecreasing function of λ; consequently

$$p_1 \leqq p_2 \leqq \ldots \leqq p_k. \tag{1}$$

For this situation of k probabilities satisfying the inequalities (1) the maximum likelihood estimates of $p_1, p_2, \ldots, p_k$ may be obtained by means of a method described in my thesis. In this thesis a more general problem is considered, where k parameters of k distributionfunctions are to be estimated, if it is known that these parameters are partially or completely ordered and moreover confined to given intervals. If the distributionfunctions are binomial, if the ordering is complete and if the given intervals are the interval [0, 1] we obtain as a special case the problem of estimating k probabilities satisfying the inequalities (1).

The estimates may be obtained in a simple way, which may be described as follows: first compute the "ordinary" maximum likelihood estimates f_i of the p_i, i.e. the maximum likelihood estimates without any restriction on the p_i. Then $f_i = a_i/n_i$. If these estimates f_i satisfy the inequalities $f_1 \leqq f_2 \leqq \ldots \leqq f_k$, the f_i are the maximum likelihood estimates of the p_i under the restrictions $p_1 \leqq p_2 \leqq \ldots \leqq p_k$. If a value of i exists with $f_i > f_{i+1}$ the i^{th} and $(i+1)^{\text{th}}$ sample are pooled and again the "ordinary" maximum likelihood estimates are computed, treating the i^{th} and $(i+1)^{\text{th}}$ sample as one sample. This procedure is carried out until a set of l samples ($l < k$) is obtained for which the ordinary maximum likelihood estimates satisfy the restrictions imposed on the p_i.

This method may be illustrated by means of the following example. Suppose $k = 4$ and

i	1	2	3	4
a_i	4	4	5	9
n_i	10	8	12	12

then we have

i	1	2	3	4
$f_i = a_i/n_i$	0.40	0.50	0.42	0.75

and for $i = 2$ we have $f_i > f_{i+1}$. Consequently the second and third sample are pooled and we obtain

i	1	{2, 3}	4
a_i	4	9	9
n_i	10	20	12
$f_i = a_i/n_i$	0.40	0.45	0.75

Thus $f_1 \leqq f_{\{2,3\}} \leqq f_4$; consequently the maximum likelihood estimates t_i of the probabilities p_i under the restrictions $p_1 \leqq p_2 \leqq p_3 \leqq p_4$ are

$$t_1 = 0.40,\ t_2 = t_3 = 0.45,\ t_4 = 0.75.$$

In this way a maximum likelihood estimate of $F(\lambda)$ at the doses used in the experiment is obtained without any assumption of a parametric nature about the distributionfunction $F(\lambda)$. From the example it may be seen that the estimates may be found by quite simple calculations. Further it is not necessary to have a large number of observations at each dose and no difficulties arise if for one or more doses none of the subjects or all subjects give a response.

References

VAN EEDEN, C., Testing and estimating ordered parameters of probability distributions, Thesis Amsterdam (1958)

FINNEY, D. J., Probit Analysis, Cambridge University Press, Cambridge (1947)

Discussion

P. ARMITAGE: I should like to ask Mrs. van Eeden whether her method can be adapted for the estimation of relative potencies, and also whether she has seen the recent work of Bartholomew, published in Biometrika, which seems closely related to her own work. Dr. Hemelrijk, I think, implied that data with arbitrarily chosen doses could not be handled by parametric methods. In fact the probit and other methods can be used for such data, and Finney has published details.

C. VAN EEDEN: In answer to the first question I would say that this paper is to be considered as a first step in applying the results obtained in my thesis to problems of bio-assay. Up to now I obtained the estimates for the distribution-function $F(\lambda)$ at the doses used in the experiment. I hope it will be possible to obtain further results such as an estimate and a confidence interval for an ED_{50} or for the differences of two ED_{50}'s.

I have read the paper of Bartholomew in Biometrika (1959), 36—48 and 328—335. His work is indeed closely related to mine. He described a test for the hypothesis that k means $\mu_1, \mu_2, \ldots, \mu_k$ of normal distributions are equal against the alternative hypoth-

esis that these means are monotone nondecreasing, whereas the fourth chapter of my thesis deals with a test for the hypothesis H_0 that k parameters of k distribution functions are monotone nondecreasing against the alternative hypothesis that H_0 is not true. For this test Bartholomew needs the estimates of the μ_i under the restriction that these means are monotone non-decreasing and this section of his work is a special case of my estimation problem with all the distributions normal, the ordening complete and all given intervals the interval $(-\infty, \infty)$.

L. MARTIN: When you pool the 2nd and 3rd group, at which abscissa do you plot the pooled $p_{\{2,3\}}$?

Is it possible to estimate a kind of ED_{50} and some confidence interval?

C. VAN EEDEN: In pooling the second and third group you obtain an estimate for p_2 *and* p_3, i.e. the estimates of p_2 and p_3 are equal in this example. So the pooled $p_{\{2,3\}}$ is plotted at d_2 and at d_3.

For the second question concerning estimates and confidence intervals for an ED_{50} I refer to my answer to the first question of Dr. Armitage.