PSYCHOMETRIKA—VOL. 33, NO. 1
MARCH, 1968

MAXIMUM LIKELIHOOD ESTIMATION OF PARAMETERS OF SIGNAL DETECTION THEORY—A DIRECT SOLUTION*

DONALD D. DORFMAN AND EDWARD ALF, JR.

SAN DIEGO STATE COLLEGE AND

U. S. NAVAL PERSONNEL RESEARCH ACTIVITY

Ogilvie and Creelman have recently attempted to develop maximum likelihood estimates of the parameters of signal-detection theory from the data of yes-no ROC curves. Their method involved the assumption of a logistic distribution rather than the normal distribution in order to make the mathematics more tractable. The present paper presents a method of obtaining maximum likelihood estimates of these parameters using the assumption of underlying normal distributions.

Ogilvie and Creelman [1966] have recently attempted to develop maximum likelihood estimates of the parameters of signal-detection theory from the data of yes-no ROC curves. Their method involved the assumption of a logistic distribution rather than the normal distribution in order to make the mathematics more tractable. Since signal-detection theory [Swets, Tanner & Birdsall, 1961; Green & Swets, 1966] assumes underlying normal distributions, their method involved changing the model into what they considered to be a more convenient mathematical model. They estimated **d′** by means of an empirical relation which they obtained between **d′** and an analogous parameter in the logistic model. This relation was found through numerical experiments on a high-speed computer. Unfortunately, a stable empirical relation could not be found between the sigma ratio of signal detection theory and the analogous parameter of the logistic model. Consequently, the problem of obtaining maximum likelihood estimates of the parameters of signal detection theory from yes-no ROC curves still remains. The present paper presents a method of obtaining maximum likelihood estimates of these parameters using the assumption of underlying normal distributions.

Solution

For the present case, signal-detection theory is employed in its general form [e.g., see Green & Swets, 1966]. This general model has been applied to such problems as recognition memory as well as signal detection [e.g.,

*This research was supported in part by grants from the National Institutes of Health, MH-10449-02, and from the National Science Foundation, NSF GS-1466.

Egan, 1958; Murdock, 1965; Parks, 1966]. In its general form, the model may be summarized as follows. The observable events of the experiment consist of two stimulus classes, S_1 and S_2, and two responses, R_1 and R_2, and a set of instructions I_k ($k = 1$ to n').*

Axiom 1. On each trial, the presentation of an S_i leads to an event x located on a unidimensional continuum.†

Axiom 2. Over an infinite set of trials, the presentation of S_i is associated with a normal distribution of such x's with mean M_i and variance σ_i^2.

Axiom 3. For each instruction, I_k, there exists a cutoff X_k, such that if $x > X_k$, the subject responds R_2, and if $x < X_k$, the subject responds R_1.

It follows from these assumptions that

$$P(R_2 \mid S_1) = 1 - F(Z_k) = F(-Z_k), \tag{1}$$

where $Z_k = (X_k - M_1)/\sigma_1$, and F is the cumulative normal function.

$$P(R_2 \mid S_2) = 1 - F(bZ_k - a) = F(a - bZ_k), \tag{2}$$

where $b = \sigma_1/\sigma_2$ and $a = (M_2 - M_1)/\sigma_2$.

The method of maximum likelihood requires that we maximize the likelihood function with respect to the parameters, a, b, and all the Z_k. In this case, this may be accomplished by differentiating the log of the likelihood function with respect to a, b and all the Z_k, setting these first partial derivatives equal to zero, and then solving the resulting set of equations. With yes-no data, the log of the likelihood function is

$$\text{Log } L = \sum_{k=1}^{k=n'} \{r_{1k} \log P_{1k} + (n_{1k} - r_{1k}) \log Q_{1k}\} + \sum_{k=1}^{k=n'} \{r_{2k} \log P_{2k} + (n_{2k} - r_{2k}) \log Q_{2k}\}, \tag{3}$$

where n_{ik} is the number of presentations of S_i under instruction I_k, n' is the number of instructions, r_{ik} is the number of R_2's to S_i under instruction I_k, $P_{1k} = F(-Z_k)$, $P_{2k} = F(a - bZ_k)$, and $P_{ik} = 1 - Q_{ik}$. Differentiating (3) with respect to a, we obtain

$$\frac{\partial \log L}{\partial a} = \sum_{k=1}^{k=n'} n_{2k} \frac{[\hat{P}_{2k} - F(a - bZ_k)]f(a - bZ_k)}{F(a - bZ_k)[1 - F(a - bZ_k)]}, \tag{4}$$

where f is the normal density function, and $\hat{p}_{ik} = r_{ik}/n_{ik}$. Differentiating (3) with respect to b,

$$\frac{\partial \log L}{\partial b} = \sum_{k=1}^{k=n'} n_{2k} \frac{[\hat{P}_{2k} - F(a - bZ_k)]f(a - bZ_k)(-Z_k)}{F(a - bZ_k)[1 - F(a - bZ_k)]}, \tag{5}$$

*An instruction usually consists of an *a priori* probability and/or a payoff matrix.

†In the general form, this continuum may or may not be a monotone transformation of the likelihood ratio.

Differentiating (3) with respect to Z_k,

$$\frac{\partial \log L}{\partial Z_k} = n_{1k} \frac{[\hat{P}_{1k} - F(-Z_k)]f(-Z_k)(-1)}{F(-Z_k)[1 - F(-Z_k)]} + n_{2k} \frac{[\hat{P}_{2k} - F(a - bZ_k)]f(a - bZ_k)(-b)}{F(a - bZ_k)[1 - F(a - bZ_k)]}. \tag{6}$$

Setting these partial derivatives equal to zero, we obtain a set of non-linear equations. The solution to such a set of equations may be obtained by an adaptation of the Newton–Raphson method [Kendall & Stuart, 1961; Rao, 1952], sometimes called the method of scoring. This approximation method was also used by Ogilvie and Creelman [1966]. Specifically, given a vector of consistent, but inefficient estimates, an improved vector of estimates is obtained from

$$\mathbf{S}_1 = \mathbf{S}_0 + \mathbf{A}^{-1}\mathbf{r}, \tag{7}$$

where $\mathbf{S}_1$ is the improved vector of estimates, $\mathbf{r}$ is the vector of first partial derivatives with the initial estimates substituted for the unknowns, and $\mathbf{A}^{-1}$ is the inverse of the matrix $\{-E(\partial^2 \log L)/(\partial\theta_t\ \partial\theta_u)\}$. Note that the elements of A are minus the expected values of the second partial derivatives, and that $\mathbf{A}^{-1}$ is the variance-covariance matrix of the estimates after the iterations are completed [Rao, 1952].

These second partial derivatives are as follows,

$$E\,\frac{\partial^2 \log L}{\partial a^2} = -\sum_{k=1}^{n'} n_{2k}w_{2k}\,,$$

$$E\,\frac{\partial^2 \log L}{\partial b^2} = -\sum_{k=1}^{n'} n_{2k}w_{2k}Z_k^2\,,$$

$$E\,\frac{\partial^2 \log L}{\partial Z_k^2} = -n_{1k}w_{1k} - n_{2k}w_{2k}b^2,$$

$$E\,\frac{\partial^2 \log L}{\partial a\,\partial Z_k} = -n_{2k}w_{2k}(-b),$$

$$E\,\frac{\partial^2 \log L}{\partial b\,\partial Z_k} = -n_{2k}w_{2k}bZ_k\,,$$

$$E\,\frac{\partial^2 \log L}{\underset{i\neq j}{\partial Z_{k_i}\,\partial Z_{k_j}}} = 0,$$

$$E\,\frac{\partial^2 \log L}{\partial a\,\partial b} = -\sum_{k=1}^{n'} n_{2k}w_{2k}(-Z_k),$$

where

$$w_{1k} = \frac{f^2(-Z_k)}{F(-Z_k)[1 - F(-Z_k)]}$$

and,

$$w_{2k} = \frac{f^2(a - bZ_k)}{F(a - bZ_k)[1 - F(a - bZ_k)]}.$$

It is interesting that w_{1k} and w_{2k} are identical to probit weights, and that this analysis is essentially a two-dimensional probit-analysis. See Finney [1962] for the one-dimensional probit analysis.

Consistent, but inefficient initial estimates can be obtained by fitting the function, $Y_k = a - bZ_k$, by eye to the observed proportions on normal-normal paper. The Y_k is the standardized normal deviate for $P(R_2 \mid S_2)$ and $-Z_k$ is the standardized normal deviate for $P(R_2 \mid S_1)$. One obtains a and b directly from this fit. To obtain the Z_k's, draw perpendicular distances from the points to the line, and the x-coordinate of the point of intersection is $-Z_k$. As n approaches ∞, these initial estimates will approach the population parameters. Under the assumption that all population proportions, P, are $0 < P < 1$, if one observed proportion is zero or one, use the appropriate vertical or horizontal distance to the provisional straight line. If both proportions of a point are zero or one (rare with large n), the initial estimate is arbitrary. This latter rule of choosing an arbitrary initial estimate of Z_k when both proportions are zero and/or one, also produces a consistent estimate (but inefficient) since $\hat{p} \to P$ as $n \to \infty$ and $0 < P < 1$.

Finally, it should be noted that in signal detection theory, the ROC curves may be summarized by $D(\Delta m, s)$, where $\Delta m = a/b$ and s is $\sigma_2/\sigma_1 = 1/b$, as presented in Green and Swets [1966]. When $\sigma_2/\sigma_1 = 1$, $\Delta m = \mathbf{d}' = a$.

Example

For illustrative purposes, maximum likelihood estimates of the parameters were obtained on the data of the four observers presented in Swets, Tanner, & Birdsall [1961]. The data of these observers are presented in Table 1. The ROC curves were generated by manipulating the payoff matrix. The computations were performed on a high-speed computer with a program written in Fortran II-D. This program provided estimates of the parameters a, b and all Z_k, their variance-covariance matrix, computed the χ^2 and log $\mathbf{L}$ on each iteration. Six iterations were performed on the data of the four observers.* Figure 1 presents the observed points and the theoretical function $Y_k = a - bZ_k$, based on the maximum likelihood estimates using standardized normal coordinates. There were four points in which one of the proportions was zero, two points for Observer 2, one for Observer 3, and one for Observer 4. These data are not presented in the figure, although the data were, of course, used in the computation of the estimates.
Each point was based upon 400 observations, 200 of S_1 and 200 of S_2.

Convergence was rapid for all four observers. Table 2 presents the initial

*This program is available upon request.

TABLE 1
The Relevant Observed Proportions of the Four Observers

	Observer 1		Observer 2		Observer 3		Observer 4	
Day	$P(R_2 \mid S_1)$	$P(R_2 \mid S_2)$	$P(R_2 \mid S_1)$	$P(R_2 \mid S_2)$	$P(R_2 \mid S_1)$	$P(R_2 \mid S_2)$	$P(R_2 \mid S_1)$	$P(R_2 \mid S_2)$
1	.32	.77	.23	.84	.18	.90	.18	.77
2	.50	.88	.25	.95	.12	.88	.35	.83
3	.65	.91	.41	.93	.33	.85	.49	.87
4	.78	.96	.66	.95	.59	.94	.74	.93
5	.05	.63	.11	.80	.01	.65	.17	.71
6	.06	.79	.03	.75	.02	.64	.12	.64
7	.20	.85	.19	.90	.15	.78	.14	.76
8	.02	.82	.02	.73	—	—	.06	.64
9	.02	.64	.00	.48	.00	.45	.00	.43
10	—	—	.04	.71	.08	.73	.05	.56
11	.19	.89	.18	.87	.14	.84	.34	.84
12	.29	.84	.25	.89	.24	.87	.34	.86
13	.03	.64	.00	.59	.02	.56	.04	.43

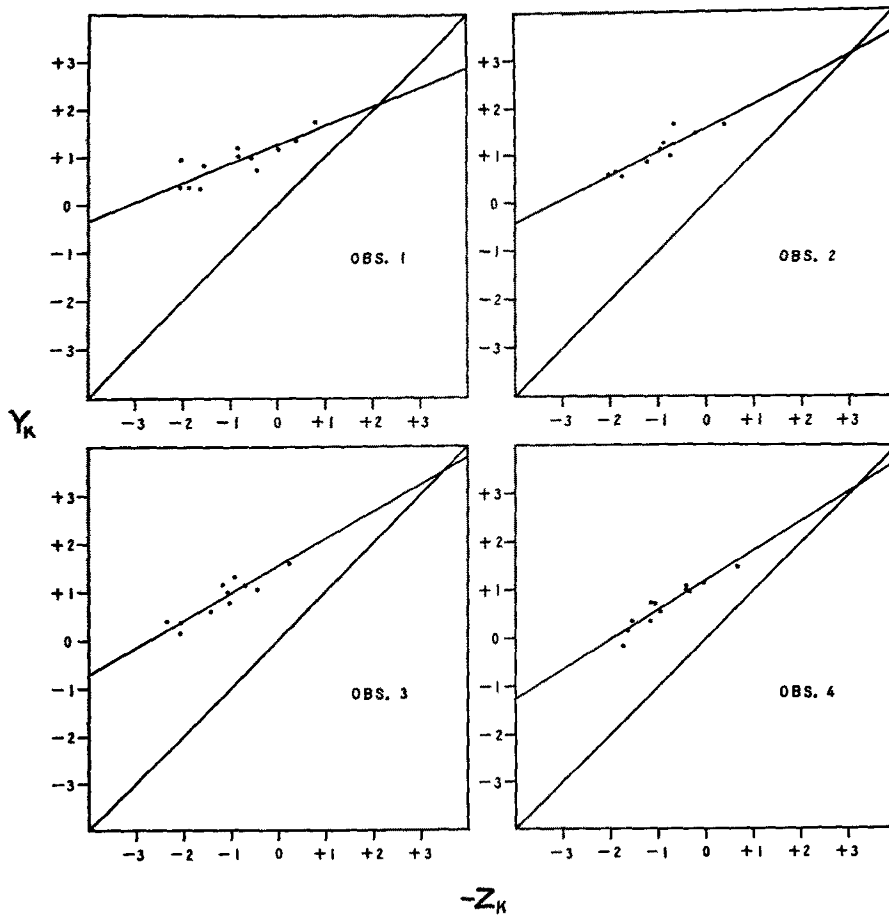

FIGURE 1
ROC Curves for the Four Observers

estimates of a and b, the χ^2 based upon these initial estimates, the final estimates of a and b, and the Δm, s, $\Delta m/\Delta s$, and the χ^2 based upon these final estimates. The χ^2 formula is as follows:

$$\chi^2 = \sum_{k=1}^{n'} n_{2k} \frac{[\hat{P}_{2k} - F(a - bZ_k)]^2}{F(a - bZ_k)[1 - F(a - bZ_k)]} + \sum_{k=1}^{n'} n_{1k} \frac{[\hat{P}_{1k} - F(-Z_k)]^2}{F(-Z_k)[1 - F(-Z_k)]}.$$

There are $n' - 2$ degrees of freedom for the χ^2, since there are $2n'$ independent normally distributed variables, $N(0, 1)$, and $n' + 2$ parameters which are estimated from the data. It should be noted that the χ^2 test of goodness-of-fit requires estimates of the parameters which are asymptotically normal and asymptotically efficient. The method of maximum likelihood gives such

TABLE 2
Results on the Four Observers

	Observer			
	1	2	3	4
Initial estimates:				
a	+1.32	+1.60	+1.41	+1.22
b	+0.41	+0.52	+0.48	+0.60
X^2 (Initial)	+65.14	+19.51	+32.88	+28.04
Final estimates:				
a	+1.27	+1.57	+1.51	+1.21
b	+0.41	+0.49	+0.55	+0.62
SDT—summary:				
Δm	+3.07	+3.20	+2.75	+1.94
s	+2.44	+2.04	+1.82	+1.61
$\Delta m/\Delta s$	+2.13	+3.08	+3.35	+3.18
X^2 (Final)	+43.82*	+15.45	+21.32*	+15.41

*$p < .05$

estimates in this case. The χ^2 should provide a sensitive index of goodness of fit with which to compare various models of the data of ROC-curves.

Sample Variances and Confidence Intervals

Table 3 presents the large-sample variances of the estimates of a and b obtained from the variance-covariance matrix. Since the parameter estimates are approximately normally distributed with large n, we may also obtain confidence intervals for the parameter estimates. Table 4 presents the 95% confidence intervals for a and b under the assumption of normality.

TABLE 3
Sample Variances for a and b

	Observer			
	1	2	3	4
Variances:				
S_a^2	.0027	.0048	.0061	.0033
S_b^2	.0016	.0028	.0034	.0026

TABLE 4
95% Confidence Intervals for a and b

Observer	a	b
1	$1.17 \leq a \leq 1.37$	$.33 \leq b \leq .49$
2	$1.43 \leq a \leq 1.71$	$.39 \leq b \leq .59$
3	$1.36 \leq a \leq 1.66$	$.44 \leq b \leq .66$
4	$1.10 \leq a \leq 1.32$	$.51 \leq b \leq .73$

Uniqueness of the Solution

There does not appear to exist a set of statistics which are jointly sufficient for the parameters. In the absence of joint sufficiency, one cannot usually determine if the solution is unique with approximation methods. However, for large n, there is a unique consistent solution to the likelihood equations [Kendall et al., 1961], where "consistent" is defined in the statistical sense. Given a set of consistent initial estimates, the procedure used in this paper yields final solutions which are also consistent [Kendall et. al., 1961]. Consequently, with large n, the approximations will tend to the ML estimates, and will therefore be asymptotically normal and efficient since the ML estimates have these properties.

REFERENCES

Egan, J. P. Recognition memory and the operating characteristic. Indiana Univer.: Hearing and Communication Laboratory, Technical Note AFCR-TN-58-51, 1958.

Finney, D. J. *Probit analysis* (2nd ed.). Cambridge: Cambridge University Press, 1962.

Green, D. M. & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Kendall, M. G. & Stuart, A. *The advanced theory of statistics,* Vol. 2. London: Griffin, 1961.

Murdock, B. B., Jr. Signal-detection theory and short-term memory. *Journal of Experimental Psychology,* 1965, **70**, 443–447.

Ogilvie, J. C. & Creelman, C. D. Maximum likelihood estimations of ROC curve parameters. Paper read at the Eastern Psychological Association, New York, April, 1966.

Parks, T. E. Signal-detectability theory of recognition-memory performance. *Psychological Review,* 1966, **73**, 44–58.

Rao, C. R. *Advanced methods in biometric research.* New York: Wiley, 1952.

Swets, J. A., Tanner, W. P., Jr. & Birdsall, T. G. The evidence for a decision-making theory of visual detection. Univ. of Mich.: Electronic Defense Group Tech. Rep. No. 40, 1955.

Swets, J. A., Tanner, W. P., Jr. & Birdsall, T. G. Decision processes in perception. *Psychological Review,* 1961, **68**, 301–340.

Manuscript received 3/1/67
Revised Manuscript received 8/31/67