A MINIMAX ESTIMATOR FOR THE LOGISTIC FUNCTION

JOSEPH BERKSON[1]
MAYO CLINIC

AND

J. L. HODGES, JR.[2]
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

One of us has discussed the use of the logistic function

$$P_i = 1 - Q_i = \frac{1}{1 + e^{-(\alpha+\beta x_i)}} \tag{1.1}$$

as a model for analyzing bioassay or other experiments with "quantal" response, and has studied the problem of estimating the parameters α and β, in several papers (see [1] and the references given therein). At each of k dose levels $x_1, \cdots, x_k$ we perform n trials. It is assumed that the number R_i of responses among the trials at dose level x_i has the binomial distribution with probability of response P_i given by (1.1), and that all trials are independent. On the basis of the observed values r_i of the random variables R_i we wish to estimate the parameters. The problem also arises in the one-parameter form, where β is taken to be known and only α is to be estimated. It is solely with this one-parameter problem that we are here concerned.

Various estimators for α may be proposed, on the basis of various general principles, such as those of maximum likelihood and minimum chi-square, these having certain desirable asymptotic properties. But realistically, what is of interest is the performance of the estimates for small or moderate sample sizes. In studying the actual performance of the estimates, it was considered necessary to use specific numerical values for the quantities involved. In [1] the mean square errors of several estimators were computed and compared for $n = 10$, $k = 3$, $x_1 = -1$, $x_2 = 0$, $x_3 = 1$, and with $\beta = \log (7/3)$. This value of β was chosen so that when $\alpha = 0$ we have $P_1 = 0.3$, $P_2 = 0.5$, $P_3 = 0.7$. Five different values of α were taken, corresponding to $P_2 = 0.5, 0.6, 0.7, 0.8$, and 0.85, and

[1] This research was supported in part by the National Science Foundation.
[2] This investigation was supported in part by a research grant (No. RG-3666) from the National Institutes of Health, Public Health Service.

the mean square errors found for the maximum likelihood estimate, minimum Pearson chi-square estimate, and minimum logit chi-square estimate, and to the estimate obtained from the last by the averaging technique of Blackwell and Rao. This estimate, the "Rao-Blackwellized minimum logit chi-square estimate," we shall denote by B. The mean square errors of these estimates were compared with each other, and with the so-called "information limit" $1/I(\alpha)$, where $I(\alpha)$ is the quantity called by Fisher the amount of information provided by the experiment.

The total number of successes $R = \sum R_i$, which is a sufficient statistic, has the frequency function

$$\phi(r, \alpha) = e^{\alpha r}(Q_1 Q_2 Q_3)^n \sum \left(\frac{7}{3}\right)^{r_3 - r_1} \binom{n_1}{r_1}\binom{n_2}{r_2}\binom{n_3}{r_3}, \tag{1.2}$$

where the sum is extended over all r_1, r_2, r_3 for which $\sum r_i = r$. The quantity

TABLE I

VALUES OF THE FREQUENCY FUNCTION $\phi(r, 0)$ AND OF SEVERAL ESTIMATES
The numbers in parentheses indicate power of 10

r	$\phi(r, 0)$	B	H_1	H_2	H
15	1.53132(−1)	0	0	0	0
16	1.42383(−1)	.13840	.137	.134	.134
17	1.14393(−1)	.27382	.275	.269	.270
18	7.92859(−2)	.41316	.415	.405	.407
19	4.72756(−2)	.55560	.557	.547	.548
20	2.41513(−2)	.70256	.704	.686	.692
21	1.05109(−2)	.85580	.856	.833	.842
22	3.86735(−3)	1.01739	1.016	.987	.997
23	1.19081(−3)	1.18973	1.185	1.149	1.163
24	3.02683(−4)	1.37583	1.367	1.323	1.324
25	6.23369(−5)	1.57963	1.567	1.513	1.540
26	1.01316(−5)	1.80647	1.792	1.725	1.744
27	1.24973(−6)	2.06360	2.048	1.980	1.985
28	1.09862(−7)	2.35824	2.379	2.300	2.262
29	6.12775(−9)	2.67949	2.821	2.741	2.664
30	1.62889(−10)	3.37000	3.370	3.370	3.366

$\phi(r, 0)$ is given in table I. From this, $\phi(r, \alpha)$ can be calculated easily by means of

$$\phi(r, \alpha) = \phi(r, 0) \frac{e^{\alpha r}(Q_1 Q_2 Q_3)^n}{[(0.3)(0.5)(0.7)]^n}. \tag{1.3}$$

We define $D(r, \alpha) = \partial \log \phi(r, \alpha)/\partial \alpha$. Then $I(\alpha)$ is the variance of $D(R, \alpha)$, easily calculated to be

$$I(\alpha) = n \sum P_i Q_i. \tag{1.4}$$

The values of B, the Rao-Blackwellized estimate, corresponding to $r = 15$ (1) 30, are given in table I. For values of $R < 15$ we have by symmetry $B(R) = -B(30 - R)$. It was found in [1] that B had the smallest mean square error among the estimates compared for all values of α which were explored. Moreover the ratio of the mean square error $E(B - \alpha)^2$ to $1/I(\alpha)$ was less than unity and in fact never exceeded 0.9. Although this in itself appeared paradoxical to some statisticians, while others accepted it as the limit of achievement of a small mean square error, certain aspects of the analysis suggested the possibility of constructing an estimator with even a smaller mean square error. It is with the exploration of this possibility that the present paper is concerned.

(We remark parenthetically that the value of this ratio, if calculated for still larger values of α than that corresponding to $P_2 = 0.9$, which was as far as the investigation [1] was extended, is found to rise again, reaching a maximum of 1.15 at $\alpha = 4.8$ corresponding to $P_2 = 0.992$. However, this is entirely due to the value of the estimate corresponding to $R = 30$, or 100 per cent response at each x. The estimates for such responses were determined in [1] by a rule the specific form of which is more or less arbitrary, the so-called $2n$ rule. If for the particular sample $R = 30$, we let $B(30) = 3.37$ instead of the value 2.9444 used in [1], the ratio has a maximum of 0.894 attained near $\alpha = 1$. Values of α beyond 2 (for which $P_2 = 0.88$), not to say values beyond 4, are of no practical interest, and no statistician would think of trying to estimate α from an experiment with 100 per cent response at each dose level. However, the mathematics is simplified if we permit the infinite range of α, and we accordingly define $B(30) = 3.37$. The values of this ratio, presently to be defined as F_B, are shown in table II.)

TABLE II

VALUES OF THE VARIANCE, THE BIAS, AND THE RISK FUNCTION

P_2	α	$I(\alpha)$	F_B	$-b_0$	$-b_1$	$-b_2$	$-b_3$	F_2	F_2'	F
.5	0	6.7	.8778	0	0	0	0	.8380	.8380	.8497
.6	.40547	6.51025	.8818	.0339	.0254	.0351	.0322	.8371	.8371	.8509
.7	.84730	5.91094	.8935	.0760	.0559	.0782	.0712	.8397	.8395	.8502
.8	1.38629	4.80096	.8820	.1393	.1018	.1428	.1234	.8643	.8515	.8492
.85	1.73460	3.99466	.8393	.1920	.1403	.1931	.1848	.8919	.8696	.8465
.9	2.19722	2.96883	.7816	.2817	.2099	.2737	.2781	.9040	.8661	.8469
.95	2.94444	1.66484	.7174	.4887	.4093	.4731	.5045	.8120	.8531	.8421
.98	3.89182	.71584	.7859	.8700	.9008	.9426	.9771	.7739	.8682	.8335
.99	4.59512	.36688	.8297	1.1882	1.4230	1.4480	1.4725	.8130	.8691	.8463
.995	5.29330	.18574	.7922					.7839	.8067	.7985

Assuming the general objective to be the attainment of a small mean square error for our estimator, some relevant considerations invite attention before a a definite plan can be laid out. If we accept the quantity $I(\alpha)$ as measuring, in some relative sense, the amount of effective data available for making the estimate of α, we may note (table II) that it is largest for $\alpha = 0$, and is progressively smaller for larger values of α. There is, so to speak, better data to work with when α is near zero. This is reflected in the general recognition that in a bioassay experiment, the function in question can best be estimated if the dosages are symmetrically disposed around the position where $P_2 = 0.5$, and that dosage arrangements very far removed from this are prohibited. This suggests that the mean square error should be considered in relation to $1/I(\alpha)$, and if the estimator is to be used in a range of conditions in which $I(\alpha)$ is not always the same, what we should seek to minimize is the product $I(\alpha)E(T - \alpha)^2$, where T is the estimate. We shall refer to the quantity $I(\alpha)E(T - \alpha)^2$ as the "risk function" of the estimate T, and denote it by $F_T(\alpha)$. In [1] the denominator $1/I(\alpha)$ of $F_T(\alpha)$ was thought of as the lower bound of the variance of an unbiased estimator. It can also be thought of as the asymptotic variance of an asymptotically efficient estimate. $F_T(\alpha)$ can be considered as the mean square error standardized to these basic variances. Now, it is not to be assumed that $F_T(\alpha)$ will be the same for different values of α. If it is large for some values of α and small for others, which shall we try to minimize? Obviously we cannot favor some one particular α, since we do not know what the true value of α will be for any experiment, the object of the experiment being just to estimate it. A principle which is applicable in these circumstances is to avoid large values of the risk. Utilizing this principle for the elaboration of a method, we shall seek to minimize the maximum value of $F_T(\alpha)$, which is to say we will apply the minimax principle utilized extensively in the works of Wald.

(We note, however, that α is a function of P_2, that is, α = logit of P_2, which is subject to control. Experiments as actually performed for an acceptable bioassay are not done with P_2 far from $P_2 = 0.5$. If an experiment yields an estimate of P_2 very far from 0.5, it is discarded and the lattice of dosages is changed toward an acceptable arrangement. Thus, in practice, we can virtually be sure that α is not extremely large.)

In contriving B, an estimate was achieved whose risk F_B never exceeded 0.894, and we wish to develop an estimator which will do better. The question naturally arises: Just how low is it possible for the maximum risk to be pushed? A main purpose of this paper is to illustrate a computational technique that gives an answer to this question. In section 2 we show that no estimate can have a maximum risk lower than 0.84. As a by-product of this investigation, we are able in section 3 to produce by direct attack an estimate with 0.85 as its maximum risk.

The computational methods, both of producing a lower bound for the maximum risk and of providing an estimate with low maximum risk, are straightforward, if somewhat lengthy. We believe they may be used profitably in other one-parameter estimation problems of practical interest.

2. A lower bound for the maximum risk

Our technique is the method of differential inequalities introduced in [7], which is based on the famous lower bound for the variance of an estimate T,

$$\sigma_T^2 \geqq \frac{[1 + b_T'(\alpha)]^2}{I(\alpha)}, \tag{2.1}$$

where $b_T(\alpha) = E_\alpha(T) - \alpha$ is the bias function of the estimate T, and $b'_T(\alpha)$ is the derivative of $b_T(\alpha)$ with respect to α. This formula is an immediate consequence of the fact that the correlation must be between -1 and 1. Indeed, $I(\alpha)$ is the variance of $D(R, \alpha)$, while $1 + b'_T(\alpha)$ is the covariance of $D(R, \alpha)$ and $T(X)$. The lower bound for the variance of estimates has a long history, going back at least to the memoir of Pearson and Filon in 1898 [8] and including the notation by Edgeworth [5]. As far as we know, the formula (2.1) was first published by Fréchet in 1943 [6]. It was developed independently by Rao [9] and by Cramér [4]. From formula (2.1) it follows that the risk function must satisfy

$$F_T(\alpha) \geqq I(\alpha)b_T^2(\alpha) + [1 + b'_T(\alpha)]^2. \tag{2.2}$$

Since R is a sufficient statistic, we need to consider only estimates which are functions of R. Furthermore, because of the symmetry of the problem we can restrict ourselves to symmetric estimates, that is, to estimates satisfying the equation

$$T(r) + T(3n - r) = 0. \tag{2.3}$$

This may be argued as follows. Suppose an estimate T does not satisfy (2.3). Let $T'(r) = -T(3n - r)$. Then $F_{T'}(\alpha) = F_T(-\alpha)$. The estimate $T^*(r) = [T(r) + T'(r)]/2$ is symmetric, so that $F_{T^*}(\alpha) = F_{T^*}(-\alpha)$. Since the average of two estimates has a mean square error not greater than the larger of the mean square errors of the estimates averaged, we have $F_{T^*}(\alpha) \leqq \max[F_T(\alpha), F_{T'}(\alpha)] = \max\,[F_T(\alpha), F_T(-\alpha)]$. Thus the maximum risk of T^* cannot exceed the maximum risk of T. In seeking estimates with low maximum risk, we therefore need consider only symmetric estimates. They have the important property of being unbiased at $\alpha = 0$.

Let us replace the differential inequality (2.2) by the differential equation

$$c = I(\alpha)b_T^2(\alpha) + [1 + b'_T(\alpha)]^2, \tag{2.4}$$

where c is a positive constant less than 1. This equation, together with the side condition $b(0) = 0$. determines the function b at least over a certain interval centered at the origin. The solution starts with $b'(0) = \sqrt{c} - 1$; this determines $b(\alpha)$ for small α, where (2.4) then determines $b'(\alpha)$, and so forth. If we set the constant c too small, $-b'(0)$ will be too large, and the negative bias will grow too rapidly, until at a finite point, say $A(c)$, we shall have $b'[A(c)] = -1$ and $b[A(c)] = \{c/I[A(c)]\}^{1/2}$. Then the solution can proceed no further, showing that (2.4) does not have a solution over the infinite range. On the other hand, if c is large enough, $-b(\alpha)$ will grow slowly enough so that $I(\alpha)$, which tends to 0, can eventually suppress the term $I(\alpha)b^2(\alpha)$, and we get a solution for all α. Let c_0 denote the dividing point, so that (2.4) can be solved for all α when $c > c_0$, but not when $c < c_0$.

The value of c_0 can be determined by numerical integration to any desired degree of accuracy. We try $c = 0.8$, and find that (2.4) rapidly explodes, showing $0.8 < c_0$. We try $c = 0.85$, and obtain a solution, so that $c_0 \leqq 0.85$. Successively we find that $c = 0.83$ fails but $c = 0.84$ succeeds. As two figures is adequate for our purposes, we take $c_0 = 0.84$. The corresponding bias function b_0 is shown in table II. Since we can with each integration bisect the interval in which c_0 is known to lie, it does not take many trials to pin down c_0 with considerable accuracy. In fact, three more integrations will show that $0.836 < c_0 < 0.837$, though the two-figure value 0.84 is used in the sequel.

The significance of c_0 is that it is a lower bound for the maximum risk of all estimates, so that the numerical work just described tells us that no estimate for α can have a maximum risk as small as 0.83 (more accurately, 0.836). Since the estimate B has maximum risk $F_B = 0.894$ (table II) we now know that it cannot be greatly improved.

The proof of the property claimed for c_0 is as follows. Let T be any symmetric estimate. Its bias function b_T is an analytic function of α and has $b_T(0) = 0$. Suppose, contrary to our claim, the maximum of $F_T(\alpha)$, say c_1, is less than c_0. Then, by virtue of (2.2),

$$I(\alpha)b_T^2(\alpha) + [1 + b_T'(\alpha)]^2 \leqq c_1 \tag{2.5}$$

for all α. Thus $b_T'(0) < b_0'(0)$ so that $b_T < b_0 < 0$ in a neighborhood to the right of $\alpha = 0$. Indeed, we can never have $b_T(\alpha) = b_0(\alpha)$ for any $\alpha > 0$. For if so, by continuity there would be a first $\alpha' > 0$ with $b_T(\alpha') = b_0(\alpha')$, and by Rolle's theorem an α'', where $0 < \alpha'' < \alpha'$, with $b_T'(\alpha'') = b_0'(\alpha'')$ and $b_T(\alpha'') < b_0(\alpha'') < 0$, in violation of (2.5). But since $c_1 < c_0$, the solution b_1 of (2.4) with $c = c_1$ explodes at $A(c_1)$, the still smaller b_T would have exploded still earlier. Thus no estimate T can have maximum risk less than c_0.

It should be emphasized that we are not claiming that c_0 is the *greatest* lower bound for the maximum risk, and in particular we do not assert that in the logistic problem there exists an estimate whose maximum risk is as low as 0.84. In the first place, we do not know whether there is any estimate T_0 whose bias function exactly equals b_0. In the second place, the inequality (2.1) is sharp, as Fréchet points out, only if the estimate T is a linear function of $D(R, \alpha)$; and in our problem these linear functions all have rather high maximum risk.

It would be possible in fact to obtain a higher lower bound than c_0 by using the differential inequality method with the Bhattacharyya bounds [2]. Even the simplest of these, however, would involve the numerical integration of a second-order differential equation with coefficients rather complicated to compute. As we shall now produce directly an estimate with maximum risk 0.85, it hardly seems worthwhile in this problem to engage in heavy computation to close further the remaining gap.

3. Direct construction of an estimate with low maximum risk

The estimate B reflects in its development over a number of years the typical history of statistical methods. Except in simple problems, it is customary for various procedures to be proposed on intuitive grounds, or as consequences of "principles," such as maximum likelihood or minimum chi-square. Then the performances of these competitive procedures are evaluated on the basis of some criterion of excellence. The final stage is an attempt to discover a method which is optimum with regard to the accepted criterion, or at least nearly so.

We suspect that recent advances in computing technology will tend to curtail this process, and that there will be developed methods for the direct attack on practical statistical problems. By a mixture of theoretical, numerical, and ad hoc devices, statisticians may attempt to construct methods having the character of good performance. We shall in this section illustrate such an attempt and obtain, as a by-product of the lower bound found in section 2, an estimate whose risk function is comparable to that of the estimate B, and like it has a maximum risk near to the lower bound c_0.

Our method rests on the fact that we have in b_0 a function which is an idealized bias function. It is reasonable to hope that an estimate whose bias function is b_0, or nearly so, will have a maximum risk nearly c_0. The task of constructing an estimate with a specified bias function, while not trivial, seems much easier than that of constructing an estimate with specified maximum risk.

Several ways of producing an estimate with bias approximately b_0 could be suggested. For example, we could consider a series expansion of odd powers of $(r - 3n/2)$, and determine the coefficients so that the bias is b_0 at several values of α. However, in this problem at least, a more simple-minded approach works well enough. We want an estimate, say H, such that

$$E_\alpha H(R) - \alpha = b_0(\alpha). \tag{3.1}$$

If H were linear, this would mean

$$H[E_\alpha(R)] = \alpha + b_0(\alpha), \tag{3.2}$$

where $E_\alpha(R) = n \sum P_i$. We pretend that H is linear, to obtain the estimate H_1 defined by

$$H_1[n \sum P_i] = \alpha + b_0(\alpha) \tag{3.3}$$

when α takes on such values that $n \sum P_i$ is an integer. We plot $\alpha + b_0(\alpha)$ against $n \sum P_i$ for specific values of $\alpha = \text{logit } P_2$. By linear interpolation we assign as $H_1(r)$ the value of $\alpha + b_0(\alpha)$ corresponding to integral values of $n \sum P_i = r$. The method breaks down for $R = 3n$, as $\sum P_i < 3$, so we employ the ad hoc value $H_1(3n) = 3.37$ used above, which is determined to control the risk for large values of α. The estimate H_1 is shown in table I, and its bias function b_1 in table II. We see that the bias agrees fairly well with b_0 unless α is quite large.

We can now improve the agreement by repeating the above procedure. We plot the error $b_1 - b_0$ against $n \sum P_i$ and correct H_1 to obtain the estimate H_2 (table I). Its bias b_2 (table II) agrees with b_0 to within 0.008 over the range $0.1 < P_2 < 0.9$. Further correction seems unnecessary.

The risk function F_2 of H_2 is shown in table II. We see that its maximum is about 0.90 near $P_2 = 0.9$. The large value of the risk F_2 for extreme values of

P_2 are however almost entirely due to the values of H_2 corresponding to large R. We note that $H_2(29) = 2.741$, well above the value of $\alpha = 2.19722$ corresponding to $P_2 = 0.9$. An ad hoc adjustment of $H_2(29)$ to the value 2.592 produces the risk function F_2', with maximum of 0.870.

Our position is now this: starting with the idealized bias function corresponding to constant risk $c = 0.84$, we have produced an estimate whose risk is indeed near 0.84 for small values of α, but which rises to a maximum of 0.87 for large α. The difficulty is of course that the lower bound on which the idealized bias function is based, is not actually attained. In achieving the risk 0.84 for small α, the estimate is too concentrated about 0, and as a result has larger risk for large α.

It is natural to think of trying a somewhat larger value of c, in the hope that a lower maximum risk will result. We accordingly put $c = 0.85$, and go through a computation analogous to that outlined above. The result is the estimate H given in the last column of table I, which has the risk function F shown in the last column of table II. The maximum risk has indeed been reduced from 0.87 to 0.85. We believe that, to two decimals, this is the minimax risk, though the argument of section 2 shows only that the minimax risk is not less than 0.836.

We conclude by indicating an alternative technique that would produce a similar result, and which might be advantageous in some problems. The equality corresponding to (2.1) is

$$\sigma_T^2 = \frac{[1 + b_T'(\alpha)]^2}{I(\alpha)\rho^2(T, D)}. \tag{3.4}$$

If we have found an estimate T which is nearly optimum, we may assume that its correlation with D is nearly the same as that of the optimum estimate. We can compute the factor $1/\rho^2(T, D)$ and use (3.4) instead of (2.1) in the method of differential inequalities. The whole procedure may then be iterated if desired.

4. Comparison of some estimators

In the paper by one of us previously mentioned [1], a comparison was made between several estimators on the basis of their mean square errors. These estimators were: the maximum likelihood, the minimum Pearson chi-square, the minimum logit chi-square and the Rao-Blackwellized minimum logit chi-square. It seems to the point to review this comparison against the background of the estimator H developed here. In addition we will include the Spearman-Karber estimator which has recently been studied in detail by Brown [3].

The maximum likelihood estimate yields ∞ as estimate for α with the sample $r = 30$. In the comparisons made previously, the samples yielding infinity by maximum likelihood were omitted in computing the statistics of all estimators. In the present development, for the estimators B and H_2, this sample was assigned the estimate $\alpha = 3.37$. Therefore, in the comparisons made here, in computing the mean square errors, this same value was given to the maximum likelihood and minimum χ^2 estimates for the sample in question, and the total sampling population was included. The probability of the sample $r = 30$ increases as P_2 and α are increased. It seems excessively artificial to consider dosage arrangements in which the probability of this sample—for which the maximum likelihood estimate is actually infinity and therefore must be omitted or the estimate given an arbitrary value—is appreciable and which, in any case, is far outside the range of an acceptable practical experiment. We therefore limit the comparisons to dosage arrangements in which the probability of $r = 30$ is less than 5 per cent. At $P_2 = 0.95$, this probability is 15 per cent, so $P_2 = 0.9$, $\alpha = 2.197$, at which the probability is 2 per cent, was the upper limit of the experiments compared.

TABLE III

COMPARISON OF ESTIMATORS

P_2	0.5		0.6		0.7		0.8		0.85		0.9	
Estimators	m.s.e.	Risk F	m.s.e.	Risk F	m.s.e.	Risk F	m.s.e.	Risk F	m.s.e.	Risk F	m.s.e.	Risk F
H	.1268	.8497	.1307	.8509	.1438	.8502	.1769	.8492	.2119	.8465	.2853	.8460
B	.1310	.8778	.1354	.8818	.1512	.8935	.1837	.8820	.2101	.8393	.2572	.7816
Maximum likelihood	.1578	1.0575	.1639	1.0673	.1867	1.1034	.2507	1.2038	.3189	1.2740	.4010	1.1904
Minimum Pearson χ^2	.1385	.9281	.1436	.9346	.1622	.9589	.2124	1.0198	.2619	1.0461	—	—
Minimum logit χ^2	.1366	.9152	.1422	.9260	.1593	.9413	.1975	.9482	.2281	.9109	.2787	.8271
Spearman-Karber	.0481	.3223	.0783	.5096	.1912	1.1300	.4955	2.3787	.8341	3.3319	1.5194	4.5108

The comparisons are shown in table III, which gives the values of risk function F, and also the mean square errors. The maximum value for F, being lowest for the estimator H(0.851 at P_2 about 0.6) is seen to be largest for the Spearman-Karber estimate (4.5 at P_2 about 0.9). [The value of F for the Spearman-Karber becomes even larger for $P_2 > 0.9$, reaching a maximum of 5.6 at P_2 about 0.95. However it is to be noted that the smallest value for F also is achieved by the Spearman-Karber (0.32 at $P_2 = 0.5$)]. The maximum likelihood estimator has a higher value of F, and also of the mean square error, than any of the other estimators compared except the Spearman-Karber, over the entire range of comparison $0.1 \leqq P_2 \leqq 0.9$.

5. Summary

If some estimators T of a parameter θ are to be compared as to relative worth, an indefinite number of criteria are conceivable. A criterion classic in the history of the theory of errors is the mean square error, m.s.e. $= E(T - \theta)^2$. If the m.s.e. of T_1 is smaller than the m.s.e. of T_2 for all values of θ, then T_1 is, by this criterion,

unequivocally better than T_2. If, however, for some values of θ, T_1 has the smaller m.s.e., while for other values of θ, T_2 has the smaller m.s.e., a question arises as to how they may be compared. Conceivably, if one knew the a priori probability distribution of θ, the mean of the m.s.e. of T_1 and of T_2 might serve as the criterion. However, in any real situation, the a priori distribution of θ is seldom definable, not to speak of being precisely determinable. Another criterion which suggests itself, related to that used extensively in the work of Wald, is to minimize the largest value of some function of the m.s.e. (minimax principle).

We consider estimating the logistic parameter α, with β known, for the case with three equally spaced x, ten at each x. A risk function $F_T(\alpha)$ is defined as the ratio of the m.s.e. of T to $1/I(\alpha)$, where $I(\alpha)$ is Fisher's amount of extractable information, and the criterion considered for the choice of T is the maximum value of $F_T(\alpha)$, for all possible α.

(1) A lower bound for the maximum of $F_T(\alpha)$ is evaluated. (2) An estimator H is devised for which the maximum value of $F_T(\alpha)$ almost achieves this bound. (3) The estimator H and several other estimators of interest are compared with respect to their F and their m.s.e.

REFERENCES

[1] J. Berkson, "Maximum likelihood and minimum χ^2 estimates of the logistic function," *J. Amer. Statist. Assoc.*, Vol. 50 (1955), pp. 130–162.

[2] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistics," *Sankhyā*, Vol. 8 (1946), pp. 1–14 and 201–218.

[3] B. W. Brown, Jr., *Some Properties of the Spearman Estimator in Bioassay*, Technical Report No. 6, University of Minnesota, School of Business Administration.

[4] H. Cramér, *Mathematical Methods of Statistics*, Princeton, Princeton University Press, 1946.

[5] F. Y. Edgeworth, "On the probable errors of frequency constants," *J. Roy. Statist. Soc.*, Vol. 71 (1908), pp. 662–668.

[6] M. Fréchet, "Sur l'extension de certaines evaluations statistiques au cas de petits enchantillons," *Rev. Inst. Internat. Statist.*, Vol. 11 (1943), pp. 182–205.

[7] J. L. Hodges, Jr., and E. L. Lehmann, "Some applications of the Cramér-Rao inequality," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 13–22.

[8] E. Pearson and L. N. G. Filon, "Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation," *Philos. Trans. Roy. Soc. London, Ser. A*, Vol. 191 (1898), pp. 229–311.

[9] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37 (1945), pp. 81–91.