Statistical properties of Pollack's PEST procedure*

PAULA RENDLEMAN, RICHARD M. ROSE, and DAVIDA Y. TELLER†
University of Washington, Seattle, Washington 98105

The bias and variability of threshold estimates produced by Pollack's PEST procedure were studied by means of repeated computer simulations of PEST runs. The influence of the size of the exit criterion, the bumping-into-zero rule, and three methods for obtaining the estimate from the data are discussed. Pollack's PEST procedure is found to result in a highly variable number of trials per run, and to be approximately as good as a simpler staircase procedure in terms of the bias and variability of the estimates obtained.

In a sizable proportion of psychophysical experiments, an attempt is made to determine the level of an independent variable that leads to a particular response probability. For example, an E may wish to determine the stimulus intensity necessary for a S to say "Yes, I detect the stimulus" on 50% of the trials, or the intensity necessary for him to respond correctly with some specified probability between 50% and 100% in a two-alternative forced-choice experiment. The intensity level found in such an experiment is often called the threshold.

The methods psychophysicists use to determine thresholds can be divided into constant-stimulus or fixed psychophysical methods, and adaptive or titration methods. In the classical fixed methods, trials are run at each of several fixed stimulus-intensity levels, and a response probability is estimated for each. The threshold is then obtained by interpolation along a psychometric function fit to the response-probability estimates. Adaptive methods do not require equal numbers of trials at all stimulus levels; rather, the current intensity level is determined by the sequential history of the experimental run.

There is a variety of adaptive methods—e.g., the block up-and-down method (Stuckey, Hutton, & Campbell, 1966), the staircase or up-and-down method (Dixon & Mood, 1948; Cornsweet, 1962), and the various procedures known as PEST (Parametric Estimation by Sequential Testing; Taylor & Creelman, 1967; Pollack, 1968). These methods all follow the same basic pattern. The E begins by running a block of trials at some initial stimulus level, and after a certain number

of trials at that level, a block of trials is begun at a new stimulus level. The choice of each new level depends on the history of the run, and the rules for changing levels are designed so that the sequence of stimulus intensities tends towards the level that the E desires to estimate. At the end of the run the desired stimulus level is estimated by one of a variety of methods of calculation.

Taylor and Creelman (1967) provide an extensive discussion of the PEST procedure. This method is usually used as a two-alternative forced-choice procedure. In PEST the stimulus intensity is not increased or decreased by a fixed amount; instead, this amount is variable and depends upon recent responses in the run. Pollack (1968) advocates a simplified version of PEST's complex sequential rules, and claims that in a two-alternative forced-choice psychoacoustical task it provides a quick estimate of the stimulus level associated with the 0.75 point on the psychometric function.

In a previous paper (Rose, Teller, & Rendleman, 1970), we have studied the properties of yes-no and two-alternative forced-choice staircases by means of repeated computer simulations of staircase runs. The reader is referred to that paper for detailed discussion of the methods and problems of staircase simulations. In the present project we have studied Pollack's version of the PEST procedure. In particular, we have addressed four issues: (1) the influence of the size of the exit criterion on the bias and variability of the estimate; (2) the effect on bias and variability of three different methods of calculating the threshold estimate—the method used in PEST, the mean of the intensities used in the run, and the peak-and-valley method suggested by Wetherill and Levitt (1965); (3) possible effects introduced by Pollack's bumping-into-zero rule (defined below); and (4) the relative usefulness of Pollack's version of PEST as compared to staircase procedures.

POLLACK'S PEST PROCEDURE

Pollack (1968) has specified a set of

rules for using the PEST procedure in two-alternative forced-choice tasks. Pollack's rules may be restated as follows:

(1) Starting Intensity Rule

The initial stimulus intensity presented is set at a level which rough preliminary tests indicate is well above the threshold.

(2) Rules for Changing Levels

In Pollack's procedure a trial block consists of two to four successive trials at a fixed intensity level. At the end of the block, the stimulus level is increased following two incorrect responses (out of two, three, or four trials); the level is not changed if three correct responses out of four trials occur; and the level is decreased if four correct responses out of four trials occur. After a trial block is completed, the E uses these rules to determine whether or not to change levels and in which direction, and begins a new trial block of two to four successive trials at that level.

(3) Step Size Rules

Reversal rule. The difference in stimulus intensity between successive stimulus levels (hereafter called the step size) is halved following each reversal of step direction, i.e., when an increase in stimulus level occurs after a decrease, or vice versa.

Doubling rule. The step size is doubled upon the occurrence of a third successive step in the same direction, i.e., after three successive increases in stimulus intensity or three successive decreases in stimulus intensity. Otherwise the step size is the same as the previous step size.

Maximum rule. Before a run begins, a maximum step size is determined. Whenever the doubling rule calls for a step size greater than this maximum, the step size is set equal to the maximum.

Initial step size rule. The initial step size is always the maximum step size.

(4) Bumping-into-Zero Rule

In most experimental situations, the stimulus intensity is measured with a logarithmic scale, such as a decibel scale, and changes in stimulus level are stepped off in logarithmic units. In such cases, the step-size rules never call for a stimulus intensity at or below zero. Occasionally, a linear scale is used, as in the time-difference experiments performed by Pollack (1968). When this is done, changes in stimulus level are stepped off along a linear, rather than a logarithmic, scale. Then it is possible for the step-size rules to call for a change in stimulus to some level below zero on the stimulus scale. If this occurs, then under Pollack's rules the step size is successively halved until the stimulus level called for is above the zero level.

*This research was supported in part by Undergraduate Research Participation Grant No. GY-6090 from the National Science Foundation, by PHS Grant No. MH17761-01 from the National Institute of Mental Health, and by PHS Grant No. EY-00421-02 from the National Eye Institute. Computer funds were provided by the Department of Psychology, University of Washington.

†Requests for reprints should be sent to R. M. Rose, Department of Psychology, University of Washington, Seattle, Washington 98105.

(5) Exit Rule and PEST Estimator

A run is terminated if the step size upon the next trial is less than an arbitrary preset criterion, called the *exit criterion*. The stimulus level that would have been presented on the next trial is the estimate of the threshold used by Pollack. This estimate will be called the PEST estimate.

Outcome Expectations

What probability level should one expect Pollack's procedure to estimate? We know of no explicit derivation of the mean and variance of this estimator. (It is for this reason that we have run computer simulations of the procedure.) Pollack (1968) and Wales and Blake (1970) suggest that these rules yield an estimate of the 0.75 point on the psychometric function. Wetherill and Levitt (1965) and Rose et al (1970) have suggested that a reasonable expectation is that a given method will yield an estimate of the *balance point*; that is, that level at which the probability of an increase in the stimulus level equals the probability of a decrease in it.

It is easy to demonstrate that, with Pollack's rules for changing levels, the balance point occurs at the 0.734 point on the psychometric function. Suppose that a level, i , has just been selected and that the associated point on the psychometric function is $p(i)$. Then, the probability that the next stimulus level will be lower than i is the probability of four correct responses, or $p(i)^4$. Let C denote a correct response, and $\bar{C}$ denote an incorrect response. Then the probability that the next stimulus level will be higher than i is the probability of occurrence of one of the following response tuples: $\bar{C}\bar{C}$, $C\bar{C}\bar{C}$, $\bar{C}C\bar{C}$, $\bar{C}\bar{C}C$, $CC\bar{C}\bar{C}$, $\bar{C}CC\bar{C}$, $\bar{C}\bar{C}CC$. That probability is

$$[1 - p(i)]^2 + 2p(i)[1 - p(i)]^2 + 3p(i)^2[1 - p(i)]^2.$$

The point of stability at which the probability of a stimulus increase is the same as the probability of a stimulus decrease is the point at which

$$p(i)^4 = [1 - p(i)]^2 + 2p(i)[1 - p(i)]^2 + 3p(i)^2[1 - p(i)]^2.$$

The solution to this equation is $p(i) = 0.734$.

It is, of course, an empirical question whether or not a given procedure will yield threshold estimates whose average is equal to the calculated balance point. If it does, the procedure will be said to be *unbiased*; if it does not, it will be said to be *biased*.

Fig. 1. Schematic representation of a PEST run resulting in a spuriously low estimate, due to repeated application of the bumping-into-zero rule (Pollack, 1968). Response probabilities are represented on the ordinate and trials on the abscissa. A 1.25% criterion is assumed, and the ordinate is marked off in units the size of the exit criterion. It is assumed that all responses happen to be correct. The diagram traces out the stimuli presented as a function of trials. The stars indicate the points at which the bumping-into-zero rule is called for, and the numbers indicate in each case the number of exit criterion units by which the stimulus intensity is reduced.

METHOD

The Psychometric Function and Bumping into Zero

The function relating probability of a correct response to stimulus level is called the psychometric function. In symmetrical two-alternative forced-choice procedures, the probability of a correct response is always at least 0.5, because the S can be right that often by guessing. The psychometric function, then, must lie in the probability interval, $[0.5, 1.0]$.

Rose et al (1970) have shown that in staircase simulations, the use of linear and normal-integral psychometric functions yield essentially identical results. Hence in the present simulation of the two-alternative PEST procedure, we have used a linear psychometric function; that is, the function took on the value, 0.5, for all stimulus levels less than some level I_0 ; took on the value, 1.0, for all stimulus levels greater than a higher level, I_1 ; and rose linearly from 0.5 to 1.0 in the interval, $[I_0, I_1]$.

Pollack's bumping-into-zero rule is needed if a linear stimulus scale is used, and the PEST rules call for negative stimulus values. The closer the lower bound, I_0 , of the psychometric function is to the zero of the stimulus scale, the more likely it is that the bumping-into-zero rule will be called into play. In the most extreme case, it is possible that I_0 (the

highest stimulus level with response probability 0.5) could coincide with the zero on the stimulus scale. In this case the bumping-into-zero rule would be used relatively often. At the other extreme, there may be many physically realizable values of the stimulus below I_0 , in which case the bumping-into-zero rule would never be called into play. (It is interesting to note that a logarithmic scaling of stimulus values guarantees that the latter case will obtain, and, conversely, that the bumping-into-zero rule accomplishes something akin to a logarithmic scaling of stimulus values as the zero stimulus level is approached.)

It seemed likely that the use of the bumping-into-zero rule could lead to some artifacts, particularly to spurious meetings of the exit criterion. Thus, one might expect that the estimates obtained with Pollack's PEST rules would be biased low. (An example of a sequence that would result in a spuriously low estimate is diagrammed in Fig. 1.) To test this possibility, we have used the two extreme cases of the psychometric function described above: (1) a function for which I_0 is the lowest stimulus value available, resulting in frequent use of the bumping-into-zero rule; (2) a function with an unlimited number of stimulus values available below I_0 , with which the bumping-into-zero rule would never be

Table 1
Unbounded Psychometric Function

Exit Criterion	PEST		Mean-Intensity		Peak and Valley		Number of Trials		
Percent of $[I_0, I_1]$	Threshold Estimate	SD	Threshold Estimate	SD	Threshold Estimate	SD	Mean	SD	SD/Mean
1.25	.733	.059	.727	.058	.715	.057	129.8	71.8	.55
2.5	.729	.063	.727	.050	.713	.053	100.4	53.2	.53
5.0	.721	.080	.716	.080	.702	.067	78.8	45.2	.57
10.0	.708	.110	.693	.111	.684	.099	60.2	33.2	.55
20.0	.700	.146	.697	.148	.704	.136	44.7	25.9	.58

Table 2
Truncated Psychometric Function

Exit Criterion	PEST		Mean-Intensity		Peak and Valley		Number of Trials		
Percent of $[I_0, I_1]$	Threshold Estimate	SD	Threshold Estimate	SD	Threshold Estimate	SD	Mean	SD	SD/Mean
1.25	.717	.062	.732	.048	.729	.054	110.8	62.1	.56
2.5	.713	.070	.729	.053	.724	.061	85.6	46.5	.54
5.0	.709	.078	.730	.057	.727	.071	61.0	33.9	.56
10.0	.704	.091	.732	.067	.734	.110	43.8	27.7	.63
20.0	.690	.112	.736	.084	.746	.134	30.5	16.5	.54

used. These two cases will be termed, respectively, the *truncated* and the *unbounded* psychometric functions.

Input Specifications

To use Pollack's rules, the E must initially specify (1) the exit criterion, (2) the starting intensity, and (3) the maximum step size. The exit criterion suggested by Pollack was 5% of the interval (I_0, I_1). We ran simulations using each of the following exit criteria: 1.25%, 2.5%, 5%, 10%, and 20% of the interval (I_0, I_1). We used the starting intensity that Pollack suggested for best results. Taking the 5% exit criterion as an intensity unit, the starting stimulus was always 31 units above I_0 . The maximum step size was 32 times the size of the exit criterion.

Computer Simulation

The PEST procedure was simulated on a computer using a random-number generator. The input consisted of the PEST rules, the input specifications, and either the unbounded or the truncated psychometric function. The computer used the response probabilities of the psychometric function to generate correct and incorrect responses, thus essentially running itself as a S.

After the exit criterion was reached, the computer estimated the threshold level by three different techniques. It first used the PEST estimator, that is, the stimulus level that would have been presented on the trial following termination of the run. Second, it used the technique analogous to the technique often associated with the staircase procedure, namely, the average of

all the stimulus levels visited after the first three trial blocks. The third estimate of threshold was calculated by the peak-and-valley technique (Wetherill & Levitt, 1965), that is, the average of those stimulus levels at which reversals in step direction occurred.

The PEST procedure was repeated 300 times for each set of input specifications, each repetition yielding the above three estimates of the threshold. The computer then calculated for each of the three estimators the mean and standard deviation of each set of 300 estimates, and the mean and standard deviations of the number of trials per run.

RESULTS

The results for the five exit criteria and the three scoring procedures are shown in

Tables 1 and 2. Table 1 presents the data generated by using the unbounded psychometric function, i.e., the case with sufficient stimulus levels below I_0 such that the bumping-into-zero rule was never called into play. Table 2 presents the data generated by using the truncated psychometric function, i.e., the case in which I_0 was the lowest stimulus level available.

If the procedures indeed yield an estimate of the balance point, the average threshold estimate should be at the intensity associated with the 0.734 probability of making the correct response. To permit an easier evaluation of the results, the mean intensity estimates have been transformed into the corresponding response probabilities on the psychometric function.

The standard deviation of each of the estimates presented in Tables 1 and 2 is also expressed in probability units. Thus, for example, from Table 1, a ± 2 SD interval about the mean for the PEST estimate with the smallest exit criterion is (0.615, 0.851).

Tables 1 and 2 also include the mean, the standard deviation, and the ratio of the standard deviation to the mean of the number of trials per simulation. In addition, the histograms in Figs. 2, 3, and 4 show the distribution of the number of trials per PEST simulation for the 1.25%, 5%, and 20% exit criteria. The solid and open bars of the histograms show, respectively, the data generated with the truncated and the unbounded psychometric functions.

DISCUSSION

Number of Trials

From Tables 1 and 2 and from Figs. 2, 3, and 4, we can see that the number of trials in a PEST run is extremely variable. For example, with an exit criterion of 5%

Fig. 2. Histograms of the number of trials per run for 300 simulations using an exit criterion of 1.25%.

Fig. 3. Histograms of the number of trials per run for 300 simulations using an exit criterion of 5%.

in 300 simulations, the number of trials per run varied from 20 to 310 (Fig. 2). This large variability of the number of trials from run to run would be significant in an actual experimental situation. Due to the rules of the PEST procedure, neither the E nor the S knows beforehand how long the experimental session will last. Our results show that the number of trials per run can occasionally become very large, leading to very long sessions.

Exit Criteria

The variability of the estimates associated with the two largest exit criteria (Table 1) is too large for these exit criteria to be used in experiments from which precise estimates are required. Take, for example, an exit criterion of 10%. If we compute an interval of ± 2 SD about the mean of the PEST estimate and of the

mean-intensity estimate, we obtain the intervals, (0.500, 0.928) and (0.500, 0.915). In both cases the lower limit of the interval is below I_0 , the point of the psychometric function at which the probability of the correct response drops to 0.5. Thus, the threshold estimates produced by exit criteria of 10% and 20% are clearly too variable for experimental use. This result implies that at least 60 to 80 trials are necessary for a PEST run to produce meaningful results. Conversely, an E who finds himself using fewer than 60 to 80 trials on a majority of his runs is probably using an exit criterion that is large in relation to the interval, $[I_0, I_1]$. This result is at odds with Pollack's (1968) suggestion that a reasonable estimate of the threshold can be obtained in as few as 12 to 14 trials. The remainder of our discussion will be restricted to the results produced by exit criteria of 1.25%, 2.5%, and 5% of $[I_0, I_1]$.

It is not a surprise that refining the exit criterion (i.e., making it smaller and thus providing a more precise estimate of threshold) increases the average number of trials per run, since a more precise criterion must be met before a run may be terminated. If successive PEST trials were independent, then the effect of increasing the number of trials run by some factor, k , would be to decrease the standard deviation of the estimate by a factor equal to $\sqrt{k}$. The average reduction that occurred in our simulations can be calculated from the figures in Table 1. When the number of trials per run is increased from 45 to 130, one would expect a 41% reduction in the standard deviation if trials were independent. However, the average reduction is 60%. Thus, refining the exit criterion leads to an increased number of

trials per run, but the increased trials buy a lot in terms of the decreased variability of the estimates. Figure 5 provides a plot of the variability of the PEST estimate as a function of the average run length.

Estimators and Bias

Tables 1 and 2 show that the threshold estimates produced by all three scoring methods of estimation are below the 0.75 point and, furthermore, most are biased slightly low of 0.734, the balance point. The case most like Pollack's (truncated function, exit criterion of 5%, PEST estimator) appears in our simulations to give an estimate of the 0.709 point on the psychometric function.

The PEST estimator is simply that stimulus intensity that would have been presented on the trial following the meeting of the exit criterion. Calculation of the estimate thus makes no direct use of most of the data collected during a PEST run. Nevertheless, comparison of the means and standard deviations across the three estimators shows that the PEST estimator is virtually as good as the mean-intensity and peak-and-valley estimators (see Table 1). Use of the mean-intensity and peak-and-valley estimators does not lead to significant improvement over the PEST estimator.

Unbounded vs Truncated Psychometric Function

Table 2 presents the data generated when the truncated psychometric function and the bumping-into-zero rule were used. A comparison with the data generated by the unbounded psychometric function (Table 1) shows, contrary to our expectations, no important differences for exit criteria of 1.25%, 2.5%, and 5%. We had expected the bumping-into-zero rule to have a significant effect on the computer simulations. Because the rule restricts the range of the psychometric function (and thus the levels at which the E may test), it can create an artificial meeting of the exit criterion due to successive halving of the step size upon approaching the zero-intensity level. We expected that such spurious sequences would contribute a negative bias and increased variability to the estimates. (See Fig. 1 for an example of one such sequence which would terminate in a false exit due to the bumping-into-zero rule and lead to a very low estimate of the threshold.) Evidently, sequences of this sort occur with sufficiently low probability that the bias and variability are not affected by them.

CONCLUSION

The results of our simulations reveal that PEST, at least with the particular set of

Fig. 4. Histograms of the number of trials per run for 300 simulations using an exit criterion of 20%.

Fig. 5. Comparison of estimates generated by Pollack's PEST procedure to staircase estimates. Percentages refer to PEST exit criteria.

rules suggested by Pollack, has some properties that have not been generally recognized. First, the number of trials in a PEST run is quite variable. Thus, with a choice of stimuli and of exit criterion that results in 80 trials being run on the average, runs extending over more than 160 trials occur 10% of the time. This is a major disadvantage of the PEST technique, particularly so for real observers, whose internal state varies over time.

Second, our manipulations of PEST, i.e., the three different estimators and the elimination of the bumping-into-zero rule, did not result in significant improvement over the technique Pollack suggests for best results. Our suspicions that the bumping-into-zero rule would increase the negative bias proved unjustified. However, we found that our simulations required considerably more trials than Pollack's estimate of 12 to 14; indeed, 60 to 80 trials were found to be the minimum that would provide a reliable estimate. Further, our results also show that Pollack's PEST rules do not quite track the 0.75 point but estimate a point on the psychometric function biased slightly low of 0.75. A similar point has been made previously for other two-alternative forced-choice adaptive methods (Rose et al, 1970; Taylor & Creelman, 1969; Wales & Blake, 1970; Wetherill & Levitt, 1965).

We are now in a position to compare the PEST estimator with the fixed-trial two-alternative forced-choice staircase techniques studied in the previous paper (Rose et al, 1970).¹ Figure 5 presents plots

of the estimated variance for three different staircase procedures as a function of the number of trials per run. The parameter of these three functions is the number of intervals between adjacent stimuli necessary to span the interval, $[I_0, I_1]$; this number takes on the values 3, 5, and 10. In order to allow comparison of PEST and staircase methods, Fig. 5 also shows the estimated variance for the PEST estimator (using the unbounded psychometric function) plotted against the average number of trials per run.

For small numbers of trials, the staircase method has lower variability, although both methods may be too variable to be used in precise work. For larger numbers of trials, the PEST estimator gives variability roughly comparable to that produced by the five-interval staircase condition. It is noteworthy that Pollack's PEST rules do not appear to lead to a significant improvement over considerably simpler fixed-trial staircase procedures; and the PEST procedure has the added disadvantage that the number of trials per run is highly variable.

We conclude with an explicit statement of the limited nature of the present research. First, the originators of the PEST procedure use rules somewhat different from those used by Pollack. In particular, Taylor and Creelman (1967) use a more complicated version of the PEST technique, which employs a Wald sequential test for changing levels and slightly different step-size rules. In addition, Taylor and Creelman use an

unbounded psychometric function, and so have no use for a bumping-into-zero rule. The rules used in our simulations originated from Pollack's version of PEST, and our results should not be applied indiscriminately to other versions of PEST.

Finally, as in the previous paper (Rose et al, 1970), our conclusions are based on study of a simple (and probably overly simple) stat-subject. We have assumed the existence of no sequential dependencies between responses and of stability in time of the response probabilities (i.e., no learning, adaptation, or attention shifts). Violations of each of these assumptions may be found to some degree one place or another in the psychophysical literature. Insofar as real Ss differ from our stat-subject, caution should be used when our results are applied. It is conceivable that in an actual experiment a S could generate estimates with less bias than those generated by our stat-subject. However, it seems unlikely a priori that human Ss would produce estimates with less variability than we have reported here.

REFERENCES

- CORNSWEET, T. N. The staircase method in psychophysics. *American Journal of Psychology*, 1962, 75, 485-491.
- DIXON, W. J., & MOOD, A. M. A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association*, 1948, 43, 109-126.
- POLLACK, I. Methodological determination of the PEST (Parametric Estimation by Sequential Testing) procedure. *Perception & Psychophysics*, 1968, 3, 285-289.
- ROSE, R. M., TELLER, D. Y., & RENDLEMAN, P. Statistical properties of staircase estimators. *Perception & Psychophysics*, 1970, 8, 199-204.
- STUCKEY, C. W., HUTTON, C. L., & CAMPBELL, R. A. Decision rules in threshold determinations. *Journal of the Acoustical Society of America*, 1966, 40, 1174-1179.
- TAYLOR, M. M., & CREELMAN, C. D. PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 1967, 41, 782-787.
- TAYLOR, M. M., & CREELMAN, C. D. Some pitfalls in adaptive testing: Comments on "Temporal integration and periodicity pitch." *Journal of the Acoustical Society of America*, 1969, 46, 1581.
- WALES, R., & BLAKE, R. R. Rule for obtaining 75% threshold with the staircase method. *Journal of the Optical Society of America*, 1970, 60, 284-285.
- WETHERILL, G. B., & LEVITT, H. Sequential estimation of points on a psychometric function. *British Journal of Mathematical & Statistical Psychology*, 1965, 18, 1-10.

NOTE

1. Rose et al (1970) expressed the variability of these staircase estimators in a slightly different form in order to compare them to yes-no staircases.

(Accepted for publication July 6, 1970.)