

Observations on a maximum likelihood method of sequential threshold estimation and a simplified approximation

PHILLIP L. EMERSON
Cleveland State University, Cleveland, Ohio

In sequential methods of measuring thresholds, the stimuli are chosen from trial to trial with the goal that the information from each trial be highly relevant to the objective. The presentation of many stimuli far above or far below threshold is a waste of time, because the responses to such stimuli have little bearing on the question of the location of the threshold. Efficiency of threshold estimation (Taylor, 1970) can be formulated in terms of the tradeoff between the standard error of the threshold estimate and the average number of trials necessary to achieve a given small standard error. A sequential method, such as the well-known PEST (Percentile Estimation by a Sequential Threshold method) of Taylor and Creelman (1967), can be more efficient by a large factor than one of the classical methods, such as constant stimuli. However, even among the various sequential methods that are in use, there are fairly large differences in efficiency. For example, the number of trials needed to achieve a given standard error using a maximum likelihood method (Pentland, 1980; Shelton, Picardi, & Green, 1982) is, in some cases, as small as half those needed with PEST and some variations (Findlay, 1978) of the PEST method.

In the maximum likelihood method of sequential threshold estimation, a new estimate of the threshold is obtained after every trial. That estimate is then used on the following trial to determine the stimulus to be presented. In the simplest cases, the stimulus value is always taken equal to the threshold estimate itself. Thus, a powerful statistical estimation method is used to concentrate stimuli densely in the region of the threshold, to maximize the relevance of the data. That seems to explain the superior efficiency.

Although the maximum likelihood method was conceived and used for theoretical purposes early in the development of sequential threshold methods (Smith, 1961; Wetherill, 1963), it has not been very popular as a practical scheme for measuring thresholds. This is somewhat curious, since it seems to be more efficient than more popular ones. Some hints toward an explanation may be inferred from com-

ments by a few psychophysical researchers. Hall (1981) mentioned a problem of lapses of attention on the part of the subject. If the subject gives a negative response to an easily identifiable stimulus, the current maximum likelihood estimate of the threshold is sent to a high level from which it then descends gradually with a number of easily identifiable stimuli. Taylor, Forbes, and Creelman (1983) suggested that the maximum likelihood method, among some others, suffers in not presenting enough easy exemplars of the stimulus to maintain a high level of motivation and, perhaps, refresh the subject's memory as to what he is supposed to be looking or listening for. They also mentioned other possible artifacts, such as lapses of attention. Shelton et al. (1982) mentioned another problem that is somewhat the same, although they did not seem to regard it as very serious. The maximum likelihood method converges in a few trials so rapidly toward the vicinity of the threshold that a naive subject gets little familiarizing experience with easy exemplars of the stimulus. On the other hand, Pentland (1980) mentioned using the method in certain visual experiments, and waxed quite enthusiastic about the rapid convergence, and the like. Very few researchers have reported using the maximum likelihood method in actual experiments, so it would seem premature to pass judgment. The hinted drawbacks mentioned above may reflect genuine disadvantages, or they may merely imply that customary experimental techniques need to be revised in order that the full advantages of the superior efficiency be realized.

A discouragement to some of those who might like to try the maximum likelihood method of sequential threshold estimation is that it requires quite a lot of computation between trials. Essentially, for each trialwise decision on the value of the stimulus to be presented next, it requires a thorough analysis of all data accumulated so far. A close approximation to the maximum likelihood method is reported here—one that can be implemented with a large reduction of the computations between trials. Indeed, the trial-to-trial operation is reduced to (1) the observance of the response on the preceding trial, and (2) the decrease or increase of the stimulus by a step size given in Table 1. The method reduces to the simplest of up-down methods (Dixon & Mood, 1948), except that the step size is not constant—it must be looked up in Table 1. This scheme is in the general class of Robbins-Monro processes (Robbins & Monro, 1951).

The Method

This method is for determining yes-no thresholds, that is, 50th percentile levels. Roughly, the method of using Table 1 is to (1) choose a lower bound, X_0 , and

This paper was written while the author was on professional leave from Cleveland State University during part of 1984. The ideas stem from research begun while the author was supported by a research initiation award from Cleveland State University during part of 1969. The author's mailing address is: Department of Psychology, Cleveland State University, Cleveland, OH 44115.

Table 1
Step Sizes for the Simplified Approximation to the Maximum Likelihood Method

Second Digit	First Digit of Trial Number				
	0	1	2	3	4
0	25.0000	.3216	.1298	.0798	.0575
1	12.5000	.2810	.1221	.0768	.0559
2	6.2500	.2494	.1152	.0741	.0544
3	3.1261	.2243	.1092	.0715	.0530
4	1.5926	.2040	.1039	.0691	.0516
5	.9362	.1867	.0990	.0668	.0504
6	.6746	.1719	.0945	.0647	.0491
7	.5348	.1593	.0903	.0628	.0480
8	.4413	.1481	.0865	.0609	.0469
9	.3739	.1383	.0830	.0591	.0458

an upper bound, X1, on the range of stimulus values to be used; (2) determine the range, $R = X1 - X0$; (3) take the first stimulus to be at midrange, $(X0 + X1)/2$; and (4) proceed with stimulus step sizes as given in Table 1, rescaled by the factor of $R/50$. The sign of the increment to the stimulus value should be taken as negative if the preceding response was yes, or positive if it was no. Some refinements of this procedure are discussed later in the light of some empirical observations. With the true maximum likelihood method, as with this simplified approximation to it, the convergence toward the vicinity of the threshold is very rapid in early trials, even though one may be quite conservative in specifying a broad range from $X0$ to $X1$. After homing in on the vicinity of the threshold, the pattern then gradually shifts to one that resembles a standard Robbins-Monro process (Wetherill, 1963) in which the step sizes are proportional to the reciprocal of the trial number.

The maximum likelihood method that is simulated by the use of Table 1 is a generalization of that outlined by Pentland (1980). The assumed psychometric function is the logistic $1/[1 + \exp\{k(m - X)\}]$, where X is the stimulus value, m is the threshold to be estimated, and k is a slope parameter. This expression for the psychometric function is assumed to give the probability of a yes response in a yes-no psychophysical paradigm. In Pentland's (1980) method, a value of k is assumed a priori, so that only the parameter m remains to be estimated. This is a simplification of the more general method (Hall, 1981) in which both parameters must be estimated simultaneously. It is quite a useful simplification, but what is the basis for the choice of the a priori value of k , and what are the consequences of mismatches between the assumed and true values of k ?

Empirical Observations

Casual observations of the maximum likelihood method in action indicated quite regular patterns in the course of changes of step sizes in a sequence of trials. Therefrom arose the idea of simulating the method by some simpler process, to reduce the con-

siderable computational effort. For $k = 1$ and $R = 50$, 500 Monte Carlo runs were performed, in which the starting value (the center of the range) was randomly placed from 5 to 10 logit units away from the true threshold. The observed step sizes were averaged over the 500 runs for each of the first 49 trials, and the entries in Table 1 are just those average step sizes from the true maximum likelihood method. Table 1 provides a very simple and useful sequential method of measuring thresholds. In order to compare its performance with that of the true maximum likelihood method, further Monte Carlo experiments were conducted.

The first experiments were with the true maximum likelihood method, and the results are shown in Figure 1. The lowest curve is for the ideal case in which the a priori assumed value of k is equal to the true value. This curve matches that of Pentland (1980) quite closely, so it may be inferred that the present attempt to match his conditions was successful. This curve was generated from 500 Monte Carlo simulations, and it shows the expected steady decrease of the standard error of estimate of the threshold with increasing trials.

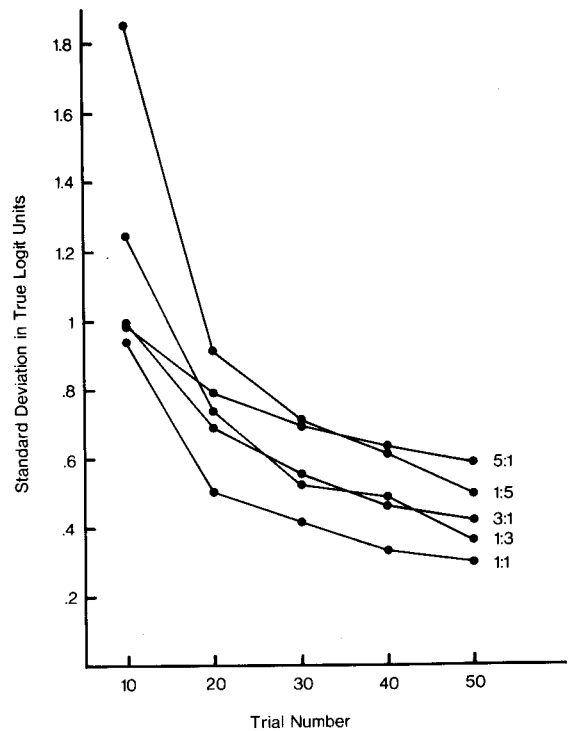


Figure 1. Standard deviation of threshold estimates as a function of the number of trials, with the maximum likelihood method of sequential threshold estimation. Curves are identified on the right by the mismatch ratio between the assumed and true values of the slope parameter of the logistic function. A perfect match is designated as 1:1, a 5-to-1 overestimate as 5:1, etc. For the 1:1 case, data were based on 500 Monte Carlo runs. For the other cases, each curve is based on 100 Monte Carlo runs.

The other curves in Figure 1 were each obtained from 100 Monte Carlo runs for various cases of mismatch between the assumed and true k values. Other researchers had claimed that the effects of mismatch were not very serious for mismatches up to about a 4-to-1 ratio. We now have fairly detailed quantitative evidence in support of this assertion, and Figure 1 also provides some additional insight about the maximum likelihood method. There is one fact not presented in Figure 1 that should be noted carefully and kept in mind. For the condition of a 5:1 overestimate of k , there was small but notable bias in the threshold estimates, resulting from the off-center initial starting values of m . In the other mismatch cases, it seemed quite clear, from the mean threshold estimates, that there was no bias. Therefore, the 5:1 condition, alone among these cases, should be considered inadmissible.

Generally, the effect of mismatch is to elevate the standard error of the threshold estimate. The effects of mismatch in one direction are approximately the same as for the other, except when bias occurs with severe overestimation of k . However, there is a hint that the two corresponding curves cross each other somewhere between 20 and 50 trials. Thus, underestimates seem to achieve better results on later trials, and overestimates on early trials. In view of the bias that occurs for extreme overestimates, one may conclude that underestimation is more conservative than overestimation of the slope parameter, k . Further details of Figure 1 will be relevant in comparisons now to be made of the true maximum likelihood method and the simplified approximation to it based on Table 1.

Figure 2 shows the results of the same Monte Carlo experiments performed with the simplified method associated with Table 1. Of first concern is the comparison of the performances of the two methods in the ideal case in which the assumed and true values of k are equal. It is quite remarkable that the simplified method suffers very little in this comparison. The standard errors differ at most (on Trial 20) by a factor of 20%. They differ by less than one logit unit on Trials 10 and 20, by less than a quarter of a logit unit on Trial 30, and are almost indistinguishable on Trials 40 and 50. To put the comparison in another way, the numbers of trials required to reach the same standard error criterion differ at most by 5 (for a criterion of .5 at Trials 20 and 25), and typically differ only by about 1 or 2.

In considering the effects of mismatch of assumed and true k , it must be reported again that there was bias for the case of a 5:1 overestimate, but not for the other cases, just as with the true maximum likelihood method. From making comparisons of individual curves in Figures 1 and 2, it seems that the simplified method does just about as well as the true maximum likelihood method for cases of underestimation (1:3 and 1:5). The points on the 1:3 curve in

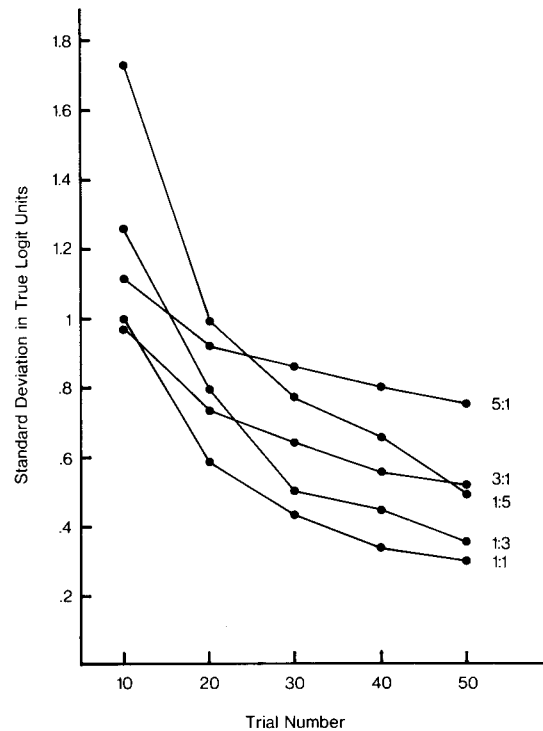


Figure 2. Standard deviation of threshold estimates as a function of the number of trials, with the Robbins-Monro approximation to the maximum likelihood method. Legends and conditions are the same as for Figure 1.

Figure 2 are actually lower than those in Figure 1 on Trials 30, 40, and 50. The differences are slight, though, and probably attributable to sampling error. For overestimates of the slope parameter value, the simplified method does seem to suffer a bit, and especially for the 5:1 overestimate. It may be that there is a tradeoff between variance and bias of the estimate, with these conditions of overestimation of the slope. With the simplified method, as with the true maximum likelihood method, underestimation of k would seem to be the more conservative policy.

Another remarkable outcome of comparing Figures 1 and 2 is the overall similarity of the configurations. The crossover of curves for reciprocal mismatches is even more apparent in Figure 2. Wetherill (1963) observed this crossover phenomenon in a more standard form of the Robbins-Monro process, where it appeared as an analysis of variance interaction between trial number and a slope parameter. The similarities of the patterns of bumps and dips of the curves in Figures 1 and 2 are exceedingly curious, although perhaps of little practical importance. Different seed numbers were used for the random number generator at the beginnings of all 10 of the Monte Carlo runs on which these 10 curves are based. Therefore, the apparent correspondence of the dips and bumps of the corresponding curves does not seem attributable to that sort of artifact.

Empirical Slope Information

It is perhaps true that many experienced psychophysical researchers are able to make good prior estimates of the slopes of psychometric functions. For my own benefit, and for the benefit of others who do not feel confident in their subjective guesses, further results from the Monte Carlo simulations are presented in Figure 3. These results can be used in a process of building up information about the value of the true slope parameter, k . The process applies just as well to the true maximum likelihood method and to the present simplified method of sequential threshold estimation.

Generally, sequential methods that are highly efficient for estimating the threshold or location parameter of a psychometric function, yield little slope information directly (Wetherill, 1963). Thus, it would seem that there is not much point in trying to use the data collected in a single sequence of trials to estimate k when that sequence is designed for the estimation of m . However, if several threshold estimates from short sequential runs are made under the same conditions, all using the same assumed value of k , then the standard deviation of those threshold estimates can provide useful information about the true value of k . Thus, very brief pilot experiments can help to ensure that one is operating along the more efficient of the curves in Figure 2, for the purposes of a following main experiment. The empirical Monte Carlo data used in Figure 3 are actually the same as those in Figure 2. However, the standard deviations are now expressed in logit units with respect to the assumed value of k . True logit units were used in Figure 2, for comparisons of efficiency. The procedure works, with Table 1, as follows.

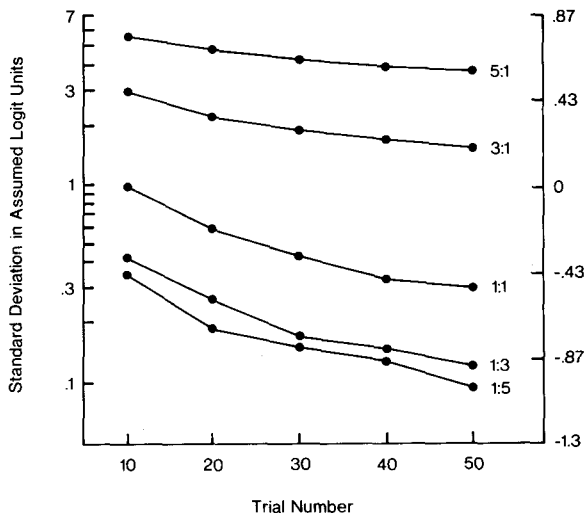


Figure 3. The data of Figure 2 replotted in assumed logit units rather than true logit units, to facilitate rough empirical slope estimation from brief pilot experiments. Logarithmic spacing is used to separate the curves, and the labels on the right-hand scale are common logarithms of the standard deviations.

The first step is to make a guess of the value of k . The slope of the logistic psychometric function is $k/4$, so this guess for k may be taken as four times a guessed slope. Another basis for a guess is $k = 4 / (X1 - X0)$, where $X1$ and $X0$ are the upper and lower limits of a stimulus range that seems almost certain to contain the threshold. The second step is to determine a stimulus range, R , as $R = 50/k$, for the guessed value of k . Then the upper limit of the stimulus range will be $m + R/2$ and the lower limit $m - R/2$, where m is an initial guess of the threshold level. The third step is to make several empirical threshold estimates under the same conditions, using the same number of trials, say 20, by the simplified method associated with Table 1. The standard deviation of these, multiplied by the assumed value of k , is then compared to the curves in Figure 3 at the appropriate trial number. The label on the nearest curve then is an estimate of how the assumed value of k relates to the true value. For example, if assumed k were .23 and the empirical standard deviation were 10.21, for estimates based on 20 trials, then the product, 2.35, would be found to be quite near the curve labeled 3:1. The inference is that the assumed k was too large by a factor of about 3. Therefore, a revised estimate is $.077 (.23/3)$.

A repetition of this whole pilot procedure with the new estimate of k may reveal whether the process is having the desired effect. Repetitions should theoretically converge toward the appropriate point on the curve labeled 1:1 in Figure 3. However, sampling errors of the estimated standard deviation would ordinarily prevent the exact achievement of that goal. In practice, one would probably be quite pleased with an indicated ratio between 1:1 and 1:3. Indeed, because of the sampling error problem, and the potential bias with overestimates of k , an indicated underestimate of 1:3 might be preferred over a perfect match.

Summary of the Simplified Procedure

The empirical findings have provided information that can help a great deal toward the optimization of the use of Table 1, or of the maximum likelihood method itself. For using Table 1, the first step is to adopt an estimate of the value of k . The method described in the preceding section can be used, if there is insufficient prior information about the value of k . The range of stimuli, R , is then determined as $R = 50/k$, for the adopted value of k , and the upper and lower limits of the stimuli are $m + R/2$ and $m - R/2$, where m is the initial estimate of the threshold. Then the step sizes in the sequence of trials are those given in Table 1 multiplied by $R/50$. Because of the way R is determined as described above, this rescaling factor of $R/50$ is equal to $1/k$, and the stimulus limits are $m + 25/k$ and $m - 25/k$. If either of these limits falls beyond the actual stimulus values that

can be produced practically, the solution is *not* to revise the value of R (or k). Instead, the proper solution is to omit trial 1, and consider starting with the step size of Trial 2 in Table 1. It may also be necessary to omit Trial 2, etc. The rule is that one may start with the trial whose step size (rescaled) is not more than half the distance from m to either of the limits of the practically available stimuli.

Table 1 can be extended at either end very simply for approximate purposes. Though there may not be much point in starting with larger steps than that determined by the 12.5 figure for Trial 1, it is clear that the sequence starts with a halving of step sizes. Thus, one could start with a step size of 25, as indicated in Table 1 for "Trial 0." Extension at the other end of the table for more than 49 trials is simple, because the process is essentially a standard asymptotically optimal Robbins-Monro process (Wetherill, 1963) by the 50th trial. For a smooth transition to trial numbers, n , greater than 49, take the step size to be $2/(n-5)$. The standard unmodified Robbins-Monro process mentioned above would specify $2/n$, but the difference is small. These step sizes must be rescaled in the same way as are those in Table 1, by the factor of $1/k$.

The Monte Carlo simulations revealed also that the standard error of estimate of the threshold on any given trial is approximately the step size for the trial multiplied by the square root of the trial number. This rule holds quite well at and beyond the fifth trial when the assumed and true values of k are equal and when one has not been forced to start with some trial other than Trial 1 in Table 1. Violations of those conditions would elevate the standard error somewhat above the value given by this simple calculation. Mild violations should have little effect, so stopping rules in the sequential process can be formulated usefully in terms of this simple estimate of the standard error. The number of trials is known a priori when such a stopping rule is used, and that is a convenience in the planning of experiments.

Conclusion

For the yes-no paradigm, good performance of the maximum likelihood method is obtained under mismatches of the assumed and true slope up to a factor of about 4 to 1. Bias in the obtained threshold estimates is obtained when the slope is overestimated by a factor of 5 to 1. For underestimates and less extreme overestimates, no bias was evident, but efficiency decreased approximately equally for equal

reciprocal mismatches. Overestimates and underestimates seemed to differ, however, in the trend of efficiency loss as a function of trial number. Underestimates suffered most on early trials, and overestimates most on later trials. Underestimation is more conservative than overestimation.

A case of the Robbins-Monro procedure developed especially to simulate the maximum likelihood method is almost as efficient, under most conditions, as the maximum likelihood method itself. Its behavior is quite similar to that of the maximum likelihood method in all of the above respects. It results in a drastic reduction of the computations between trials.

To increase efficiency in any main experiments to follow, there is a fairly simple method of obtaining slope information from brief pilot experiments.

The method described here is for the yes-no paradigm of threshold estimation when the threshold is defined as the 50th percentile. A different procedure would be required for forced choice when the threshold is defined differently, such as the 75th percentile.

REFERENCES

- DIXON, W. J., & MOOD, A. M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association*, **43**, 109-126.
- FINDLAY, J. M. (1978). Estimates on probability functions: A more virulent PEST. *Perception & Psychophysics*, **23**, 181-185.
- HALL, J. L. (1981). Hybrid adaptive procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America*, **69**, 1763-1769.
- PENTLAND, A. (1980). Maximum likelihood estimation: The best PEST. *Perception & Psychophysics*, **28**, 377-379.
- ROBBINS, H., & MONRO, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics*, **29**, 400-407.
- SHELTON, B. R., PICARDI, M. C., & GREEN, D. M. (1982). Comparison of three adaptive psychophysical procedures. *Journal of the Acoustical Society of America*, **71**, 1527-1533.
- SMITH, J. E. K. (1961). Stimulus programming in psychophysics. *Psychometrika*, **26**, 27-34.
- TAYLOR, M. M. (1970). On the efficiency of psychophysical measurement. *Journal of the Acoustical Society of America*, **49**, 505-508.
- TAYLOR, M. M., & CREELMAN, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, **41**, 782-787.
- TAYLOR, M. M., FORBES, S. M., & CREELMAN, C. D. (1983). PEST reduces bias in forced choice psychophysics. *Journal of the Acoustical Society of America*, **74**, 1367-1374.
- WETHERILL, G. B. (1963). Sequential estimation of quantal response curves. *Journal of the Royal Statistical Society (B)*, **25**, 1-48.

(Manuscript received June 18, 1984;
accepted for publication June 22, 1984.)