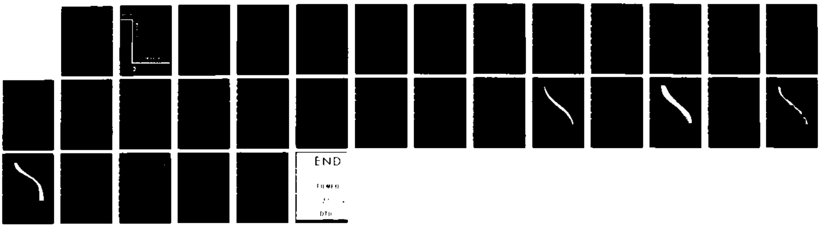
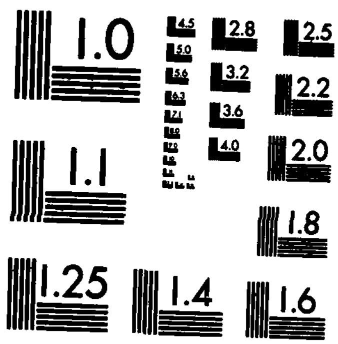
AD-A126 230 CONFIDENCE ENVELOPES FOR MONOTONIC FUNCTIONS: PRINCIPLES DERIVATIONS AND..(U) MCFANN GRAY AND ASSOCIATES INC SAN ANTONIO TX D THISSEN ET AL. MAR 83 1/1

UNCLASSIFIED AFHRL-TP-82-37 F41689-81-C-0012 F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFHRL-TP-82-37

10

AIR FORCE

HUMAN RESOURCES

ADA126230

CONFIDENCE ENVELOPES FOR MONOTONIC FUNCTIONS: PRINCIPLES, DERIVATIONS, AND EXAMPLES

By

David Thissen

Department of Psychology
University of Kansas
Lawrence, Kansas 66045

Howard Wainer

T-254
Educational Testing Service
Princeton, New Jersey 08541

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235

March 1983

Interim Report for Period July 1981 – December 1981

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

This paper has been reviewed and is approved for publication.

JANOS B. KOPLYAY
Contract Monitor

NANCY GUINN, Technical Director
Manpower and Personnel Division

J. P. AMOR, Lt Col, USAF
Chief, Manpower and Personnel Division

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFHRL-TP-82-37	2. GOVT ACCESSION NO. AD-A126 230	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* CONFIDENCE ENVELOPES FOR MONOTONIC FUNCTIONS: PRINCIPLES, DERIVATIONS, AND EXAMPLES		5. TYPE OF REPORT & PERIOD COVERED Interim July ~~1981~~ – December 1981 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR *(s)* David Thissen Howard Wainer		8. CONTRACT OR GRANT NUMBER *(s)* F41689-81-C-0012
9. PERFORMING ORGANIZATION NAME AND ADDRESS McFann-Gray & Associates, Inc. 5825 Callaghan Road, Suite 225 San Antonio, Texas 78228		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2313T137
11. CONTROLLING OFFICE NAME AND ADDRESS HQ Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE March 1983 13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* Manpower and Personnel Division Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235		15. SECURITY CLASS *(of this report)* Unclassified 15.a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of this abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

curves
envelopes
item response theory
latent trait theory
monotonic functions
standard errors
statistical theory

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The bounds of plausible behavior for a multiparameter monotonic function can be expressed as a stochastic confidence envelope around the function. This paper presents the derivation of a general method for constructing such bounds. The bounds described are first specialized to the linear case, yielding results identical to known results found by other means. The method in the more general nonlinear case is then illustrated on three logistic item response models.

DD Form 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 Jan 73

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

AFHRL Technical Paper 82-37 March 1983

CONFIDENCE ENVELOPES FOR MONOTONIC FUNCTIONS: PRINCIPLES, DERIVATIONS, AND EXAMPLES

By

David Thissen

University of Kansas

Howard Wainer

Educational Testing Service

Reviewed by

William E. Alley
Chief, Force Utilization Branch
Manpower and Personnel Division

Submitted for publication by

Janos B. Koplyay
Chief, Manpower and Force Management Systems Branch
Manpower and Personnel Division

This publication is primarily a working paper.
It is published solely to document work performed.

Preface

This research was completed under the Manpower and Force Management thrust and the Force Acquisition and Distribution subthrust. This is part of a continuing effort to improve assessment of personnel qualifications.

The authors wish to express their appreciation to Dr. Benjamin Fairbank, Jr., McFann-Gray & Associates, Inc., and to Dr. Janos Koplyay and Dr. Malcolm Ree of the Manpower and Personnel Division of the Air Force Human Resources Laboratory.

Table of Contents

List of Figures

Confidence Envelopes for Monotonic Functions: Principles, Derivations, and Examples.

I. Curve Envelopes

In seminal work published in 1929, Working and Hotelling developed the formulae for the extrapolation of the concept of the confidence interval to regression lines, to provide a "confidence envelope" enclosing the population line with probability $1 - \alpha$. Such an envelope for a regression function $F(x; \underline{\theta})$ in which $\underline{\theta}$ is a parameter vector has also been called a confidence band (by Miller, 1966). It is defined by functions $L(x; \hat{\underline{\theta}})$ and $U(x; \hat{\underline{\theta}})$ (illustrated in Figure 1) satisfying the relationship

$$P\{U(x;\hat{\underline{\theta}}) \geq F(x;\hat{\underline{\theta}}) \geq L(x;\hat{\underline{\theta}}),\ \forall\ x\} = 1 - \alpha$$

The envelope problem for linear models has been treated extensively by Working and Hotelling (1929), Roy (1957), Miller (1966), and others; but all use the intimate interrelationships of linear models with multinormal error to develop computational formulae for $U(x; \hat{\underline{\theta}})$ and $L(x; \hat{\underline{\theta}})$. This paper describes an algorithm for computing confidence envelopes satisfying the above equation for any function $F(x; \hat{\underline{\theta}})$ which is monotonic in its parameters. This technique will be applied to linear regression, to obtain the classical result; then illustrations will be provided using nonlinear regression.

It is assumed that there are parameter estimates $\hat{\underline{\theta}}$ for which the sampling distribution is multivariate normal with mean $\hat{\underline{\theta}}$ and covariance matrix Σ. Only multi-parameter curves will be considered; for one-parameter curves, which are conditionally monotonic (in the sense defined below) the problem is quite simple. In the multi-parameter situation, a central $(1 - \alpha)$-confidence region for the parameters, called the $(1 - \alpha)$-Highest Density Region (HDR) by Novick and Jackson (1977), is easy to describe: it is defined by the ellipsoid containing $100(1 - \alpha)\%$ of the posterior density, where all of the points within the ellipsoid have higher density than all of the points outside the ellipsoid. The two-dimensional case is illustrated in Figure 2.

The HDR for the general multivariate normal density is defined most easily in terms of that for the "standard" multinormal distribution: $\underline{z} \sim N(\underline{O}, I)$. The HDR for the

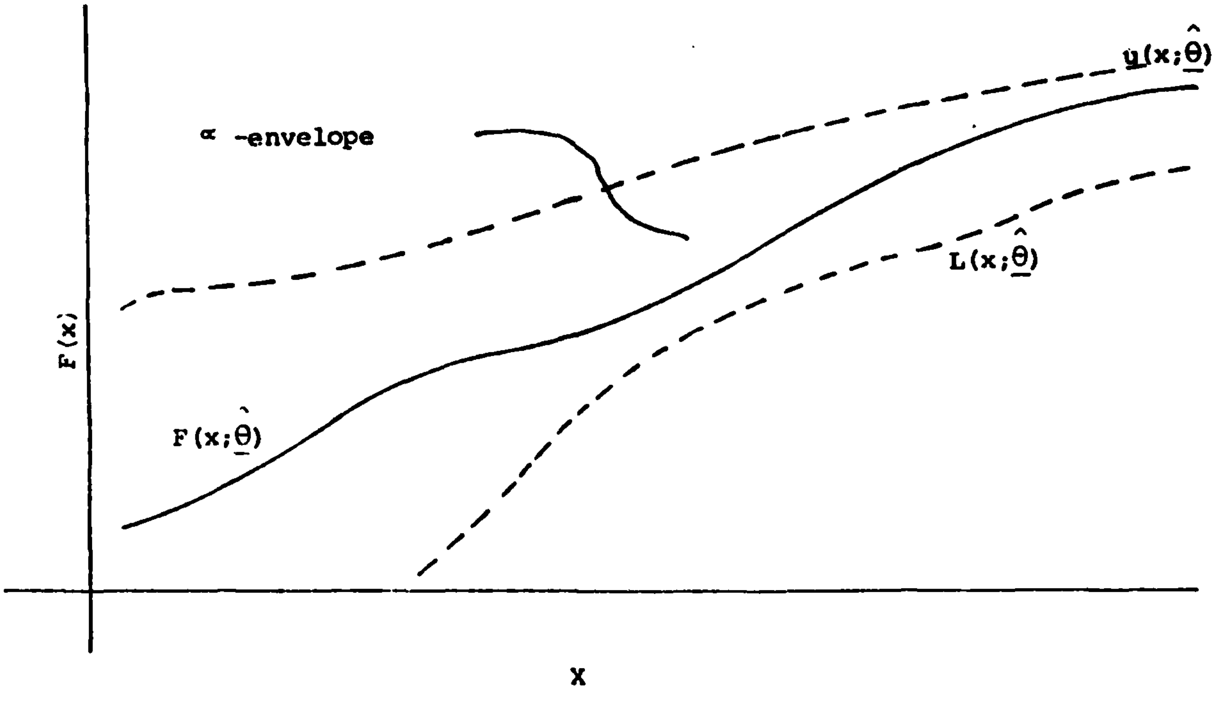


Figure 1. A monotonic curve with a hypothetical confidence envelope.

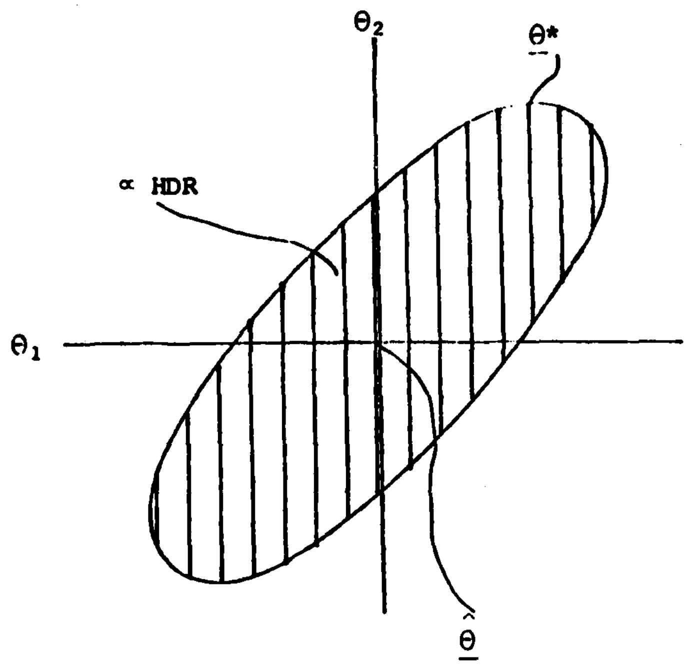


Figure 2. Ellipse showing highest density region.

k-dimensional standard multinormal is defined by the hypersphere (or circle, for the two-dimensional illustration in Figure 3) with radius equal to $\sqrt{\chi^2_{(1-\alpha)}(k)}$, where $\chi^2_{(1-\alpha)}(k)$ is the upper 100 (1 - α) % critical value for the χ^2 distribution with k degrees of freedom. Since any multinormal variables may be orthonormalized by the transform $\underline{z} = C^{-1}(\underline{\theta} - \hat{\underline{\theta}})$, general multinormal $\underline{\theta}^*$ may be defined as

$$\underline{\theta}^* = C\underline{z}^* + \hat{\underline{\theta}} \tag{1}$$

(in which C is the Cholesky factor of Σ such that $\Sigma = CC'$). Thus, the HDR for the general multinormal in Figure 2 is an ellipsoid defined by equation (1) and the condition

$$\underline{z}^{*\prime}\underline{z}^* = \chi^2_{(1-\alpha)}(k).$$

Since the radius of this hypersphere (or circle in the two dimensions of Figure 3) is fixed, the points $\underline{z}^*$ may be defined conveniently in hyperspherical (or circular) coordinates as functions exclusively of the angle $\underline{\phi}$.

Each point in the parameter space defines a fitted line; the lines defined by the parameter values within the HDR in Figure 2 are the lines which must lie within the (1 - α)-envelope in Figure 1. Therefore, the problem of computing the boundaries of the (1 - α)-envelope becomes that of computing, for each value of x, the value of $L(x; \hat{\underline{\theta}})$ which is the minimum $F(x; \underline{\theta})$ for any $\underline{\theta}$ in the HDR, and $U(x; \hat{\underline{\theta}})$ which is the maximum $F(x; \underline{\theta})$ in the HDR.

If the function $F(x; \underline{\theta})$ is "conditionally monotonic" in its parameters, where "conditional monotonicity" means that for parameter θ_i,

$$F(x;\theta_i \mid \theta_{j\neq i}) > F(x;\theta_i + \delta \mid \theta_{j\neq i})$$

or

$$F(x;\theta_i \mid \theta_{j\neq i}) < F(x;\theta_i' + \delta \mid \theta_{j\neq i})$$

for all positive δ, then the minimum and maximum values of $F(x; \underline{\theta})$ for the $\underline{\theta}$ within the HDR are always <u>on the boundary</u> of the HDR. Thus, computing $L(x; \hat{\underline{\theta}})$ and

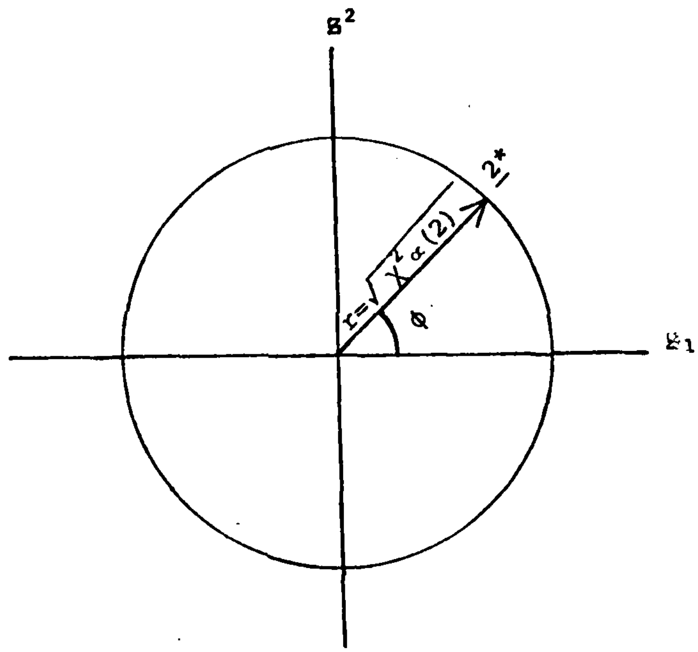


Figure 3. Circle illustrating the meaning of phi, for the two dimensional case.

$U(x; \underline{\hat{\theta}})$ is a matter of locating the points $\underline{\theta}^*$ on the ellipsoid in Figure 2 which minimize and maximize $F(x; \underline{\theta})$ for values of x. The points $\underline{\theta}^*$ are all specified by angles in $\underline{z}$-space and the transformation in equation (1).

Thus, to compute values of the function $L(x; \hat{\underline{\theta}})$ for a series of values of x, and to make plots as in Figure 1, one minimizes for each value x_0:

$$F(x_0; \underline{\theta}) = F\{x_0; G[H(\underline{\phi}; \chi^2_{(1-\alpha)}(k)); \hat{\underline{\theta}}, \Sigma]\}$$

$$= F(x_0; \underline{\phi}), \text{ for short,}$$

in which

$$\underline{\theta} = G(\underline{z}; \hat{\underline{\theta}}, \Sigma) = C\underline{z} + \hat{\underline{\theta}}$$

and

$$\underline{z} = H(\underline{\phi}; \chi^2_{(1-\alpha)}(k))$$

represent the conversion of circular or hyperspherical coordinates $[\underline{\phi}', r = \sqrt{\chi^2_{(1-\alpha)}(k)}]$ into cartesian coordinates $\underline{z}$. Minimization is in $\underline{\phi}$-space, with dimensionality k - 1. Consequently, for a two-dimensional problem, one solves for the scalar angle ϕ in Figure 3 which gives the values of $\underline{\theta}^*$ in Figure 2 which minimize $F(x_0; \underline{\theta})$ in Figure 1; that is the value of $L(x_0; \hat{\underline{\theta}})$. After repeating this process frequently enough to plot $L(x_0; \hat{\underline{\theta}})$, one <u>maximizes</u> $F(x_0; \underline{\theta})$ over $\underline{\theta}^*$ to obtain $U(x_0; \hat{\underline{\theta}})$.

The Linear Case

The derivation of the envelope for the linear case is more complex than the usual arguments (as in Miller, 1966); however, generalization of this argument to the nonlinear case is straightforward. One specifies that

$$F(x; \underline{\theta}) = \theta_1 x + \theta_2 \qquad (2)$$

is a linear function in which $\underline{\theta} = \theta_1, \theta_2$ are the slope and intercept parameters; for convenience, the mean of the (fixed) x-values is zero and their variance is 1. One assumes that $\hat{\underline{\theta}}$ represents the maximum likelihood estimate (MLE) for $\underline{\theta}$ from a large sample of n observations. The sampling distribution for $\hat{\underline{\theta}}$ is approximately N ($\hat{\underline{\theta}}$, Σ) in which

$$\Sigma = (\sigma^2/n)I$$

is the negative inverse expected value of the matrix of second derivatives of the loglikelihood, and σ^2 is the error variance. Thus,

$$C = (\sigma/\sqrt{n})I$$

L (x; $\hat{\underline{\theta}}$) is obtained by minimizing (2) for each value of x over values of $\underline{\theta}$ so that

$$\underline{\theta} = C\underline{z}^* + \hat{\underline{\theta}} \qquad (3)$$

in which the $\underline{z}^*$ satisfy the condition

$$\underline{z}^{*\prime}\underline{z}^* = \chi^2_{(1-\alpha)}(2). \qquad (4)$$

The value of $h = \sqrt{\chi^2_{(.95)}(2)}$ is $\sqrt{6} \approx 2.45$, to give a 95%-envelope. The condition specified by (4) is met by $\underline{z}^*$ obtained as

$$\underline{z}^{*\prime} = [\,h\cos\phi \quad h\sin\phi\,] \qquad (5)$$

The lower bound of the 95%-envelope L (x; $\hat{\underline{\theta}}$) is given by the minimum of the linear function (2) with the constraint that $\underline{\theta}$ satisfies (3) in which $\underline{z}^*$ satisfies (5). The minimization is in ϕ-space.

The required minimum is obtained where the derivative of (2) with respect to ϕ is zero; that is

$$\partial F/\partial\phi = x\partial\theta_1/\partial\phi + \partial\theta_2/\partial\phi = 0 \qquad (6)$$

Since

$$\partial \underline{\theta} / \partial \phi = C \partial \underline{z} / \partial \phi$$

and

$$\partial \underline{z}' / \partial \phi = h[-\sin \phi \quad \cos \phi],$$

it follows that

$$\partial \underline{\theta}' / \partial \phi = (h\sigma / \sqrt{n}) \; [-\sin \phi \quad \cos \phi]$$

and

$$\partial F / \partial \phi = (h\sigma / \sqrt{n}) [-x \sin \phi + \cos \phi] \qquad (7)$$

The partial derivative (6) is zero when (7) is zero; (7) is zero when

$$1/x = \sin \phi / \cos \phi = \tan \phi .$$

So

$$\phi = \tan^{-1} (1/x) \qquad (8)$$

Solutions of ϕ between 0° and 180° give maximum; solutions between 180° and 360° give minimum. Such a closed-form result is simpler than those to be obtained in subsequent sections on nonlinear models. To plot L (x; $\hat{\underline{\theta}}$) compute ϕ using (8), $\underline{z}^*$ as in (5), $\hat{\underline{\theta}}$ as in (3) and F from (2) and plot the corresponding hyperbola.

Or, to obtain the customary form given by Working and Hotelling (1929) and Miller (1966), note that for ϕ satisfying (8)

$$x \cos \phi + \sin \phi = \pm [1 + x^2]^{1/2} \qquad (9)$$

plus for maximum and negative for minimum.

(Geometric proof of (9) is obvious using any triangle satisfying (8).) Using the identity in (9) and substituting (3) in (2) (keeping the terms separate), one obtains

$$L(x;\underline{\hat{\theta}}) = \hat{\theta}_1 x + \hat{\theta}_2 - h\sigma/\sqrt{n}\ [1+x^2]^{1/2} \qquad (10)$$

where the first two terms are from $\underline{\hat{\theta}}$ in (3) and the last term is the product of $C\underline{z}^*$ and x. This is the usual form for the lower bound, as in Miller (1966, p. 111), with h (the root of the large-sample χ^2) replacing $\sqrt{2F}$ and x standardized. Thus, for the linear case, the trigonometry implied by ϕ is computationally optional; it serves only if it is desirable to know the parameter values associated with some point on the lower bound. The trigonometry will be required in the nonlinear case, however. The upper bound simply reverses the negative sign in (10).

A Simple Nonlinear Example

In the case of logistic regression of a binary response on some fixed x, $F(x;\ \underline{\theta})$ is the logistic function

$$F(x;\ \underline{\theta}) = 1\ /\ [1 + \exp(-u)]\ ,$$

$$u = \theta_1 (x - \theta_2)$$

in which θ_1 is the logit-slope and θ_2 is the location parameter, estimated so that the sampling distribution of $\underline{\hat{\theta}}$ is $N(\underline{\theta},\ \Sigma^2 = CC')$. $F(x;\ \underline{\theta})$ is a minimum wherever u is minimized, and $F(x;\ \underline{\theta})$ is maximized wherever u is a maximum. To minimize (maximize) F, minimize (maximize) u. Only minimization will be considered to locate $L(x;\underline{\hat{\theta}})$; to maximize for $U(x;\underline{\hat{\theta}})$, reverse the signs. Minimize u with respect to ϕ.

The minimum of u for a particular x_o is obtained where the derivative

$$\partial u/\partial\phi = (x-\theta_2)\partial\theta_1/\partial\phi - \theta_1\partial\theta_2/\partial\phi = 0. \qquad (11)$$

Since

$$\underline{\theta} = C\underline{z} + \underline{\hat{\theta}}$$

$$\partial\underline{\theta}/\partial\phi = C\partial\underline{z}/\partial\phi,$$

$$\underline{z}' = [h\cos\phi \quad h\sin\phi], \text{ and}$$

$$\partial\underline{z}'/\partial\phi = [-h\sin\phi \quad h\cos\phi]$$

in which $h = \sqrt{\chi^2_{(1-\alpha)}(2)}$, the derivative in (11) represents a straightforward computation. Unlike the linear example, there is no closed-form solution in this case, but with the second derivative

$$\partial^2 u/\partial\phi^2 = (x - \theta_2)\partial^2\theta_1/\partial\phi^2 - 2\partial\theta_1/\partial\phi \; \partial\theta_2/\partial\phi - \theta_1\partial^2\theta_2/\partial\phi^2$$

the function u is easily minimized (over ϕ) by Newton-Raphson for a set of values of x, and the results may be plotted as $L(x; \hat{\underline{\theta}})$.

A More Complex Example

In a more complex sort of logistic regression sometimes used in mental test theory, a lower asymptote θ_3 is also estimated,

$$F(x; \underline{\theta}) = \theta_3 + (1-\theta_3)\{1/[1+\exp(-u)]\}$$

$$= \theta_3 + (1-\theta_3)P, \text{ and}$$

$$u = \theta_1(x - \theta_2)$$

and, for notational convenience,

$$Q = 1 - P$$

where P indicates the probability of a correct response (Lord, 1980).

Again, assume that $\underline{\theta}$ has been estimated so that the sampling density is $N(\hat{\underline{\theta}}, \Sigma = CC')$. Since there are three parameters in this model, minimization and maximization of $F(x_o; \underline{\theta})$ must be with respect to a two-dimensional angle $\underline{\phi}$, and all of the derivatives are more complex. Nevertheless, $L(x_o; \hat{\underline{\theta}})$ is obtained at the point at which $\partial F/\partial\underline{\phi} = \underline{0}$, where

$$\partial F/\partial \underline{\phi} = (1 - \theta_3)\ PQ\ \partial u/\partial \underline{\phi} + Q\partial\theta_3/\partial \underline{\phi}$$

and

$$\partial u/\partial \underline{\phi} = (x - \theta_2)\partial\theta_1/\partial \underline{\phi} - \theta_1\partial\theta_2/\partial \underline{\phi}\ .$$

As before,

$$\underline{\theta} = C\underline{z} + \hat{\underline{\theta}},$$

so

$$[\partial\underline{\theta}/\partial\underline{\phi}] = C\ [\partial\underline{z}/\partial\underline{\phi}]$$

and since

$$\underline{z}' = [\ h \sin\phi_2 \cos\phi_1 \quad h \sin\phi_2 \sin\phi_1 \quad h \cos\phi_2\ ]$$

it follows that

$$[\partial\underline{z}/\partial\underline{\phi}]' = \begin{bmatrix} -h \sin\phi_2 \sin\phi_1 & h \sin\phi_2 \cos\phi_1 & 0 \\ h \cos\phi_2 \cos\phi_1 & h \cos\phi_2 \sin\phi_1 & -h \sin\phi_2 \end{bmatrix}$$

in all of which $h = \sqrt{\chi^2_{(1-\alpha)}(3)}$

With the matrix of second derivatives (which is entirely too big and complex to reproduce here), the function $F\ (x_0; \underline{\phi})$ can be minimized over $\underline{\phi}$ to give $L\ (x_0; \hat{\underline{\theta}})$, and maximized to give $U\ (x_0; \hat{\underline{\theta}})$. The minimization problem is fairly ill-conditioned for some values x_0, so both a conditioned Newton-Raphson algorithm and multiple starting values must be used to locate the minimum.

Plotting $L\ (x; \hat{\underline{\theta}})$ can be accomplished using the following system. Start at the leftmost x_0 to be plotted, and successively (iteratively) compute $L\ (x; \hat{\underline{\theta}})$ as the minimum $F(x; \underline{\phi})$ for each value of x_0 to be plotted, always using the value of $\underline{\phi}$ for the previous (left-adjacent) x_0 as the starting value. When the rightmost x_0 is reached, reverse the procedure and re-compute each point, starting from $\underline{\phi}$ for the right-adjacent x_0. In the flattest regions, the results for the two passes sometimes differ. Since the process is one of minimization, to get $L(x_0; \hat{\underline{\theta}})$, retain the lower value. The values for $U(x_0; \hat{\underline{\theta}})$ are computed with the same algorithm, with the signs reversed.

II. N-line Plots

While plots of the $(1-\alpha)$-envelope define a $(1-\alpha)$-confidence interval for a fitted curve, they do not graphically display the density of curves within the envelope. For that purpose, there are N-line plots, which are plots which (approximately) fill the $(1-\alpha)$-envelope with curves, in proportions roughly corresponding to their likelihood. N-line plots consist of plots of N lines drawn from parameters randomly drawn from the sampling distribution for the parameters. N-line plots are easiest to describe by a series of examples.

A One-Parameter Example: 25-Line Plots

An extremely simple variety of logistic regression is sometimes used in mental test theory; it makes use of a model of the form

$$F(x; \theta) = 1 / [1 + \exp(-u)],$$

$$u = x - \theta$$

in which θ is a location parameter, and the only estimated parameter in the model. Since this one-parameter function is monotonic in its parameter, the computation of the $(1-\alpha)$-envelope for the fitted curve is trivial, as mentioned above: if $\hat{\theta}$ is the estimate of θ, with a standard error of σ, then the upper boundary for the 95%-envelope is $U(x; \hat{\theta}) = F(x; \hat{\theta} - 1.96\,\sigma)$; and the analogous lower boundary for the envelope is $L(x; \hat{\theta}) = F(x; \hat{\theta} + 1.96\,\sigma)$. A typical 95% envelope for such a function is plotted in Figure 4, with the modal (fitted) curve in the center of the envelope. The envelope plot describes, with its boundaries, the central 95% confidence interval for the fitted curve.

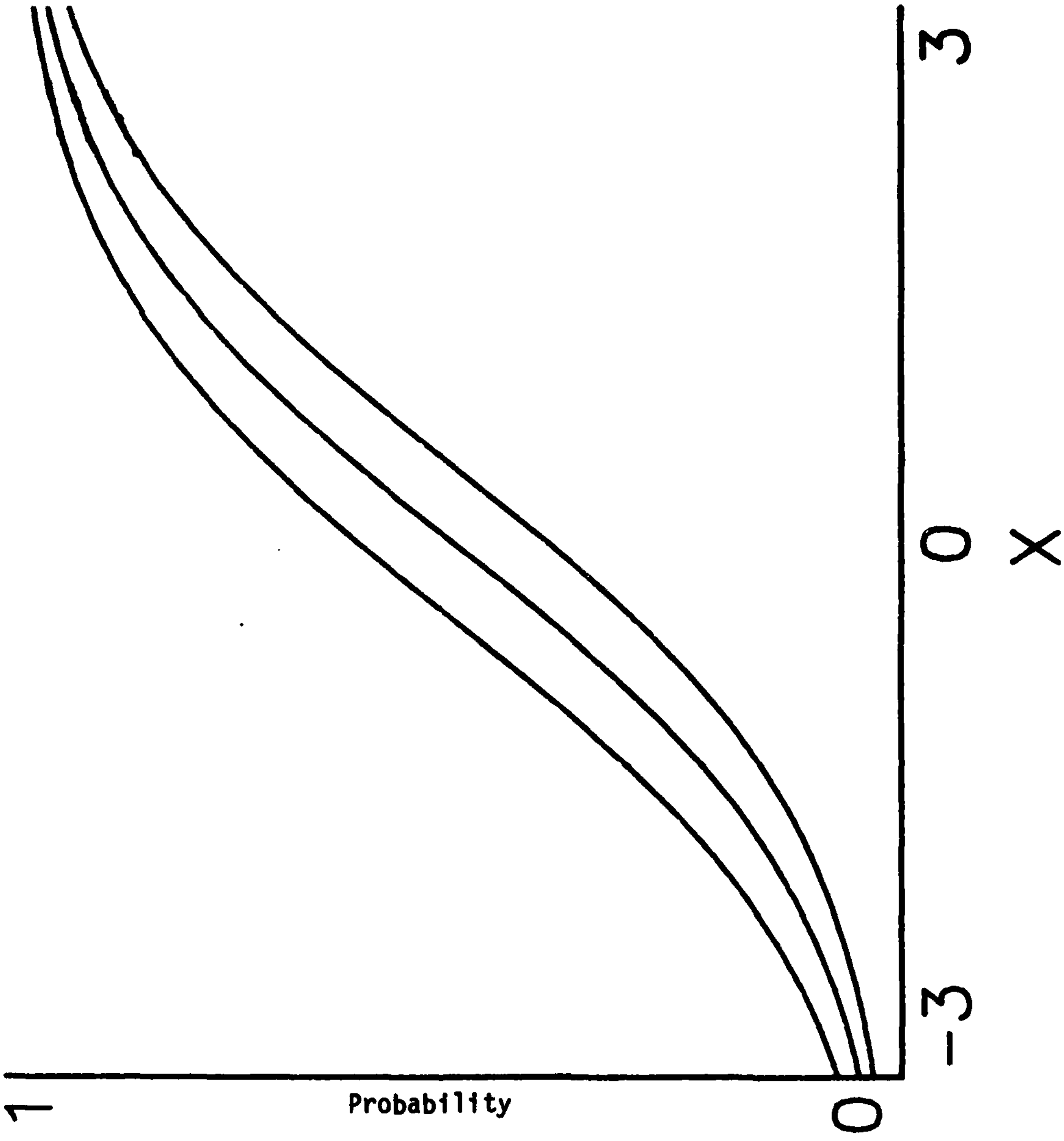


Figure 4. One-parameter model; a=1.0, b=0.0, c=0.0, n=100, 95% confidence envelope. In this and in subsequent figures the letters a, b, and c indicate the parameters of a logistic item response model (Lord, 1980), and so correspond to the vector $\underline{\theta}$ discussed in the text. Parameter a is related to the steepness of the curve at its steepest point, b indicates the position of the curve along the X axis, and c is the lower asymptote.

But can the density for the fitted curve be "filled in" a bit? Figure 5, following the assumption that the density for θ is $N(\hat{\theta}, \sigma^2)$, shows plots of the curves corresponding to 25 random deviates from a normal density with mean $\hat{\theta}$ and variance $\hat{\sigma}^2$. Note that such randomly sampled curves are dense near the curve corresponding to $\hat{\theta}$ and fill the 95%-envelope in Figure 4. Thus, a 25-line plot is a "filled-out" envelope plot, even though the density of lines is less near the edges.

The envelope described by the upper and lower extreme curves of the 25-line plot is, in expectation, precisely equal to the true 95% envelope. This is because the distance between the upper and lower boundaries of the 95%-envelope plot is determined by plotting $F(x; \hat{\theta} + 1.96\ \sigma)$ and $F(x; \hat{\theta} - 1.96\ \sigma)$ i.e., plotting F for two values of θ, centered on $\hat{\theta}$, which have a range of $2\ (1.96)\ \sigma$. The random deviates which yield the 25-line plot are also centered on $\hat{\theta}$, and the expected value of their range is $2\ (1.96)\ \sigma$. (The expected value of the range of a sample of N = 25 standard normal deviates is $3.93 \doteq 2\ (1.96)$ (Pearson and Hartley, 1956).) Thus, the boundaries of the 95%-envelope and the expected values of the highest and lowest curves of a 25-line plot are in the same place. The 25-line plot gives a clearer graphic description of the envelope for the curve.

A Two-Parameter Example: 85-Line Plots

A two-parameter logistic regression, in which

$$F(x; \underline{\theta}) = 1/\ [1 + \exp(-u)]\ , \text{ and}$$

$$u = \theta_1 (x - \theta_2)$$

provides more interesting N-line plots, but they are more costly to make. The vector $\underline{\theta}$ is assumed estimated with sampling density $N(\hat{\underline{\theta}}, \Sigma)$. If the upper and lower extreme curves of the N-line plot are to equal (in expectation) the boundaries of the 95%-envelope, the appropriate N-line plot should have as its (squared) range on any line through the density (shown in Figure 2), the Mahalanobis distance given by

$$(\underline{\theta} - \hat{\underline{\theta}})'\ \Sigma^{-1}\ (\underline{\theta} - \hat{\underline{\theta}}) = \chi^2_{(1-\alpha)}(2)\ .$$

That is, on any line through the two-dimensional space in Figure 2, the expected range of the projections of the random deviates of the N-line plot should be the same as the

Figure 5. One-parameter model; a=1.0, b=0.0, c=0.0, n=100, a 25-line plot.

distance between the two boundaries of the HDR on that line. Every value of x has as its upper and lower ($1 - \alpha$)-envelope boundaries the parameters at the intersections of the HDR boundary with some line in $\underline{\theta}$-space. So the requirement that the expected value of the range of the projections of the random sample on any such line be equal to the HDR boundary difference gives the desired result: the expected location of the upper and lower N-line curves are along the upper and lower ($1 - \alpha$)-envelope boundaries.

For the two-dimensional case, N (regrettably) is 85. This is determined by noting that the $\underline{\theta}$ -space of Figure 2 is a projection of the $\underline{z}$-space of Figure 3, and the 95%HDR boundary of Figure 3 has radius equal to $\sqrt{\chi^2_{(.95)}(2)} = \sqrt{6.0} = 2.45$. Thus one makes N-line plots from the two-space by sampling standard normal deviates in the orthonormal space (from the density in Figure 3), transforming them to have the $\underline{\theta}$ -density in Figure 2, and plotting the resulting curves. So the expected value of the range of the standard normal z's sampled must be 2 (2.45), to give the N-line plot the required breadth. Pearson and Hartley's (1956) Table 27 indicates that the required sample size is N = 85. Figure 6 shows an illustrative 85-line plot; Figure 7 is the corresponding 95%-envelope plot. Note that the 85-line plot shows clearly how the odd shape of the 95%-envelope arises from the distribution of curves. Flat curves make up $U(x; \hat{\underline{\theta}})$ on the left and $L(x; \hat{\underline{\theta}})$ on the right, whereas steep curves make up $L(x; \hat{\underline{\theta}})$ on the left and $U(x; \hat{\underline{\theta}})$ on the right.

Eighty-five lines can be tedious to plot. Twenty-five line plots are equivalent (in expectation) to only 85%-envelopes, but, as Figure 8 shows, give much of the detail available in 85-line plots.

Lots-of-Line Plots

By an argument similar to that of the preceding section, N-line plots for three-parameter models require N = 235 to match (in expectation) 95%-envelope plots. Figure 9 shows a 235-line plot for the three-parameter logistic model described above, and Figure 10 is the 95%-envelope plot. Two hundred thirty-five lines take a plotter a long time. (The width of a 25-line plot is equivalent (in expectation) to a 72%-envelope.)

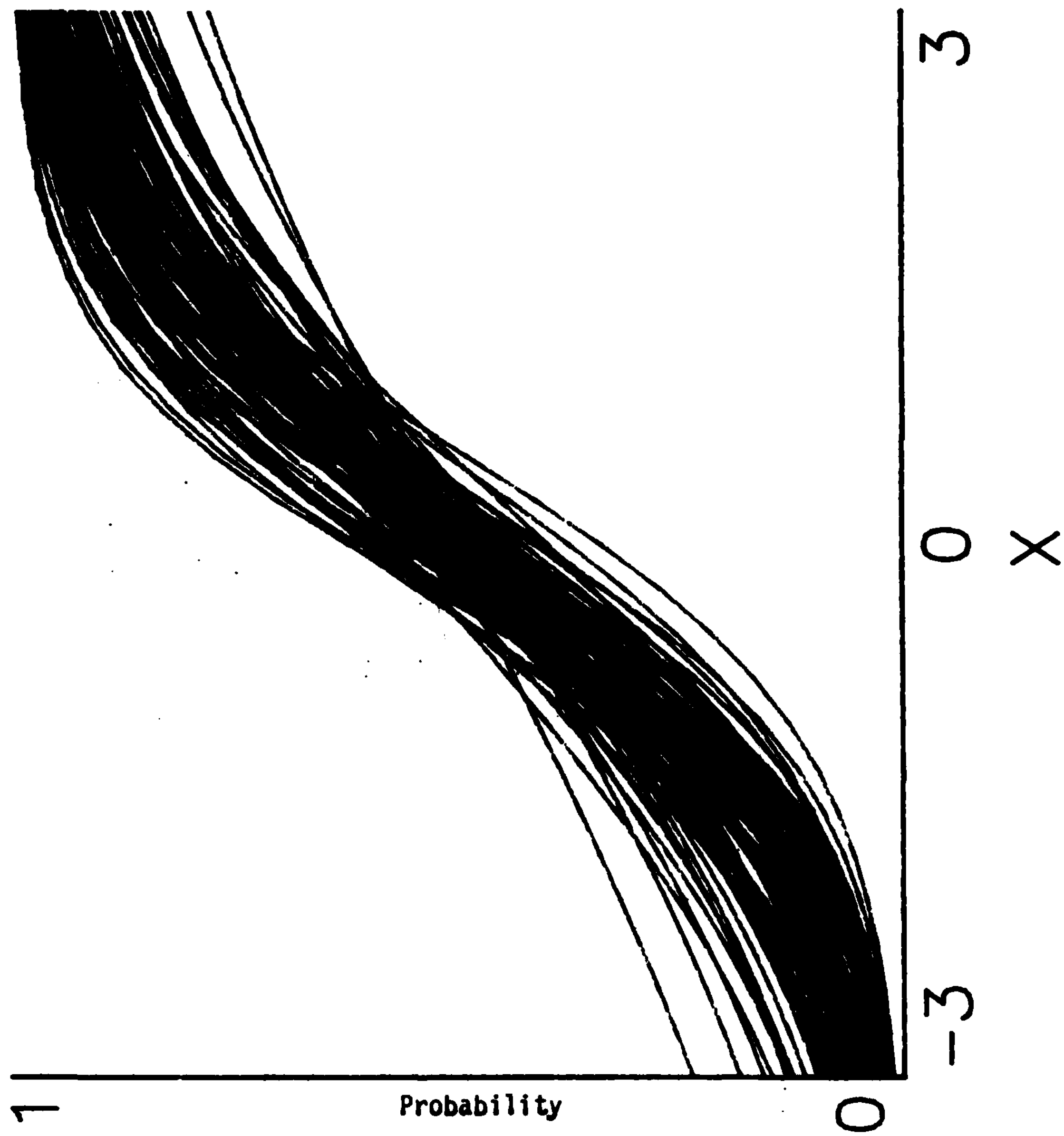


Figure 6. Two-parameter model; a=1.0, b=0.0, c=0.0, n=100, an 85-line plot.

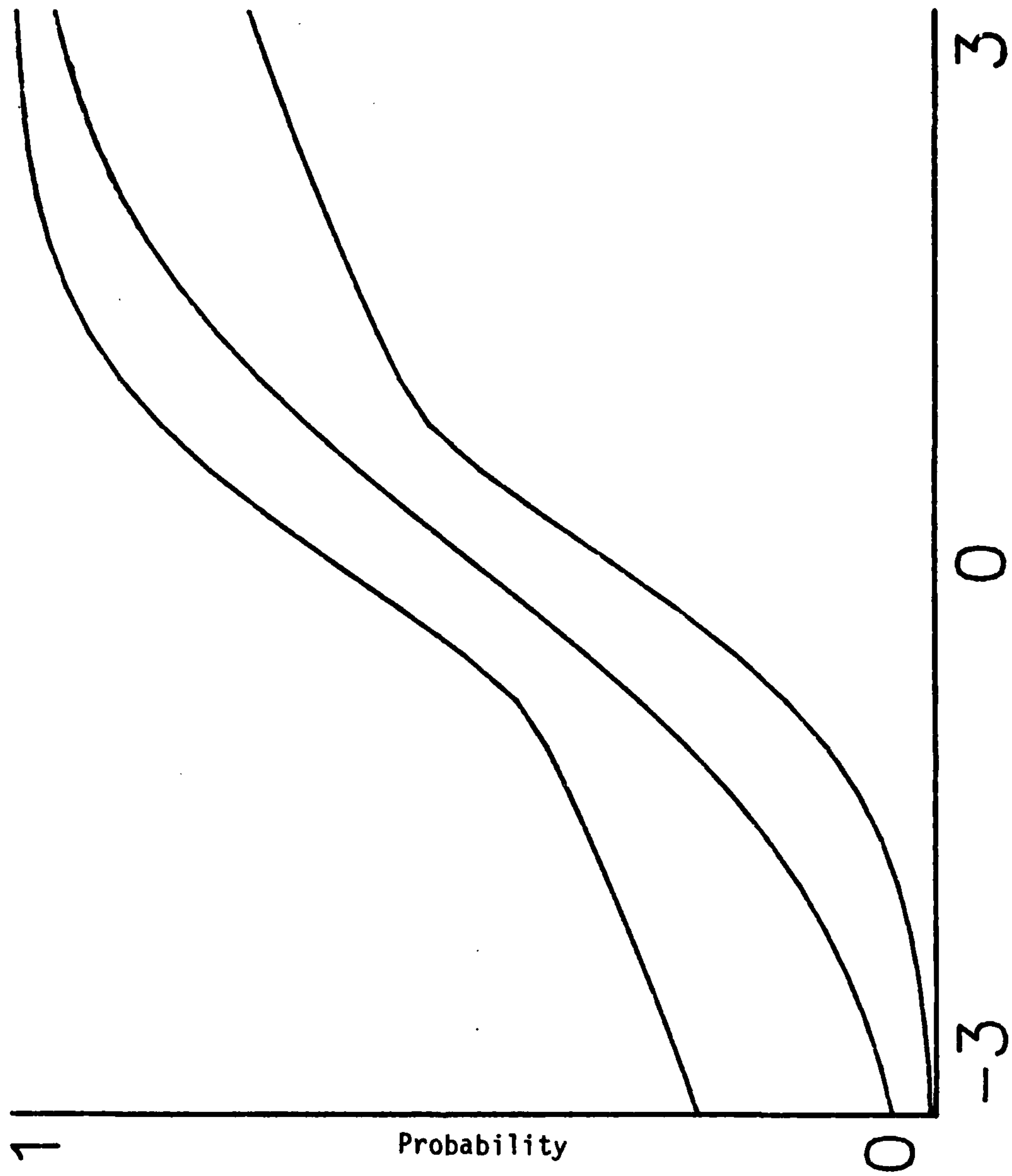


Figure 7. Two-parameter model; a=1.0, b=0.0, c=0.0, n=100, 95% confidence envelope.

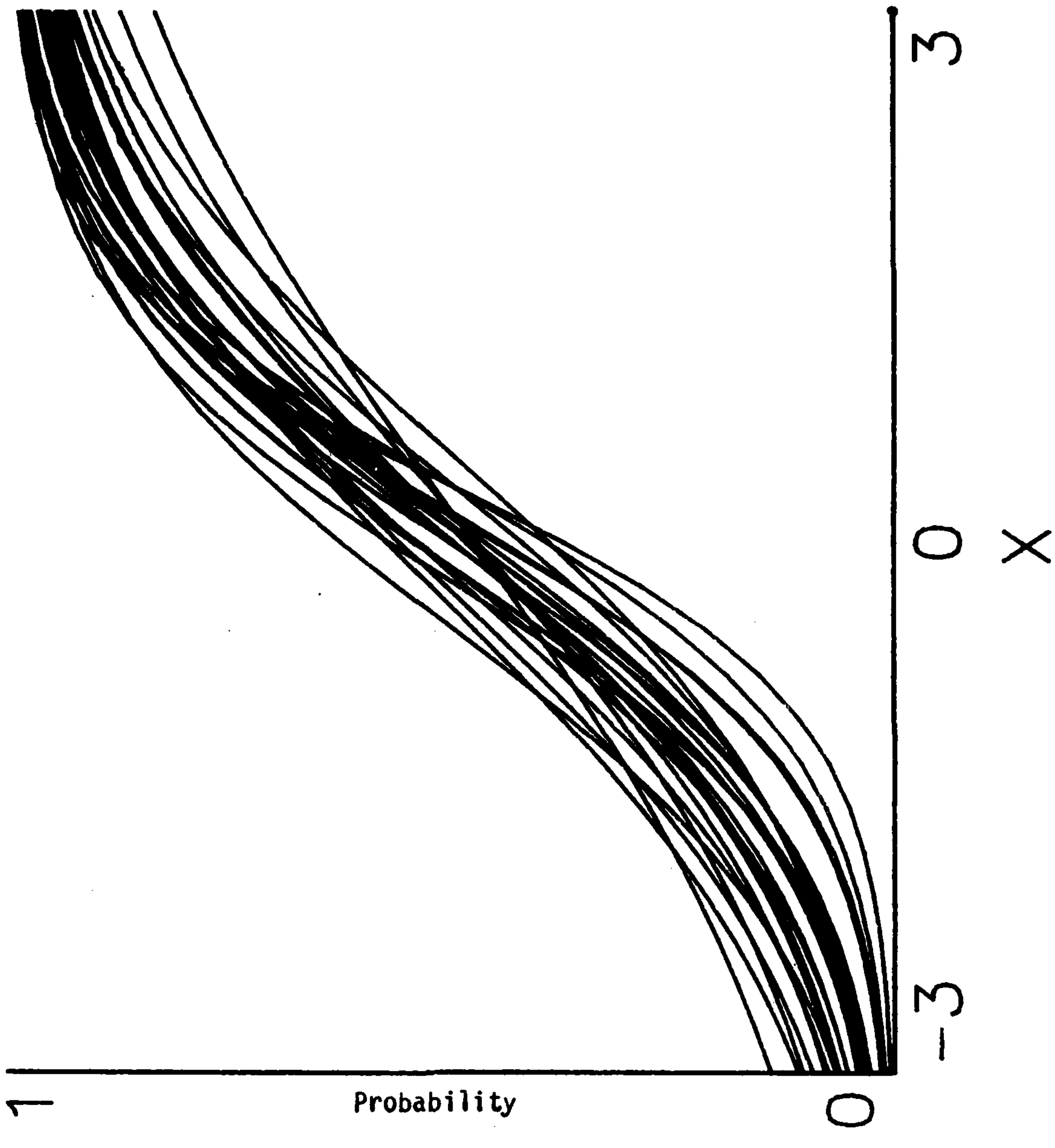


Figure 8. Two-parameter model; a=1.0, b=0.0, c=0.0, n=100, a 25-line plot.

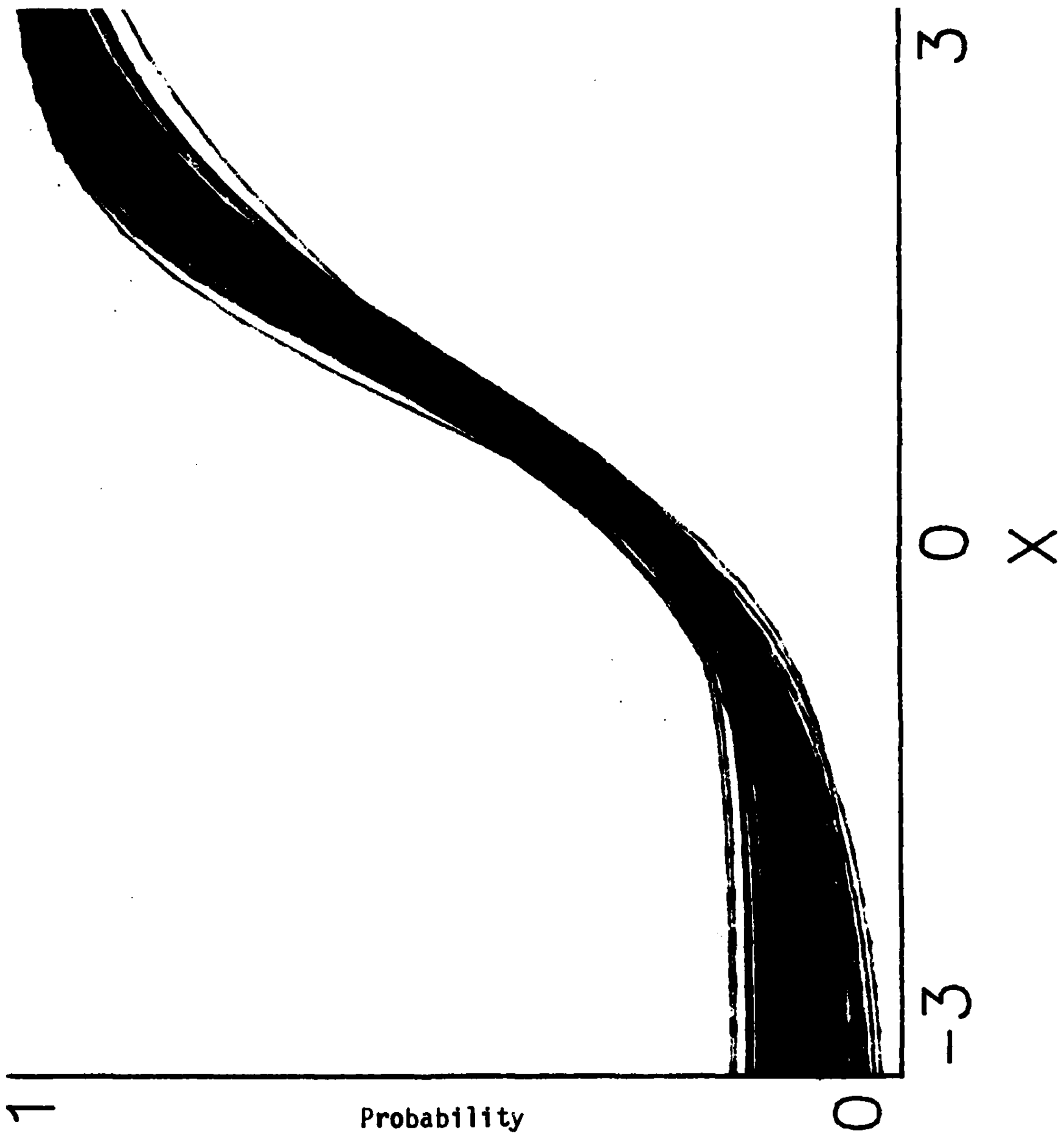


Figure 9. Three-parameter model; a=1.5, b=1.0, c=0.1, n=1,000, 235-line plot.

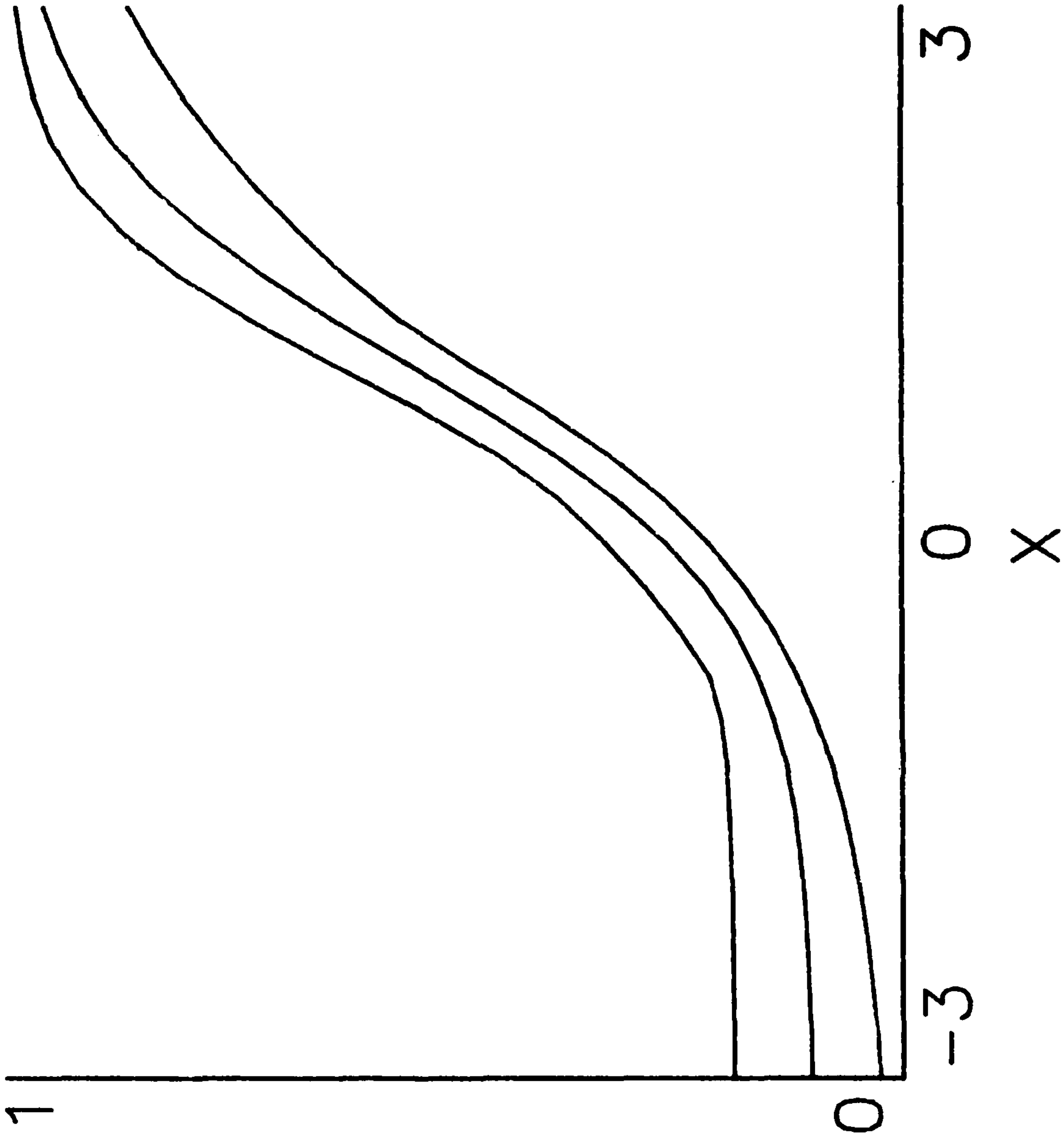


Figure 10. Three-parameter model; a=1.5, b=1.0, c=0.1, n=1,000, 95% envelope.

N-line plots become cumbersome as dimensionality increases. Some alternatives to 95%-envelope-equivalent N-line plots may be suggested. The simplest is to combine the (1 - α)-envelope itself (computed as in Section I) with a small number of randomly sampled curves to give some idea about the structure. Another possibility is to abandon the dependence of N-line plots on extreme-value statistics. One could plot a small number of lines, and then, for each value of x, estimate the standard deviation of $F(x; \underline{\theta})$. That estimate could be multiplied by the factor appropriate to the dimensionality of the problem (2.45 for two, 2.8 for three, and so on) and added to (and subtracted from) the modal curve to give projections of the (1 - α)-envelope. The standard deviation could be estimated robustly, to depart indeed from the dependence on extreme-value statistics in the methods above.

On the other hand, one could (less robustly, but more quickly) depend more on extreme value statistics, using half-N-line plots. To make a half-N-line plot, plot the number of randomly sampled curves needed to have the expected range equal to half the desired range (0.98 for one dimension, 1.23 for two, 1.4 for three, and so on; N's = 4, 6, 8. . .). The data analyst then must visually multiply the width of the plot by two to see the desired result, which, with such small N's, would be unreliable. Still, it would be quick.

Conclusion

The methodology derived and illustrated here provides an operational connection between the stochastic variation in the parameter space and the observed variation in the function space. This explicit connection provides a mechanism for the establishment of confidence bounds around non-linear functions. Such bounds were not available with previous techniques.

Earlier work (Thissen & Wainer, 1982) showed that the standard errors of item parameters for commonly used models were so large as to render use of the models impractical in some cases even with moderately large sample sizes. The techniques herein developed can be used to determine the consequences of these large standard errors on the estimation of the overall response function. Such measures are relevant for many practical applications of item response theory (such as computer adaptive testing).

Recent investigations (Lord and Stocking, personal communication) have shown that the response surface being explored for the three-parameter logistic function boundaries is poorly behaved, with local extrema being encountered more often than seldom. Consequently the bounds based on the methods of this paper may be too narrow. This limitation is not too serious, for one typically finds that the error bounds are often too broad for the delicate decisions required. Thus if the bounds are perhaps narrower than they should be and are still too broad, this limitation is unlikely to be critical. We have not found these same local extrema problems with the one or two parameter models.

References

Lord, F. Applications of item response theory to practical testing problems. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Miller, Jr., R. G. Simultaneous statistical inference. New York: McGraw-Hill, 1966.

Novick, M. R., & Jackson, J. E. Statistical methods for educational and psychological research. New York: McGraw-Hill, 1977.

Pearson, E. S., & Hartley, H. O. Tables of the incomplete beta-function. London: Cambridge University Press, 1956.

Roy, S. N. Some aspects of multivariate analysis. New York: John Wiley & Sons, 1957.

Thissen, D. & Wainer, H. Some standard errors in item response theory. Psychometrika, 1982, in press.

Working, H. & Hotelling H. Application of the theory of error to the interpretation of trends, Journal of the American Statistical Association, 1929, Suppl. 24: 73-85.

4 -

DT