Perception & Psychophysics
1978, Vol. 23 (2), 181-185

Notes and Comment

Estimates on probability functions: A more virulent PEST

J. M. FINDLAY
University of Durham
South Road, Durham DH1 3LE, England

In many psychophysical situations, a sigmoid shaped "sensitivity" or "psychometric" function appears to relate the probability of an observer's response and the level of some physical variable. For example, in a detection experiment with stimuli presented to an observer on discrete trials, such a relationship frequently obtains between the probability of detection on a trial and the intensity of the stimulus.

A common experimental problem is to estimate the parameters of such a function from a set of observations. In traditional psychophysical procedures (e.g., the method of constant stimuli), the observer is presented with stimuli at intensity levels which have been selected prior to the commencement of the experiment. Parameter estimation from the data can be carried out using optimum statistical procedures (e.g., Finney, 1948). However, these methods suffer from the drawback that the observer is often presented with stimuli far removed from the range of interest for which the detection probability is either near zero or near unity. Little information is gained from these trials, and consequently, the methods are relatively slow. With increasing use of computers in the control of psychophysical experiments, some attention has been given to more flexible "staircase" methods in which the choice of intensity on a particular trial is dependent upon the subject's previous responses. Taylor and Creelman (1967) have put forward a method which can be used for the purpose of estimating the level corresponding to a particular probability. This method is called PEST (or Parameter Estimation by Sequential Testing). It is the purpose of this note to suggest ways in which this procedure can be improved to give faster or more accurate estimates. Estimates of the degree of improvement are obtained from computer simulations. The situation which will be considered is the symmetrical one in which the probability of response varies from 0% to 100% and the level to be estimated is that for which the probability is 50%.

I should like to thank Arthur Still for helpful comments on an earlier version of this paper.

THE SEQUENTIAL TESTING ALGORITHM

The procedure used is to present the observer with a number of trials at one particular testing level (e.g., stimulus intensity level), during which a cumulative count is taken of the number of trials with positive and negative responses. After an appropriate number of trials, determined by a rule of the algorithm (Rule A), a step is made to a new stimulus intensity level whose value is computed by a second rule of the algorithm (Rule B) in a way that depends upon the history of the subject's responses. Further trials are given at this level until Rule A requires another step, and so on. A stop rule (Rule C) determines when the sequence terminates. The operation of this procedure is now set out in a rather more general form with the specific rules used in PEST given in parentheses.

Sequence Initiation

Two parameters are selected either arbitrarily or on the basis of a priori information. These are the initial stimulus level, and the initial step size, I, which determines the change in level when a step is made.

Trial Presentation

A stimulus is given and the observer's response recorded. Following this, a check is made using Rule A to see if the stimulus level is to be changed for the subsequent trial. (In the original PEST, Rule A took the form that the level should change if the observed number of positive responses deviated from the value expected at the target point by some constant amount, W. Taylor and Creelman (1967) state that this is equivalent to a Wald sequential likelihood ratio test. W is an important parameter. If it is large, a larger amount of evidence must be accumulated at one level before a step is made, and vice versa).

Steps

When a change of level is demanded by Rule A, then Rule B is invoked to set the new level. The step size made will be dependent upon the initial step size and the sequence history. The direction of the step depends on whether more or less positive responses occur than the expected number. (In PEST, Rule B consists of four subrules: (1) The step size is halved if the direction of change is reversed. (2) No change is made in the step size for the second step in a given direction. (3) and (4) For steps beyond the second in a given direction, the step size will be doubled, although a subsidiary rule ensures that doubling does not always occur at the third step. It is

possible to express this procedure succinctly by noting that the step size is equal to $I/2^n$, where I is the initial step size and n is a single parameter, the "number of effective reversals," carried forward from the sequence history. It is initially 0, increases by 1 every time a reversal of step direction occurs, and decreases by 1 every time a "doubling" occurs).

Sequence Terminations

Rule C states when the sequence terminates. (In PEST, this rule terminates the sequence when the step size falls below some preset minimum, i.e., when the parameter n as defined above exceeds a particular value).

Estimation Procedure

The sequential testing procedure is always attempting to home in on the desired level, and so the final testing level (i.e., that which would have been presented on the trial occurring after Rule C operates) may be used for the estimate. A slightly more accurate, but more laborious, procedure is to use the whole of the data obtained to fit a psychometric function.

BASIS OF MODIFICATION TO PEST PROCEDURE

Two modifications to the PEST procedure are suggested. Their effect is assessed by simulations in the next section.

The first change takes account of information which is discarded in the original procedure. Rule A of PEST states that a change of level is made whenever the observed responses differ from those expected at the target position by a value greater than W. For example, if the target position is the 50% point and W = 1.0, this has the effect that a change is required as soon as the number of positive responses differs from the number of negative responses by three. Thus a step is made after three trials if the response distribution over positive and negative responses is 0-3 or 3-0; after five trials, if the response distribution is 4-1 or 1-4, etc. In the original procedure of PEST, no distinction was made between these cases. Consequently, to take an extreme example, the procedure followed the same course after the observer had made 3 positive responses out of 3 as when he had made 13 positive responses out of 23.

The latter distribution, however, would be unlikely to arise unless the level was already quite close to the 50% point, whereas the former is indicative only of a level that is likely to be displaced from the 50% point in one direction. By failing to distinguish between these cases, information is lost. There are various ways in which this information might potentially be used. One that has proved successful both in simulations and in experiments with human subjects is the following. Whenever the number of trials at a particular level exceeds some preset value, then an increment is made in n, the number of effective reversals. This has the consequence that subsequent steps are smaller, with the intention of giving more rapid convergence.

The second modifcation also concerns Rule A of PEST. As Taylor and Creelman point out, use of a small value of W results in rapid convergence to near the target level, but the final accuracy suffers because there is a significant probability of a step in the wrong direction when the testing level is close to the target. For large W, the converse occurs; more trials are required at each level, but the steps are more frequently in the correct direction. A straightforward compromise appears to be to gradually increase W as the series progresses. It has proved possible to implement this successfully by making W a function of n.

SIMULATIONS

Two factors are of concern in assessing the efficiency of any procedure designed to establish a psychophysical target. These are, first, the accuracy of the target estimate, and second, the time (i.e., the number of trials) needed to make the estimation. There will normally be a tradeoff between the speed and the accuracy of a procedure.

A suitable way to indicate the potency of a particular procedure and thus compare procedures is to show how response accuracy is related to the number of estimation trials. This is effected in the simulation results below. The results on any particular trial are subject to the usual sampling vicissitudes, and so the estimates given are derived from 500 simulation trials. For each trial, a particular stop rule is used and a figure stored for the final target setting and the number of trials taken to reach that point. Setting accuracy is then estimated as the standard deviation of the 500 trial settings, and this is plotted against the mean trial number at which the stop rule was invoked. A record is also taken of the mean of the final settings to check for bias. The stop rule of PEST involves termination of the sequence when the step falls below a certain size. In all cases discussed here, step size is directly related to the variable n, which is the number of effective reversals. On each trial, the simulations take a record of the setting point and trial number when n first reaches 2, 3, 4, and 5.

The simulations were made in a way quite similar to that of Taylor and Creelman (1967). The assumed psychometric function was the logistic function:

$$P(+) = \frac{1}{1 + e^{-L}},$$

where P(+) gives the probability of a positive response as a function of L, the stimulus level. The logistic function has the property that the probability of a positive response changes in a symmetrical ogive from near zero for large negative values of L to near unity for large positive values. It has the value .5 at L = 0 and changes from the value of .2689 to that of .7311 as L changes from −1 to +1.

In the simulation trials, the target level is the 50% probability point. An initial value of L was chosen using a randomization procedure to lie at some point between 5 and 10 logit units from the target (the probabilities of all values within this interval being distributed uniformly). The initial step size was 5 logit units. This was chosen to represent a typical practical situation, and it seems unlikely that much generality is sacrificed by this restriction, although it is obvious that more trials would be required for an initial point further removed from the target.

RESULTS OF SIMULATIONS

The simulation procedure described above was carried out using the original PEST procedure and the various modifications discussed. Examination of the results showed that in general all the procedures tested appeared to give unbiased estimates. The bias was assessed by examining the difference between the mean of the levels reached on the 500 simulation trials and the true target level (L = 0). Since the simulation trials always approach the target from the same direction, any systematic bias in settings should be revealed. In only one case is there a substantial difference from the target level. This is the setting after two "reversals" with M = 4 and W = 2 (Figure 2). The reason, in retrospect, is obvious. In this case, the result of the modification is that n is incremented at each level tested (since at least four trials are required before a step can be made at W = 2), and so the procedure cannot reach the target level without at least three steps. The settings in the same simulation after three "reversals" also show a slight bias. The magnitude of this is approximately .3 logit units, which is less than half the standard deviation of the settings. In all other cases, the bias found is less than .2 logit units at n = 2 and less than .12 logit units in all other cases. Although marginally greater than the figures expected from calculations of a standard error, these values are obviously of no practical significance.

The first modification considered is the one which concerns Rule A of PEST—relating the step size to

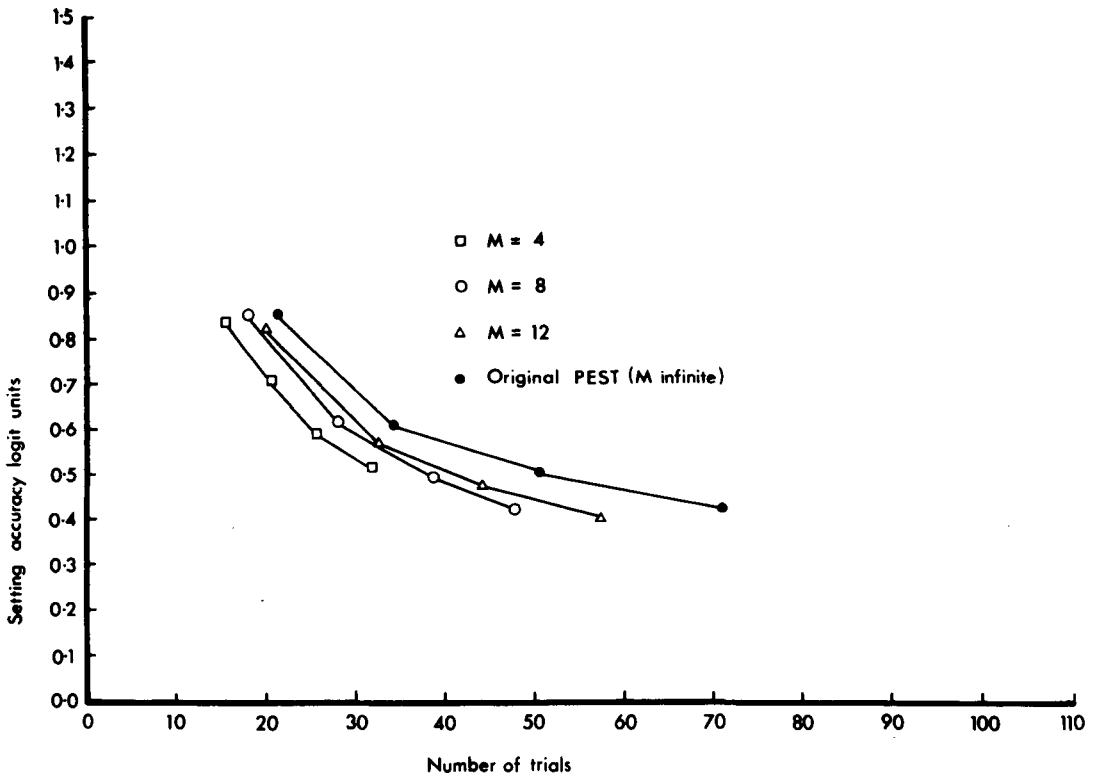

Figure 1. Dependence of accuracy of the estimation procedure on the number of trials. The points on each graph give the setting accuracy and the average number of trials when the stop rule is set for, successively, 2, 3, 4, and 5 "effective reversals." Each point shows the average of 500 simulation trials. The figure shows the effect of Modification 1 in which the "number of effective reversals" is incremented when the number of trials at a level exceeds M. Results are shown for various values of M. The criterion for changing level (W = 1) gives fast but relatively inaccurate convergence.

the number of trials, N, taken at the previous stimulus level before the rule was invoked. In the original version, the step size was independent of this number, but, as argued earlier, this results in an unnecessary loss of information. The modifications made here all take the form of making the parameter n (the number of effective reversals), which is carried forward in the trial sequence, depend upon N in the following fashion:

$$n = \text{integer part of } N/M,$$

where M is a constant for a particular simulation. Thus, if M = 8, the value of n is increased by 1 if 8 or more trials have occurred at the previous stimulus level, by 2 if 16 or more trials occurred, and so on. The original PEST may be taken as the limit of this procedure as M becomes large. Figures 1 and 2 show the effect of these changes. In Figure 1, the value W, used in the Wald test invoked as Rule A to determine whether a change in stimulus level should be made, is set at 1, which gives fast but relatively inaccurate decisions. In Figure 2, the value of W is set to 2, which corresponds to a slower, but more accurate, procedure.

Figure 3 shows that simulation results in which the second modification is incorporated. In this, the value of W does not remain constant but changes as the sequence proceeds, so that, in effect, more evidence is needed before a change in level is made as the stimulus level approaches the target level. The results given here are from a procedure in which W was made to depend upon n in the following way:

$$W = .5 \quad n \leqslant 1$$

$$\frac{n}{2} \quad n > 1.$$

This function has no precise theoretical justification and is almost certainly not optimal, but it appeared appropriate on the basis of earlier simulations.

DISCUSSION

It is clear from the simulation results that both modifications do, indeed, lead to an improvement in the speed/accuracy function. For the first modification, the simulations have been carried out for various values of the parameter M, which is the number of trials at a given level before the "number of effective reversals" is incremented. The power of the procedure

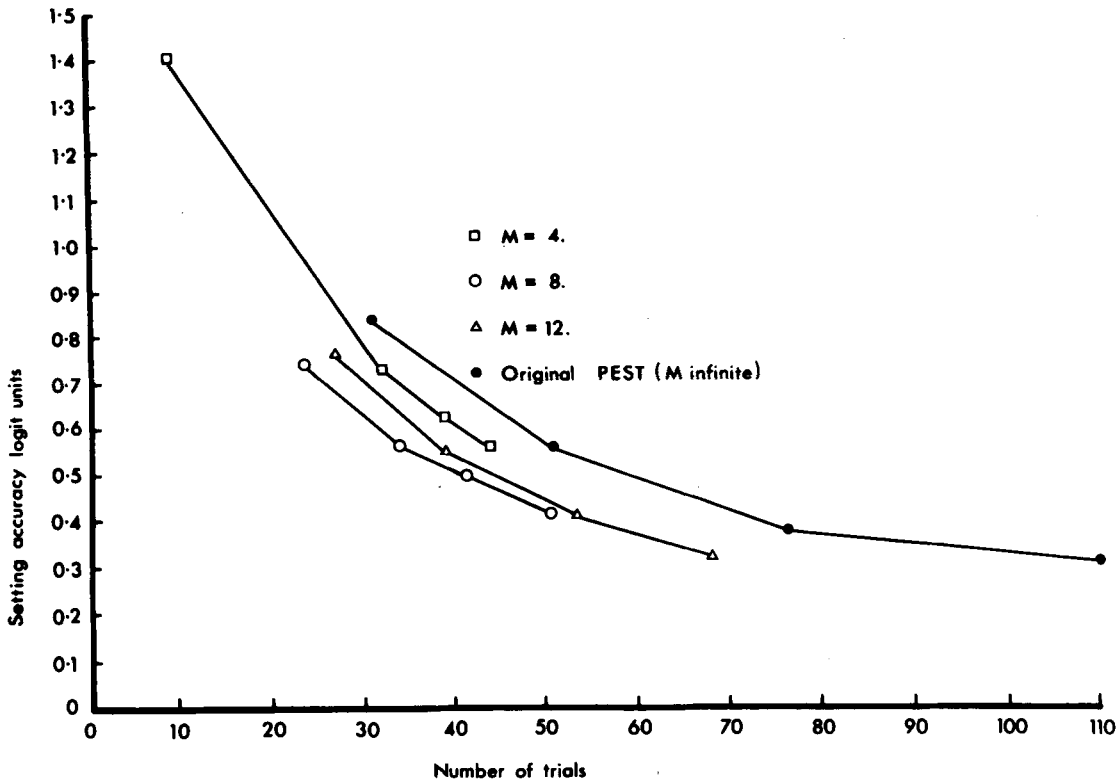

Figure 2. Dependence of accuracy of the estimation procedure on the number of trials. The points on each graph give the setting accuracy and the average number of trials when the stop rule is set for, successively, 2, 3, 4, and 5 "effective reversals." Each point shows the average of 500 simulation trials. The figure shows the effect of Modification 1 in which the "number of effective reversals" is incremented when the number of trials at a level exceeds M. Results are shown for various values of M. The criterion for changing level (W = 2) gives relatively slow, but accurate, convergence.

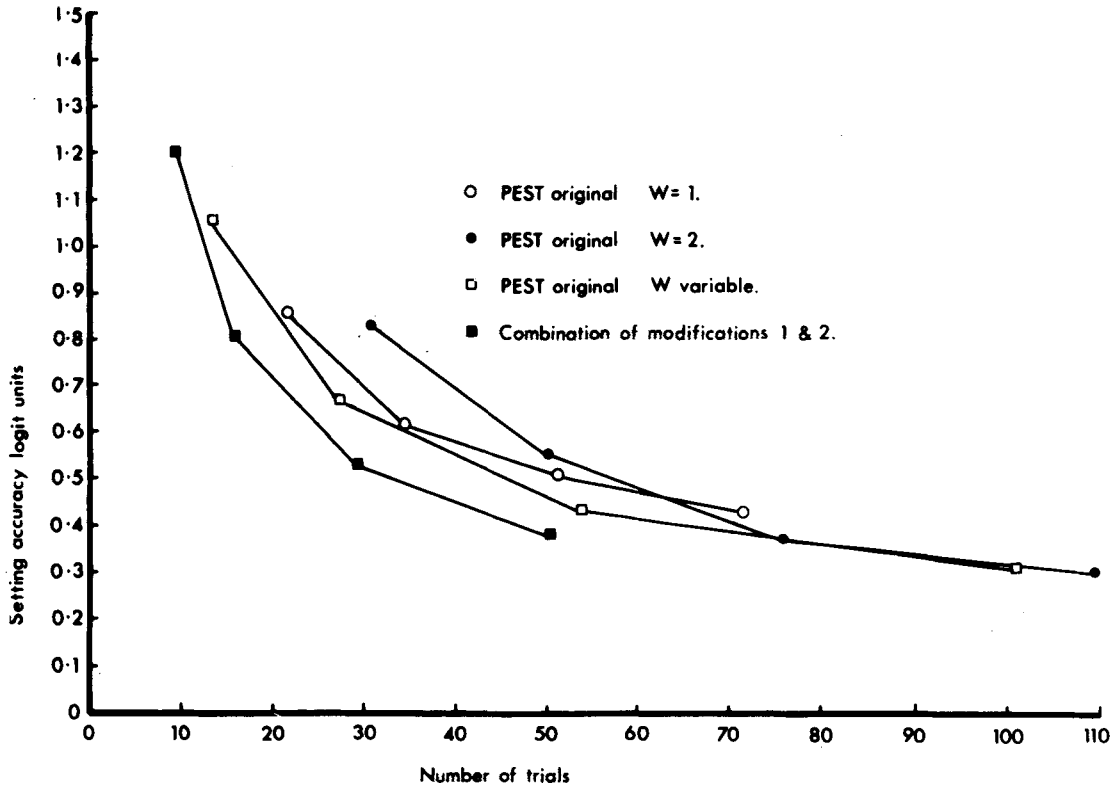

Figure 3. Dependence of accuracy of the estimation procedure on the number of trials. The points on each graph give the setting accuracy and the average number of trials when the stop rule is set for, successively, 2, 3, 4, and 5 "effective reversals." Each point shows the average of 500 simulation trials. The figure shows the effect of introducing Modification 2 in which the criterion for a change of step (W) varies during the procedure.

is shown to be dependent on the value of M, and in both cases tested, an optimum value of M occurs. For the fast-inaccurate case (W = 1), the optimum appears to be around 4, whereas for the slow-accurate case, a value around 8 gives the best performance. The degree of improvement is quite considerable; it is possible to achieve a setting with a particular degree of accuracy in 15-20 fewer trials than in the original PEST formulation.

The second modification, results from which are set out in Figure 3, achieves an improvement also, although the magnitude of this is smaller. Figure 3 shows the results of a simulation in a condition where both modifications were incorporated, the speed/accuracy curve for this condition is the best that has been obtained.

The general form of the speed/accuracy curve in all simulations is of some interest. It shows that the rate of improvement in accuracy with increasing trials becomes progressively slower. This is largely a consequence of the sampling variance; even with optimally chosen intensity values, setting accuracy would be dependent upon the square root of the number of trials. It is evident that the improvement in accuracy beyond 30 or 40 trials will rarely justify the extra effort, although in practical situations, the observer's behavior might be subject to other sources of variation so that more trials might be desirable for greater reliability.

REFERENCES

FINNEY, D. J. *Probit analysis.* Cambridge, England: Cambridge University Press, 1948.

TAYLOR, M. M., & CREELMAN, C. D. PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 1967, **41**, 782-787.

Received for publication September 14, 1977;
accepted December 14, 1977.)