Perception & Psychophysics
1980, 28 (4), 377-379

Notes and Comment

Maximum likelihood estimation: The best PEST

ALEX PENTLAND
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

A common experimental problem is to determine a "threshold" for some psychometric function. This threshold is often defined by an observer's response probability as a function of an independent, physical variable. As experimenters, we would like to manipulate the physical variable in order to determine the value of the physical variable that yields threshold response as quickly as possible and with the fewest number of measurements. Rapid determination of the threshold not only minimizes the amount of effort required, but also minimizes the effects of change in the observer's criteria.

This is a common problem in many fields, and over the years several techniques for parameter estimation have been developed. Dixon and Mood (1948) developed their classical up-down method for explosives research. Cornsweet (1962) discussed the use of the staircase method in psychophysics. Wetherill (1963) examined the problem more abstractly, using the context of bioassay for discussion.

These techniques generally suffer from the problem that often little information is gained with many of the measurements, so the procedures are inefficient. In order to increase efficiency, Taylor and Creelman (1967) devised a procedure called PEST (parameter estimation by sequential testing), which reduced the number of measurements needed to reach a given level of accuracy. This procedure was later improved by Findlay (1978).

Both PEST methods depend on the information gathered to date to guide further measurements. In these methods, the sampling efficiency (the amount of information gathered by each measurement) of the standard staircase method is kept high by adjusting the staircase step size on the basis of the information already gathered. In common with all staircase techniques, assumptions must be made about the form and the range of the psychophysical function whose threshold value is being measured. Normally, the psychophysical function is assumed to be an ogive- or sigmoid-shaped function, and the range within which the threshold is located is known.

The problem of how to minimize the number of measurements required in order to determine a threshold to within a given accuracy is, in general, solved by trying to predict the setting of the independent variable that will maximize the amount of information that can be obtained from each of the measurements. Each time a measurement is to be made, all of the previous measurements can be used to obtain a maximum likelihood estimate of the setting of the independent variable that will yield the most information. Because each measurement yields the maximum amount of information, the fewest number of measurements need to be made in order to reach a given accuracy. For the normal sigmoid-shaped psychophysical function, the point that will yield the maximum amount of information about the positioning of the entire curve is the 50% response point, at which the slope of the response curve is greatest.

This maximum likelihood technique starts its search for the threshold with what is essentially a binary search, which would be the most efficient search technique if the psychophysical function were a step function. As the search narrows down to the region of uncertainty near the threshold, the maximum likelihood technique becomes more like the standard staircase method, but with variable step size.

It turns out that the computations required in order to implement this technique are substantial. This presents no difficulty when experiments are controlled by a small computer, of course, but, at minimum, a small programmable calculator is required. Some multidimensional or otherwise complicated problems will not fit on a hand calculator.

Comparison with Other Techniques: A Simulation

Two factors are of concern in assessing the efficiency of parameter estimation techniques: the accuracy of the estimate and the number of trials or measurements required to achieve that accuracy. There is always a tradeoff between the speed and accuracy of this type of procedure.

The maximum likelihood method was compared with the two previous PEST methods and the conventional staircase technique by determining how the accuracy of their threshold estimate varies with the number of measurements taken.

The simulations were made in the manner of both Findlay (1978) and Taylor and Creelman (1967) in order to allow easy comparison. The assumed psychometric function was the logistic function:

$$P(+) = \frac{1.0}{(1.0 + e^{-L})},$$

where P(+) gives the probability of a positive re-

0031-5117/80/010377-03$00.55/0

sponse as a function of L, the stimulus level. This function produces a symmetrical ogive with positive response probabilities near zero for large negative values of L and near unity for large positive values of L. It has the value .5 at L = 0 and changes from the value .269 for L = −1 to .731 for L = +1.

In the simulation, the target probability or threshold is the 50% positive response point. The threshold value was randomly placed between 5 and 10 logit units from the starting point of the maximum likelihood procedure, which is always the center of the independent variable's range. These conditions were chosen by Findlay for his simulation as being representative of a typical testing situation. With a starting point further removed from the threshold, the advantages shown by the maximum likelihood technique are increased.

The results of the simulation using the maximum likelihood technique are shown in Figure 1, along with the results obtained by Findlay in his simulation for his improved PEST, the Taylor and Creelman PEST, and the standard staircase technique. The setting accuracy is given as the standard deviation of the threshold estimate after n trials, in logit units. Each point shown is based on 500 simulation trials, in order to give a noise-free estimate of performance.

It can be seen from Figure 1 that the maximum likelihood procedure normally requires less than *half* the number of trials required by the other techniques, an advantage that becomes even greater when

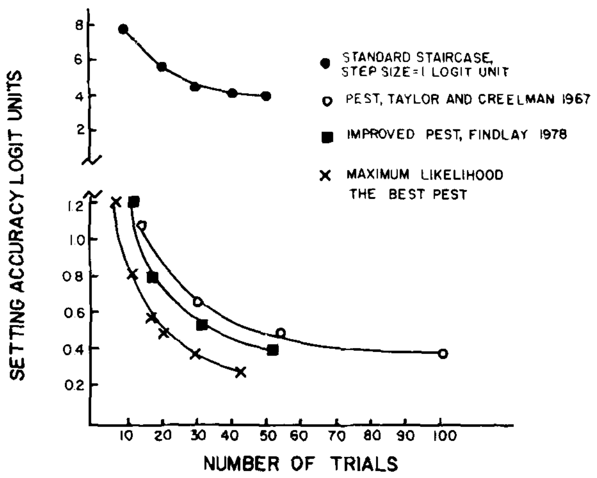

Figure 1. A comparison of the standard staircase technique, the Taylor and Creelman PEST, the Findlay improved PEST, and the maximum likelihood PEST (the best PEST). The maximum likelihood PEST can require fewer than half as many measurements to reach the same level of accuracy. The accuracy measure used here is the standard deviation of the estimator (in logit units) after n steps of the algorithm have been completed (i.e., n measurements have been taken). It is not possible to do substantially better with a standard staircase than the example shown, because reduced step size results in better asymptotic accuracy, but also in substantially more measurements being required.

higher accuracies are desired. The maximum likelihood procedure is, of course, unbiased. It is not possible to do substantially better with the standard staircase technique than the example shown, because using smaller step sizes to obtain greater asymptotic accuracy results in a much greater number of measurements' being required.

Discussion

It should be noted that this technique is as fully general as any staircase technique. Wherever a normal staircase may be used, it can be replaced with this maximum likelihood technique. Thus, several maximum likelihood staircases may be randomly interwoven to eliminate observer expectation effects, as with standard staircases.

In practice, the maximum likelihood technique has been shown not only to be more efficient than the other techniques, but also to be insensitive to occasional errors by the subject, errors concerning the range within which the threshold was assumed to exist, and errors concerning assumptions about the form of the psychophysical function. It has, therefore, proven of substantial practical use in making psychophysical measurements of many different types.

Of special note is the ease with which contrast-sensitivity functions can be measured using this technique. Using a two-alternative forced-choice paradigm, we have been able to regularly measure the contrast-sensitivity function in under 5 min, obtaining high accuracy measurements at 10 points along the function. This compares favorably with the method of Sekuler and Tynan (1977), which is modeled after the Békésy sweep-frequency audiometer. This technique may therefore be suitable for use in clinical applications.

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.

Findlay, J. M. Estimates on probability functions: A more virulent PEST. *Perception & Psychophysics*, 1978, **23**, 181-185.

Sekuler, R., & Tynan, P. Rapid measurement of contrast-sensitivity functions. *American Journal of Optometry and Physiological Optics*, 1977, **54**, 573-575.

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

Wetherill, G. B. Sequential estimation of quantal response curves. *Journal of the Royal Statistical Society*, Ser. B, 1963, **25**, 1-48.

APPENDIX

The approach taken to the problem of determining a threshold is to maximize the information gained with each measurement. If the amount of information gathered with each measurement is maximized, then the fewest possible number of measurements will be required.

For any one point, x, on the response curve, R, we have the probability, P_+, of a positive answer. Given n samples taken at x of which s were positive, our estimate of P_+ is s/n, the variance is $P_+(1 - P_+)/n$, and the width of the confidence intervals about s/n are:

$$C.I. = \frac{k\sqrt{P_+(1-P_+)}}{\sqrt{n}},$$

where k depends on the level of confidence desired (e.g., 90%, 95%, etc.). The width of the region corresponding to this confidence interval in terms of the independent variable I is (as we make k small) simply:

$$\frac{\partial I}{\partial R}\frac{\sqrt{P_+(1-P_+)}}{\sqrt{n}}. \quad (1)$$

Thus, to minimize the range of the independent variable I for a given number of samples, you should sample at the point at which Equation 1 is minimized, that is, when the variance of the response variable times the inverse of the slope of the response curve is at a minimum. In order to estimate the position of this point, we should use the maximum likelihood estimator, because it is known to be the most efficient unbiased estimator.

In the normal sigmoid-shaped psychophysical function, the maximum amount of information about the threshold is gained when sampling at or near the 50% positive response point. This is because the slope of the function is greatest there, and the variance least.

Thus our strategy is to obtain the best possible estimate of the 50% point and to sample there. This we may do by calculating the likelihood of the 50% point's being at each point within the independent variables range, and taking as our estimate of the 50% point the location that is the most likely point in that range. Thus, after n − 1 measurements, we find the n^{th} measurement, m_n, by solving:

$$m_n = \max_{x \in (a,b)} \Pr[x \text{ is } 50\% \text{ point} \mid (m_1, r_1), (m_2, r_2) \ldots (m_{n-1}, r_{n-1})],$$

where (a, b) is the range of the independent variable x, and the (m_i, r_i) denotes the results of the i^{th} measurement that was taken at value m_i of the independent variable. The value of r_i is +1 if the observer gave a positive response, and −1 if a negative response was obtained.

For the case of the sigmoid-shaped logit function, this may be rewritten as:

$$m_n = \max_{x \in (a,b)} \prod_{j=1}^{n-1} (1.0 + e^{-r_j(m_j - x)})^{-1}.$$

Before each measurement, we compute m_n, which is the maximum likelihood estimate of the position of the 50% point on the response curve, and then take the n^{th} measurement at position m_n. Note that at the start we already know that a measurement taken at point a of the independent variables range will give a negative response, and a measurement at point b will give a positive response. Thus, the first measurement is always taken at the center of the independent variable's range.

(Received for publication April 21, 1980; accepted June 20, 1980.)