Biol. Cybern. 53, 189–194 (1986)

Biological Cybernetics

Estimating the Variance of a Critical Stimulus Level from Sensory Performance Data*

David H. Foster

Department of Communication and Neuroscience, University of Keele, Keele, Staffordshire ST5 5BG, England

Abstract. Measures of sensory performance yielding a nonlinear dependence on stimulus level are often used to derive a critical stimulus level that corresponds to some criterion level of performance. Typical examples include the sigmoidal psychometric function used to estimate a "threshold" stimulus level, and the power-law increment-threshold curve used to estimate a "field sensitivity". Estimates of the variance of an estimated critical stimulus level derived from a single set of performance data are, however, infrequently reported, even though other estimates of reliability may not be available. An application of the classical "combination of observations" method is described here by which such variance estimates may be computed. The method was tested by applying it to sets of simulated psychometric-function data and increment-threshold data and comparing its results with those obtained by Monte-Carlo studies, each comprising 1000 runs. Differences between the estimated root mean variance of the estimated critical stimulus level and the "true" value were found to be not more than about 3% of the true value.

1 Introduction

Many measures of sensory performance produce nonlinear dependencies on stimulus level. Typical examples include the sigmoidal or logistic, *psychometric* function, relating the probability of a correct stimulus response to the level of the stimulus (Fig. 1a), and the power-law, *increment-threshold* function, relating the "threshold" increment of a test stimulus level to the background or field stimulus level (Fig. 2a). One purpose in obtaining such performance data is to estimate a *critical level* of the stimulus or some other factor that yields a criterion level of sensory performance. In the case of the psychometric function, this is the level of the stimulus (the "threshold") that corresponds to some target probability, for example 75% in a 2-alternative forced-choice task; in the case of the increment-threshold curve, this is the level of the field that corresponds to an increment threshold that is a fixed multiple of absolute threshold; the reciprocal of this field level is the "field sensitivity".

Estimates of the variance or standard deviation of an estimated critical stimulus level derived from a single set of performance data are not commonly reported. This may not be important where replication of the experiment is feasible and several sets of performance data may be used to obtain independent estimates of a critical stimulus level. In a number of experimental situations, however, there may not be opportunity to obtain more than one set of performance data and replication cannot then be used to assess reliability. Even when replication is possible, estimates of the variance of the individual estimates of the critical stimulus level are still needed to compute the best estimate of that level. The purpose of this communication is to draw attention to the usefulness of an approximate method belonging to the classical study of the "combination of observations". Under fairly general conditions, the method allows a good estimate to be made of the variance of an estimated critical stimulus level from performance data, and moreover the variances of any other variables (including the parameters of the model curve) depending on those data.

It may be noted that some existing computer software packages, for example those based on the Gauss-Newton algorithm, can routinely produce estimates of the variances of parameter-value estimates, as

* A portion of the data presented here was contained in a communication read at the Keele meeting of the Experimental Psychology Society, March 1985

do those procedures which transform the data so that traditional linear least-squares regression may be used. [For discussion of nonlinear least-squares regression modelling see Ratkowsky (1983) and Draper and Smith (1981, Chap. 10).] The method described here, however, is quite general. Its principal requirements are that some model- or curve-fitting procedure is available, that the function or algorithm relating the estimate of the critical level of the stimulus (or any other variable) to the experimental data is well behaved, and that the variances of the performance values are reasonably small. No other special constraints on the curve-fitting procedure or the estimate of the critical stimulus level or other variable are imposed. In particular, there is no requirement that the performance data be transformed nonlinearly (as in probit analysis), or that the goodness of fit be determined by least-squares, or that the performance values be treated as normal random variables, or that the matrix of 2$^{\text{nd}}$-order partial derivatives, the *Hessian*, should be known, or that the estimate of the critical stimulus level or other dependent variable should be expressible as a linear transformation of the parameter values of the performance curve.

2 Estimation of Variance by Combination of Observations

Let $y_1, y_2, \ldots, y_n$ be a set of n performance values measured at corresponding values $x_1, x_2, \ldots, x_n$ of the stimulus variable. For each pair of values (x_i, y_i), there is a random "error" which causes y_i to deviate from its expected value. Let T be the estimate of the critical value of the stimulus derived from the experimental data by fitting some model performance curve. Examples of this derivation are given in Sect. 4. The variable T may be regarded as a function of the y_i, $i=1,2,\ldots,n$, thus

$$T=g(y_1, y_2, \ldots, y_n). \tag{1}$$

Suppose that the variances of the y_i are σ_i^2, $i=1,2,\ldots,n$. Then provided that the σ_i are all small, the variables y_i are uncorrelated (a condition that may be relaxed at the cost of greater computational complexity), and the function g is well behaved (Lindley 1965), the variance $\operatorname{var}(T)$ of T is given approximately by

$$\operatorname{var}(T)=\sum_{i=1}^{n} (\partial g/\partial y_i)^2 \sigma_i^2, \tag{2}$$

where the partial derivatives $\partial g/\partial y_i$ are evaluated at $y_1, y_2, \ldots, y_n$.

3 Implementation

In the example applications described below, a nonlinear optimization technique modified from the *simplex* method (Nelder and Mead 1965) was used to fit the model performance curve to the simulated data to obtain, in turn, the estimate T in (1). The simplex method is robust and computationally compact, and it tends to produce a good approximate answer quite efficiently (Dixon 1972), requiring mainly additions, subtractions, and simple logical orderings, with few multiplications and no divisions. Goodness of fit was measured by the likelihood function.

The partial derivatives $\partial g/\partial y_i$ in (2) were each estimated by finite-difference approximations. The goodness of the estimates were monitored by comparing values obtained by taking forward and backward differences. The variances σ_i^2, $i=1,2,\ldots,n$, were, in the case of normally distributed y_i, assumed to be given. In those cases where prior estimates of the σ_i^2 are not available, they may be estimated from repeat measurements of the y_i at each stimulus level x_i, $i=1,2,\ldots,n$, or, provided that the σ_i^2 are all equal (usually true for increment-threshold data after the traditional log transformation) and the performance model is correct, they may be estimated by the residual sum of squares divided by the residual degrees of freedom $n-p$, where p is the number of parameters of the model. The adequacy of the model may be tested in the usual way by comparing the two estimates of the σ_i^2. In the case of binomially distributed y_i, $i=1,2,\ldots,n$, the σ_i^2 were determined from the y_i and the number of trials at each level.

4 Monte-Carlo Simulations

Computations were carried out in FORTRAN on a laboratory minicomputer (CAI Alpha LSI-2 with floating-point precision of 7 significant decimal digits) and on two mainframe computers (CDC 7600 and Cyber 176 with floating-point precisions of 15 significant decimal digits). Where appropriate, calculations were done in double-precision arithmetic. Results presented here were obtained on the mainframe computers; results for the minicomputer were similar, but rather less precise.

Two classes of psychophysical performance were simulated: the one comprising psychometric functions characterized by an underlying sigmoidal dependence of performance on stimulus level, with binomially distributed performance values, and the other comprising increment-threshold functions characterized by an underlying power-law dependence of performance on (field) stimulus level, with normally distributed performance values (after logarithmic transformation). An important property of any variance estimate is its bias, that is, the difference between its mean over successive determinations and its "true" value. The "true" value was estimated in each case by Monte-Carlo simulations of the psychophysical performance over 1000 runs. For ease of comparison of variance estimates with parameter values of the models, square roots of variance estimates (standard deviations, SDs) were tabulated. *Percentage bias* was defined as the difference between the root mean variance estimated by the combination-of-observations method and the true value, expressed as a percentage of the true value.

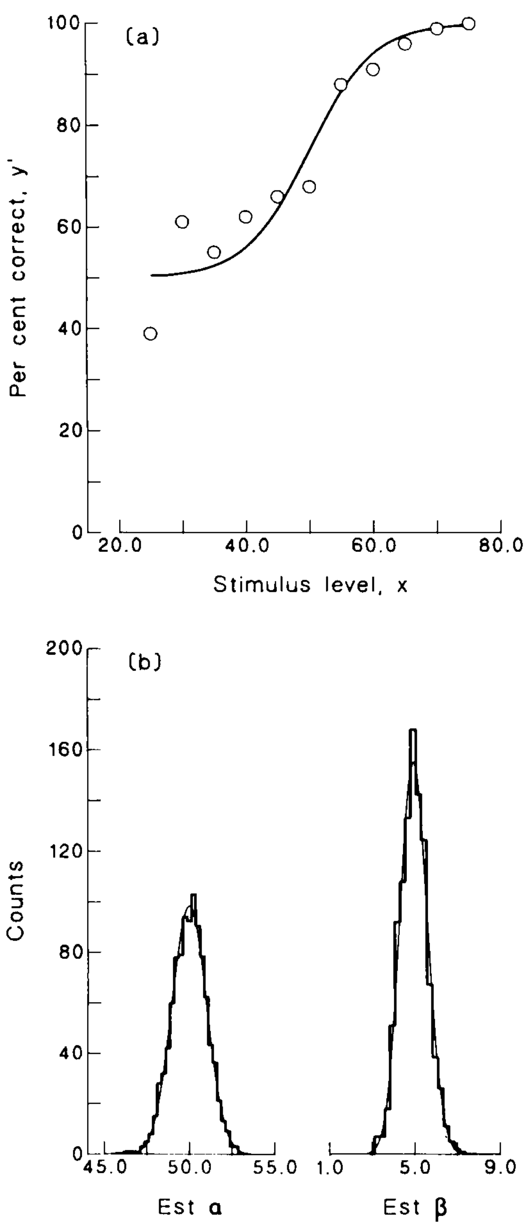

Fig. 1a and b. Sigmoidal psychometric function. **a** Illustration of simulated data set based on Eq. (3) with $\alpha = 50.0$, $\beta = 5.0$, $\gamma = 2$, and number of trials $N = 100$ at each level x. **b** Histograms of estimated values of α and β obtained over 1000 simulations

Sigmoidal Psychometric Function. A family of logistic functions was used to model psychometric-function data, thus

$$y_i = 1/\gamma + (1 - 1/\gamma)(1 + \exp(-(x_i - \alpha)/\beta))^{-1}, \quad i = 1, 2, \ldots, n, \tag{3}$$

where the parameters α and β determined the "midpoint" and "spread" of the distribution, and the constant γ, when finite, corresponded to the number of alternatives in the MAFC task giving rise to the data. (The midpoint of the function is equal to the stimulus level corresponding to the 75% performance level in a 2AFC task, and to the 50% performance level when $\gamma \to \infty$, where performance ranges from 0 to 100%.) For given α, β, and γ, the values y_i, $i = 1, 2, \ldots, n$, in (3) were replaced by values y_i', $i = 1, 2, \ldots, n$, each drawn pseudorandomly from a binomial distribution in which the number of trials was N and the probability of success in a single trial was equal to y_i. An illustration of such a data set for $\alpha = 50.0$, $\beta = 5.0$, $\gamma = 2$, $N = 100$, $n = 11$ is given in Fig. 1a. This set of simulated data was fitted by the function (3) by maximizing the likelihood over α and β, and "new" estimates of the midpoint α and, additionally, the spread β computed. One thousand such sets of data were generated, which, in turn, gave rise to 1000 estimates of α and β. Example histograms of the estimated values of α and β are illustrated in Fig. 1b. Variances of these estimates were calculated and used as the "true" values[1]. The combination-of-observations method was then applied individually to 1000 of the sets of simulated data, resulting in 1000 estimates of the variances of the estimates of α and β.

Results obtained with two different values of the parameters β and γ in the underlying function (3) are summarized in Table 1. In no case was the percentage bias in the estimated SDs of the estimates of α and β estimated by the combination-of-observations method relative to the "true" SD more than 3.1%.

In an exploratory simulation to test the effect of increasing the size of the σ_i [Eq. (2)] in Condition 2 (Table 1), the number of trials N at each level x_i was reduced from 100 to 10. Percentage bias in the estimated SDs of the estimates of α and β then increased to 11% and 14% respectively.

Power-Law Increment-Threshold Function. A family of functions proposed by Sigel and Brousseau (1982), who

1 In all cases, the distribution of estimated values did not differ significantly from normality ($p > 0.1$), and the means of the estimated values did not differ by more than 0.5% of the model values

Table 1. Comparison of standard deviations (SDs) of estimates of (a) the midpoint α, and (b) the spread β of an underlying sigmoidal function [Eq. (3)] obtained for two conditions (1,2), by direct calculation ("True SD"), based on Monte-Carlo studies of 1000 runs each, and by the combination-of-observations method ("Approx. SD"), where the root mean variance was computed over 1000 simulations. Percentage bias in the combination-of-observations method is shown. Parameter values for the underlying sigmoidal function [Eq. (3)] and number of trials N at each stimulus level x are indicated

Condition	True SD	Approx. SD	% Bias
1(a)	1.002	1.000	0.20
1(b)	0.563	0.580	3.03
2(a)	1.003	1.034	3.11
2(b)	0.654	0.636	2.83

Condition 1: $\alpha=50.0$, $\beta=10.0$, $\gamma\to\infty$, $N=100$
Condition 2: $\alpha=50.0$, $\beta=5.0$, $\gamma=2$, $N=100$

generalized the increment-threshold function described by Barlow (1958), was used to model increment-threshold data, thus

$$y_i=\alpha(x_i+\beta)^\gamma, \quad i=1,2,\ldots,n, \tag{4}$$

or, after $\log_{10}$ transformation,

$$\log y_i=\log\alpha+\gamma\log(x_i+\beta), \quad i=1,2,\ldots,n, \tag{5}$$

where the parameters α, β, and γ determined absolute threshold, the "break-point" of the curve, and the slope of the curve. For given α, β, and γ, the values $\log y_i$, $i=1,2,\ldots,n$, in (5) were replaced by values y'_i, $i=1,2,\ldots,n$, each drawn pseudorandomly from a normal distribution with mean $\log y_i$ and standard deviation $\sigma_i=\sigma=0.05$. An illustration of such a data set for $\log\alpha=1.0$, $\log\beta=7.0$, $\gamma=1.0$, $\sigma=0.05$, and $\eta=10$ is given in Fig. 2a, where $x'=\log x$. This set of simulated data was fitted by the function (5) by minimizing the residual sum of squares (equivalent to maximizing the likelihood) and estimates computed of the logarithm of the stimulus levels $x'=\xi_1$ and $x'=\xi_2$, corresponding to elevations in log(increment threshold), y', of respectively 0.3 and 1.0 above log(absolute threshold), $\log\alpha+\gamma\log\beta$. One thousand such sets of data were generated, which, in turn, gave rise to 1000 estimates of ξ_1 and ξ_2. Example histograms of the estimated values of ξ_1 and ξ_2 are illustrated in Fig. 2b. Variances of these estimates were calculated and used as the "true" values[2]. The combination-of-observations method was then applied individually to 1000 of the sets of simulated data, resulting in 1000 estimates of the variances of the estimates of ξ_1 and ξ_2.

Results obtained with two different values of the parameter γ in the underlying function (5) are summarized in Table 2. In no case was the percentage bias in the estimated SDs of the estimates of ξ_1 and ξ_2 estimated by the combination-of-observations method relative to the "true" SD found to be more than 1.8%. Note that the size of the "true" SD is of the order of the SDs of the log(increment threshold) values. Note also the decrease in the size of the "true" SD with increase (0.3 to 1.0) in the criterion elevation in log(increment threshold) and with increase in the exponent γ.

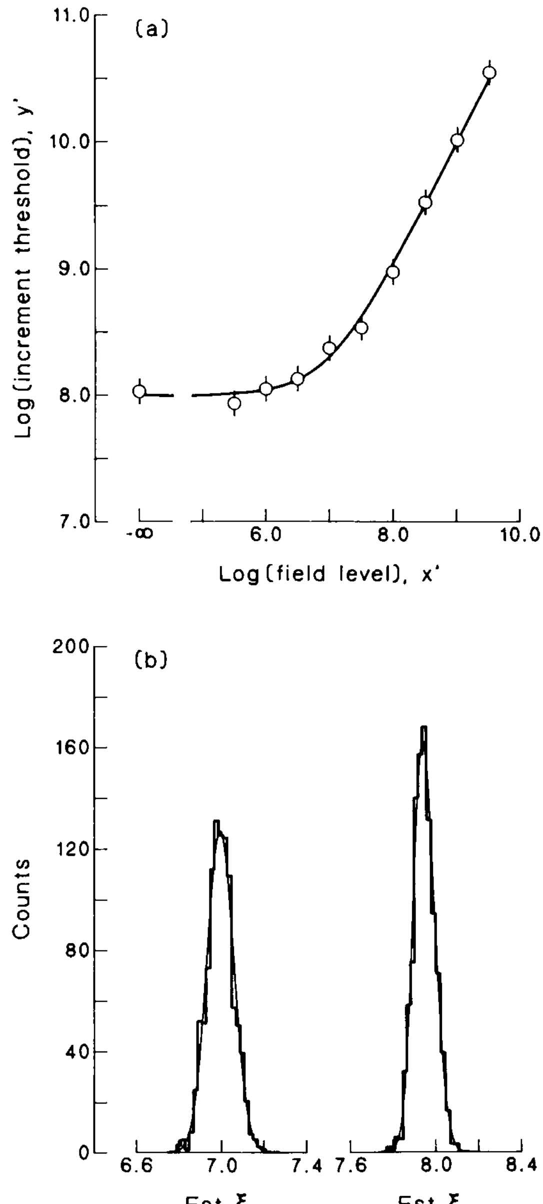

Fig. 2a and b. Power-law increment-threshold function. **a** Illustration of a single simulated data set based on Eqs. (4) and (5) with $\log\alpha=1.0$, $\log\beta=7.0$, $\gamma=1.0$, and standard deviation $\sigma=0.05$ for all y'. **b** Histograms of estimated values of log(field level) ξ_1, and log(field level) ξ_2, obtained over 1000 simulations

Table 2. Comparison of standard deviations (SDs) of estimates of (a) log(field level) ξ_1, and (b) log(field level) ξ_2, for an underlying power-law function obtained for two conditions (1,2), by direct calculation ("True SD"), based on Monte-Carlo studies of 1000 runs each, and by the combination-of-observations method ("Approx. SD"), where the root mean variance was computed over 1000 simulations. Percentage bias in the combination-of-observations method is shown. Parameter values for the underlying power-law function [Eq. (4)] are indicated

Condition	True SD	Approx. SD	% Bias
1(a)	0.0630	0.0635	0.85
1(b)	0.0492	0.0496	0.83
2(a)	0.1151	0.1172	1.78
2(b)	0.0798	0.0802	0.43

Condition 1: $\log\alpha = 1.0$, $\log\beta = 7.0$, $\gamma = 1.0$, $\sigma = 0.05$
Condition 2: $\log\alpha = 1.0$, $\log\beta = 7.0$, $\gamma = 0.5$, $\sigma = 0.05$

In an exploratory simulation to test the effect of increasing the size of the σ_i [Eq. (2)], the value of σ was increased to 0.2 on the log scale. This value is rather larger than would be expected in practice. Increases in percentage bias in the estimated SDs of the estimates of ξ_1 and ξ_2 were of the same order as those obtained for the sigmoidal psychometric function with larger σ_i.

5 Discussion

Estimates of the variance of a critical level of a stimulus variable derived in a psychophysical performance task may be useful in a number of circumstances: (1) where replications of the experiment are impossible; (2) where replications of the experiment are possible and the best (minimum-variance) estimate of a critical stimulus level is to be obtained by a weighted mean over the separate replications of the experiment; and (3) where values of the variance have an intrinsic importance, for example when abnormal values signify the presence of abnormal sensory function (compare Patterson et al. 1980).

Implicit in the present approach has been the assumption that sufficient data were available to fit a model performance curve. To make such a fit, it is not necessary to determine performance at regular intervals over the whole stimulus range. Indeed some techniques for estimating a critical stimulus level have deliberately combined selective sampling of the stimulus range with fitting a performance curve. In the case of the psychometric function, the simple adaptive procedure PEST (Parameter Estimation by Sequential Testing; Taylor and Creelman 1967) has been used to generate the testing levels and a psychometric function fitted by maximum likelihood (Hall 1981). This hybrid adaptive procedure makes fuller use of the data obtained by PEST; it allows the number of trials to be fixed; and it is relatively insensitive to inappropriate estimates of the initial stimulus level and changes in stimulus level (Hall 1981). In the case of the increment-threshold function, a procedure has been followed in which very few segments of the range were sampled and a standard template fitted to derive field sensitivities (Stiles 1978, p. 19).

The values of the parameters of the sigmoidal and power-law performance functions (3) and (4) used here were chosen to be reasonably illustrative of values that might be encountered in practice. In all the cases tabulated, the percentage error in the root mean estimated variance was not more than about 3% of the "true" value, and this error would probably be considered acceptably small in many practical situations. But, as with any method depending on Taylor-series expansions, it is possible that with some performance functions and particular critical stimulus levels the combination-of-observations method may fail. The method has, however, been used successfully in the author's laboratories for over one year with increment-threshold data and with psychometric-function data obtained by the hybrid adaptive procedure as well as by the method of constant stimuli.

The combination-of-observations method appears to have general applicability, and need not be restricted to the estimation of variances of estimates of particular stimulus levels or other dependent variables. It might be used to test different stimulus parametrizations to determine those which are more "natural" by their minimum variance (a suggestion due to Dr. R.J. Watt[1]) and, similarly, it might be used to evaluate different performance models against the variance of some test statistic (a suggestion due to Dr. P.T. Smith[1]). Caution should, however, be exercised in the general application of minimum-variance criteria if the accuracy of the model is uncertain, as Draper and Smith (1981, p. 25) have noted.

Acknowledgements. I thank P.W. Jones for helpful discussions, G.F. Pick for providing the simplex routine, and A.M. Herzberg, S.R. Pratt, and A. Tomoszek for critical reading of the manuscript.

References

Barlow HB (1958) Intrinsic noise of cones. In: Visual problems of colour, vol. 2. HMSO, London, pp 615–630

Dixon LCW (1972) Nonlinear optimisation. The English Universities Press, London, pp 82–87

Draper NR, Smith H (1981) Applied regression analysis, 2nd edn. Wiley, New York
Hall JL (1981) Hybrid adaptive procedure for estimation of psychometric functions. J Acoust Soc Am 69:1763–1769
Lindley DV (1965) Introduction to probability and statistics from a Bayesian viewpoint, Part 1: Probability. Cambridge University Press, Cambridge, pp 134–136
Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313
Patterson VH, Foster DH, Heron JR (1980) Variability of visual threshold in multiple sclerosis. Effect of background luminance on frequency of seeing. Brain 103:139–147
Ratkowsky DA (1983) Nonlinear regression modeling. Dekker, New York
Sigel C, Brousseau L (1982) Pi-4: adaptation of more than one class of cone. J Opt Soc Am 72:237–246
Stiles WS (1978) Mechanisms of colour vision. Academic Press, London
Taylor MM, Creelman CD (1967) PEST: Efficient estimates on probability functions. J Acoust Soc Am 41:782–787

Received: July 16, 1985

Dr. D. H. Foster
Department of Communication
and Neuroscience
University of Keele
Keele
Staffordshire ST5 5BG
England

Note added in proof. Dr. W. F. Bischof has pointed out the relevance of a different approach, that of the bootstrap [Efron B (1982) The jackknife, the bootstrap and other resampling plans. CBMS-NSF Regional Conference Series in Applied Mathematics, No. 38; Society for Industrial and Applied Mathematics, Philadelphia, PA] in which the variance of a statistic is computed by repeated resampling of the data.

Erratum

Estimating the Variance of a Critical Stimulus Level from Sensory Performance Data

David H. Foster

Biol. Cybern. 53, 189–194 (1986)

Two unfortunate printing errors were made in the above article:

on page 192, the footnote reference "2" on line 6 from the bottom of column 1 should have been "1",

on page 193, the footnote reference "1" on lines 13 and 16 from the bottom of column 2 should have been "*".