DOCUMENT RESUME

ED 053 394 CG 006 480

| | |
|---|---|
| AUTHOR | Offir, Joseph |
| TITLE | Some Mathematical Models of Individual Differences in Learning and Performance. Psychology and Education Series. |
| INSTITUTION | Stanford Univ., Calif. Inst. for Mathematical Studies in Social Science. |
| SPONS AGENCY | National Science Foundation, Washington, D.C. |
| REPORT NO | TR-176 |
| PUB DATE | 28 Jun 71 |
| NOTE | 109p. |
| EDRS PRICE | EDRS Price MF-$0.65 HC-$6.58 |
| DESCRIPTORS | *Computer Assisted Instruction, Elementary School Mathematics, *Individual Differences, Instructional Improvement, *Mathematical Models, *Paired Associate Learning, *Performance |

ABSTRACT

With the advance of computers, extensive work has been undertaken in the field of programmed instruction. Much effort has been invested to devise schemes of optimal instruction with respect to suitable criteria. Yet, what is needed is a theory which prescribes how learning can be improved, i.e., a theory of instruction. The present study is motivated by the absence of adequate formalization of individual and item differences. The results of the study demonstrate unequivocally that the One-Element Model (OEM) with the heterogeneity provision is still a fairly accurate model. More significant is the observation that individual differences have a first order effect on the predictive power of simple stochastic models. In addition, the hypothesis was confirmed that the heterogeneity assumption increases the predictive power of simple learning models and has a sizable effect on their learning properties. Finally, in the context of computer-assisted instruction in elementary mathematics, results demonstrated that asymptotic performance data can be accounted for successfully by probablistic automation models with few parameters. (Author/TA)

N-X
CG

# SOME MATHEMATICAL MODELS OF INDIVIDUAL DIFFERENCES IN LEARNING AND PERFORMANCE

BY

JOSEPH OFFIR

TECHNICAL REPORT NO. 176
June 28, 1971

PSYCHOLOGY & EDUCATION SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

## TECHNICAL REPORTS

### PSYCHOLOGY SERIES

### INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

(Place of publication shown in parentheses; if published title is different from title of Technical Report, this is also shown in parentheses.)

(For reports no. 1 - 44, see Technical Report no. 125.)

50 R. C. Atkinson and R. C. Calfee. Mathematical learning theory. January 2, 1963. (In B. B. Wolman (Ed.), Scientific Psychology. New York: Basic Books, Inc., 1965. Pp. 254-275)

51 P. Suppes, E. Crothers, and R. Weir. Application of mathematical learning theory and linguistic analysis to vowel phoneme matching in Russian words. December 28, 1962.

52 R. C. Atkinson, R. Calfee, G. Sommer, W. Jeffrey and R. Shoemaker. A test of three models for stimulus compounding with children. January 29, 1963. (J. exp. Psychol., 1964, 67, 52-58)

53 E. Crothers. General Markov models for learning with inter-trial forgetting. April 8, 1963.

54 J. L. Myers and R. C. Atkinson. Choice behavior and reward structure. May 24, 1963. (Journal math. Psychol., 1964, 1, 170-203)

55 R. E. Robinson. A set-theoretical approach to empirical meaningfulness of measurement statements. June 10, 1963.

56 E. Crothers, R. Weir and P. Palmer. The role of transcription in the learning of the orthographic representations of Russian sounds. June 17, 1963.

57 P. Suppes. Problems of optimization in learning a list of simple items. July 22, 1963. (In Maynard W. Shelly, II and Glenn L. Bryan (Eds.), Human Judgments and Optimality. New York: Wiley. 1964. Pp. 116-126)

58 R. C. Atkinson and E. J. Crothers. Theoretical note: all-or-none learning and intertrial forgetting. July 24, 1963.

59 R. C. Calfee. Long-term behavior of rats under probabilistic reinforcement schedules. October 1, 1963.

60 R. C. Atkinson and E. J. Crothers. Tests of acquisition and retention, axioms for paired-associate learning. October 25, 1963. (A comparison of paired-associate learning models having different acquisition and retention axioms, J. math. Psychol., 1964, 1, 285-315)

61 W. J. McGill and J. Gibbon. The general-gamma distribution and reaction times. November 20, 1963. (J. math. Psychol., 1965, 2, 1-18)

62 M. F. Norman. Incremental learning on random trials. December 9, 1963. (J. math. Psychol., 1964, 1, 336-351)

63 P. Suppes. The development of mathematical concepts in children. February 25, 1964. (On the behavioral foundations of mathematical concepts. Monographs of the Society for Research in Child Development, 1965, 30, 60-96)

64 P. Suppes. Mathematical concept formation in children. April 10, 1964. (Amer. Psychologist, 1966, 21, 139-150)

65 R. C. Calfee, R. C. Atkinson, and T. Shelton, Jr. Mathematical models for verbal learning. August 21, 1964. (In N. Wiener and J. P. Schoda (Eds.), Cybernetics of the Nervous System: Progress in Brain Research. Amsterdam, The Netherlands: Elsevier Publishing Co., 1965. Pp. 333-349)

66 L. Keller, M. Cole, C. J. Burke, and W. K. Estes. Paired associate learning with differential rewards. August 20, 1964. (Reward and information values of trial outcomes in paired associate learning. (Psychol. Monogr., 1965, 79, 1-21)

67 M. F. Norman. A probabilistic model for free-responding. December 14, 1964.

68 W. K. Estes and H. A. Taylor. Visual detection in relation to display size and redundancy of critical elements. January 25, 1965, Revised 7-1-65. (Perception and Psychophysics, 1966, 1, 9-16)

69 P. Suppes and J. Donio. Foundations of stimulus-sampling theory for continuous-time processes. February 9, 1965. (J. math. Psychol., 1967, 4, 202-225)

70 R. C. Atkinson and R. A. Kinchla. A learning model for forced-choice detection experiments. February 10, 1965. (Br. J. math stat. Psychol., 1965, 18, 184-206)

71 E. J. Crothers. Presentation orders for items from different categories. March 10, 1965.

72 P. Suppes, G. Groen, and M. Schlag-Rey. Some models for response latency in paired-associates learning. May 5, 1965. (J. math. Psychol., 1966, 3, 99-128)

73 M. V. Levine. The generalization function in the probability learning experiment. June 3, 1965.

74 D. Hansen and T. S. Rodgers. An exploration of psycholinguistic units in initial reading. July 6, 1965.

75 B. C. Arnold. A correlated urn-scheme for a continuum of responses. July 20, 1965.

76 C. Izawa and W. K. Estes. Reinforcement-test sequences in paired-associate learning. August 1, 1965. (Psychol. Reports, 1966, 18, 879-919)

77 S. L. Blehart. Pattern discrimination learning with Rhesus monkeys. September 1, 1965. (Psychol. Reports, 1966, 19, 311-324)

78 J. L. Phillips and R. C. Atkinson. The effects of display size on short-term memory. August 31, 1965.

79 R. C. Atkinson and R. M. Shiffrin. Mathematical models for memory and learning. September 20, 1965.

80 P. Suppes. The psychological foundations of mathematics. October 25, 1965. (Colloques Internationaux du Centre National de la Recherche Scientifique. Editions du Centre National de la Recherche Scientifique. Paris: 1967. Pp. 213-242)

81 P. Suppes. Computer-assisted instruction in the schools: potentialities, problems, prospects. October 29, 1965.

82 R. A. Kinchla, J. Townsend, J. Yellott, Jr., and R. C. Atkinson. Influence of correlated visual cues on auditory signal detection. November 2, 1965. (Perception and Psychophysics, 1966, 1, 67-73)

83 P. Suppes, M. Jerman, and G. Groen. Arithmetic drills and review on a computer-based teletype. November 5, 1965. (Arithmetic Teacher, April 1966, 303-309.

84 P. Suppes and L. Hyman. Concept learning with non-verbal geometrical stimuli. November 15, 1968.

85 P. Holland. A variation on the minimum chi-square test. (J. math. Psychol., 1967, 3, 377-413).

86 P. Suppes. Accelerated program in elementary-school mathematics -- the second year. November 22, 1965. (Psychology in the Schools, 1966, 3, 294-307)

87 P. Lorenzen and F. Binford. Logic as a dialogical game. November 29, 1965.

88 L. Keller, W. J. Thomson, J. R. Tweedy, and R. C. Atkinson. The effects of reinforcement interval on the acquisition of paired-associate responses. December 10, 1965. (J. exp. Psychol., 1967, 73, 268-277)

89 J. I. Yellott, Jr. Some effects on noncontingent success in human probability learning. December 15, 1965.

90 P. Suppes and G. Groen. Some counting models for first-grade performance data on simple addition facts. January 14, 1966. (In J. M. Scandura (Ed.), Research in Mathematics Education. Washington, D. C.: NCTM, 1967. Pp. 35-43.

91 P. Suppes. Information processing and choice behavior. January 31, 1966.

92 G. Groen and R. C. Atkinson. Models for optimizing the learning process. February 11, 1966. (Psychol. Bulletin, 1966, 66, 309-320)

93 R. C. Atkinson and D. Hansen. Computer-assisted instruction in initial reading: Stanford project. March 17, 1966. (Reading Research Quarterly, 1966, 2, 5-25)

94 P. Suppes. Probabilistic inference and the concept of total evidence. March 23, 1966. (In J. Hintikka and P. Suppes (Eds.), Aspects of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1966. Pp. 49-65.

95 P. Suppes. The axiomatic method in high-school mathematics. April 12, 1966. (The Role of Axiomatics and Problem Solving in Mathematics. The Conference Board of the Mathematical Sciences, Washington, D. C. Ginn and Co., 1966. Pp. 69-76.

(Continued on inside back cover)

# SOME MATHEMATICAL MODELS OF INDIVIDUAL DIFFERENCES IN LEARNING AND PERFORMANCE

by

Joseph Offir

TECHNICAL REPORT NO. 176

June 28, 1971

PSYCHOLOGY & EDUCATION SERIES



INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ACKNOWLEDGEMENTS

I wish to thank Dr. Patrick Suppes, my dissertation committee chairman, who initially introduced me to the problem of individual differences. I also wish to thank the other members of the committee, Dr. Richard C. Atkinson and Dr. Ingram Olkin, for their helpful suggestions and criticisms.

The data described in Chapter III were gathered as part of the Stanford Program in Computer-Assisted Instruction, supported by National Science Foundation Grants NSFG-18709 and NSF GJ-197.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENT i

I. INTRODUCTION

I.1 Models of Learning and Performance 1

I.1.1 Three Models of the Learning Process 1
I.1.2 An Example of a Performance Model 4

I.2 Identification of the Problematic Situation 7

II. EFFECTS ON LEARNING PROPERTIES OF HAVING CONTINUOUS DISTRIBUTIONS OVER THE LEARNING RATES

II.1 Introduction 12

II.1.1 General Remarks 13
II.1.2 The Evolution of the Method 16

II.2 Effects on Total Error Statistics of Having Independent Beta Distributions Over the Learning Rates 20

II.3 Effects on Response 4-Tuple of Having Independent Beta Distributions Over the Learning Rates 29

II.3.1 Probabilities of Response Sequences Over Trials 2 to 5 29
II.3.2 Data Analysis 36

II.4 Discussion and Conclusions 50

II.4.1 General Remarks: Mathematical Methods for the Analysis and Evaluation of Models 50
II.4.2 The Important Features of the Results 53
II.4.3 Further Research and Conclusions 57

III. PERFORMANCE MODELS FOR SIMPLE ARITHMETIC PROBLEMS

III.1 Introduction 59

III.2 Some Basic Results 60

III.2.1 The Likelihood Function and Maximum Likelihood Estimates 60

III.2.2 The Bivariate Dirichlet Distribution 62

III.2.3 The Distribution of Item Performance Rates with Homogeneous Individuals 63

CONTENTS (Continued)


| | | Page |
|---|---|---|
| III.3 | Total Error Statistics | 66 |
| III.4 | Data Analysis | 68 |
| | III.4.1 Description of the Data | 68 |
| | III.4.2 The Distribution of Item Performance Rates with Homogeneous Individuals | 74 |
| III.5 | Discussion and Conclusions | 75 |
| | III.5.1 The Conditional Models | 75 |
| | III.5.2 The Unconditional Models | 76 |
| REFERENCES | | 79 |
| APPENDIX | | 82 |

CHAPTER I

# INTRODUCTION

## I.1. MODELS OF LEARNING AND PERFORMANCE

The idea that an educational experience is comprised of two stages, learning and performance, is relatively new and very little has been written about it. Conceptually, the notion is quite simple. The learning stage takes place as long as the subject continues to update his knowledge or as long as there is a positive probability that the proportion of his correct responses will increase. This stage lasts until the subject reaches a threshold, or a steady state, beyond which improvements may be only random fluctuations. The performance stage takes place from this point onward.

In Section I.1.1 we briefly review three models of the learning process -- usually associated with Paired Associate Learning (PAL). A simple performance model (Automaton) for two rows addition problems is presented in I.1.2.

### I.1.1. Three Models of the Learning Process

The Single-Operator Linear Model (LM)

The model is represented by two equivalent equations (Atkinson, Bower and Crothers, 1965):

The probability, $p_n$, of a correct response on trial $n$ increases according to the equation

$$p_{n+1} = \alpha P_n + (1-\alpha) \tag{1.1}$$

where $\alpha$ denotes the learning rate. The initial probability $p_1$ is assumed to be $1/r$, i.e., one over the number of response alternatives. Equivalently, the probability, $q_n$, of an incorrect response on trial $n$ decreases according to the relation

$$q_{n+1} = \alpha q_n \tag{1.2}$$

The One-Element Model (OEM)

The OEM and its properties (Ibid) are derived from the following assumptions. Each item starts in the unconditioned state U. Subsequently the item may move with probability c to state L, where it is conditioned, or stays unconditioned with probability $1-c$. Until the item is conditioned there is a constant probability g that the subject will respond correctly by guessing. Once the item becomes conditioned, i.e. enters state L, the probability of a correct response is unity. The transition matrix and the response probability vector are usually presented in the following way:

$$\begin{array}{c} \\ L \\ U \end{array} \begin{array}{c} \begin{array}{cc} L & U \end{array} \\ \begin{bmatrix} 1 & 0 \\ c & 1-c \end{bmatrix} \end{array} \quad \begin{array}{c} Pr(\text{correct}) \\ \begin{bmatrix} 1 \\ g \end{bmatrix} \end{array} \tag{1.3}$$

$g = 1/r$ as in the LM case. Since both models have the same mean learning curve

$$MLC = \frac{1-g}{c} = \frac{q_1}{1-\alpha} \tag{1.4}$$

it is convenient to interchange $\alpha$ with $1-c$ and $q_1$ with $1-g$. If $g$ is fixed as above the two models have only one free parameter.

The Long-Short Model (LS-3)

This model was motivated, among other things, by PAL studies which indicated that before conditioning immediate recall of S-R pairs by a subject was nearly perfect while the proportion of correct responses decreased with the time before the next trial (Peterson, et al., 1962). The model is described in Atkinson and Crothers; part of the description is quoted in the next lines.

> "Encoding for a given stimulus item occurs at most on one trial; the probability that encoding occurs on trial $n$ given that it has not occurred on previous trial is $c$. If an item is presented that has already been encoded (either on the present trial or on an earlier trial), then with probability $a$ it goes into state L and with probability $1-a$ it goes into state S. Thus, after each presentation, an encoded item is in either state L or S, and if the item were to be presented again immediately the subject would make the correct response with probability 1. However, other events intervene from one presentation of an item to its next presentation, and during this period we assume there is a probability $f$ that an item in state S will move back to state F. We assume the value of $f$ depends upon the number and type of intervening items; also, $f$ depends upon the exposure time of the given item, for this affects the repetition

rate and hence the slope of the forgetting function (Peterson, et al., 1962).

"Given the above assumptions, it can be shown that moves among the four states are described by the following transition matrix and response probability vector:

$$\begin{array}{c} \\ L \\ S \\ F \\ U \end{array} \begin{array}{c} \begin{array}{cccc} L & S & F & U \end{array} \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ a & (1-a)(1-f) & (1-a)f & 0 \\ a & (1-a)(1-f) & (1-a)f & 0 \\ ca & c(1-a)(1-f) & c(1-a)f & 1-c \end{array}\right] \end{array} \begin{array}{c} Pr(\text{correct}) \\ \left[\begin{array}{c} 1 \\ 1 \\ g \\ g \end{array}\right] \end{array} \qquad (1.5)$$

where $g = 1/r$ ; throughout the paper we shall use $g$ to denote the guessing probability."

A special case of the LS-3 model is reduced to a two parameter version by letting $c = 1$ in Eq. (1.5). This special case will be designated as the LS-2 model.

### I.1.2. An Example of a Performance Model

As an example (Suppes, 1968), consider a stochastic automaton for column addition of two integers:

The automaton is the structure

$$\langle A, I, 0, M, Q, s_0 \rangle$$

$A = \{0,1\}$ - the set of internal states

$I = \{(m,n) : 0 \leq m,\ n \leq 9\}$ - the input alphabet

$0 = \{0,1,\ldots,9\}$ - the output alphabet

$$M(k,(m,n)) = \begin{cases} 0 & \text{if } m+n+k \leq 9 \\ 1 & m+n+k > 9 \quad \text{for } k = 0,1 \end{cases}$$

M is the transition function from $A \times I$ into $A$ .

$Q(k,(m,n)) = (k+m+n) \bmod 10$ – is the transition function from $A \times I$ into $O$ .

$s_0 = 0$ – is the initial state.

Consider first the three parameter situation $0 \leq a,b,c \leq 1$ where

$$P(M(k,(m,n)) = 0 \mid k+m+n \leq 9) = 1-a \equiv \bar{a}$$

$$P(M(k,(m,n)) = 1 \mid k+m+n > 9) = 1-b \equiv \bar{b} ,$$

i.e., if there is no "carry" the probability of a correct response is $1-a$ . If there is a carry the probability of such a transition is $1-b$ .

The third parameter is simply the output error $c$

$$P(Q(k,(m,n)) = (k+m+n) \bmod 10) = 1-c \equiv \bar{c} .$$

If $C_i$ and $D_i$ represent carries and digits in problem $i$ respectively, and if we ignore the probability of two errors leading to a correct response, e.g., transition error followed by an output error then

$$P(\text{correct answer to problem } i) = (1-c)^{D_i}(1-b)^{C_i}(1-a)^{D_i - C_i - 1}$$

We can reduce this case to a two parameter situation, $a$ and $b$ , by assuming $c$ the output error to be fixed for all items.

Different statistics may be calculated for different automata models, and provide an immediate analysis of digit by digit response. An example of such statistics is the likelihood of $n$ digit responses derived by Suppes for the automaton described above. Here, for illustration purposes, the distribution of total error is derived:

Let IS denote the internal state;

$$X_i = \begin{cases} 1 & \text{if correct response on digit } i \\ 0 & \text{otherwise} \end{cases}$$

$$P(x_i=1) = \begin{cases} \bar{c} & \text{if a) 1 is ones column digit} \\ \bar{c}\bar{a} & \text{if b) not (a), IS = 0, i.e., no carry} \\ \bar{c}\bar{b} & \text{if c) not (a), IS = 1, i.e., carry} \end{cases}$$

$n_a, n_b, n_c$ - are the number of digits under (a), (b) and (c). Then the probability of A, B and C correct responses under (a), (b) and (c) respectively is given by

$$(4) \binom{n_a}{A}\binom{n_b}{B}\binom{n_c}{C} \bar{c}^{A+B+C} c^{n_a-A} \bar{a}^B \bar{b}^C (1-\bar{c}\bar{a})^{n_b-B} (1-\bar{c}\bar{b})^{n_c-C}$$

Suppes proved in general that given any (connected) finite automaton, there is a stimulus response model that asymptotically becomes isomorphic to it.

## I.2. IDENTIFICATION OF THE PROBLEMATIC SITUATION

With the advance of computers, extensive work has been undertaken in the field of programmed instruction. Much effort has been invested to devise schemes of optimal instruction with respect to suitable criteria (e.g., Smallwood, 1967). Most of these efforts have not yielded much in the way of unequivocal results (Silberman, 1962), a situation which is symptomatic of a deeper problem that exists not only in the field of programmed instruction but in other areas of educational research. What is needed is a theory which prescribes how learning can be improved. A theory of this type has come to be called a theory of instruction (e.g., Hilgard, 1964; Bruner, 1964), as compared with a theory of learning.

Typical questions that a theory of instruction concerns itself with are: how to advance a student through a block of teaching material, when to stop presenting teaching items, what items are to be presented within a given time. Ideally this kind of question can be answered with mathematical rigour in a decision analysis frame of reference. It should be remembered, however, that the criterion for optimization is always determined subjectively beforehand.

In many works (e.g., Groen and Atkinson, 1966) an instructional system is defined as the structure $\langle C,R,H,d,u,g \rangle$ where

C - is the set of concepts to be presented

R - is a set of all possible responses made by the student

H - is a set of histories of the student's performance

d: $H \to C$ is a decision function

u: $C X R X H \to H$ is an updating function

g: is a fixed criterion for optimally given in advance

Historically, mathematical learning theory and optimization attempts have tended to ignore the structure of the stimulus set C . Items of C have been assumed to be independent and not to have a cumulative effect on the learning and to be homogeneous and not of varying degree of difficulty. Some recent attempts have been made to formally model certain prototypal tasks which occur in elementary mathematics (e.g., Suppes, et al., 1968; Offir, 1968).

In considering the response set R , most studies use only dichotomous variables 0,1 to indicate correct or incorrect responses. Many studies proceed to estimate the model's parameters and to test the model's adequacy by averaging (dichotomous responses) over ensemble of subjects in order to explain the learning or the performance that has taken place. By using quantal responses, i.e., 0,1 variables, and the like, one ignores the relationship between the structure of the stimulus set C and the full response structure of R . By so doing, it is impossible, for instance, to distinguish between relevant and irrelevant responses. There is no substantial reference to this issue in the literature.

A more serious inadequacy is the overlooking of individual response protocols and their sequential dependencies. There are few attempts to tackle this problem (see Sternberg, 1963 for references). In general, however, in most applications of learning models it is assumed that the same parameter values characterize all the subjects in the experimental group. This is further confounded by the assumption of equal initial probabilities for all subjects. Sternberg says, "it must be kept in mind when this tacit assumption of individual homogeneity is made in the application of model type, that what is tested by comparisons between data and model is the conjunction of the assumption and the model type and not the model type alone."

Glaser (1967) and particularly Sternberg, point to some implications resulting when the homogeneity assumption is not met. Many of these implications relate to the inter-subject variance which seems to be smaller for the model than for the data. Sternberg gives some references to a very few studies trying to cope with this problem. Little work has been done in which variation in the learning rate parameters is allowed.

Now since the set of histories H depends on the initial probability parameters, and since it is updated on the basis of C and R , H lacks a complete description due to the short-comings introduced in considering C and R .

The present study is motivated by this absence of adequate formalization of individual and item differences. Heterogeneity of individuals and items (compounded) will be introduced in the hope of achieving better estimation and testing of the models, and eventually better instructional procedures that may be differentially sensitive to deviations from homogeneity.

It should be clear from the preceding paragraphs that many applications could and should depend on individual and item differences. One example (Matheson, 1964) points out, in a one-parameter situation, how a teaching system based on this kind of consideration improves its teaching performance as successive students are taught by it.

The most recent example of allowing the parameters of the model to vary with students and items in order to develop an optimal teaching procedure is described by Laubsch (1969). Laubsch partitioned the learning rate parameters of the RTI learning model (a more general model than the LM and the OEM) into subject and item components where the effects of the components on the composite parameter were almost additive (cf., fixed-effects ANOVA). Since the RTI has two parameters (composite), for $m$ items and $s$ subjects, $2(m+s)$ parameter estimates were needed to specify the learning parameters for $ms$ subject-items. Under the numerical maximum likelihood procedure Laubsch suggested, the approach becomes unrealistic for most practical situations -- even on the fastest computer.

Nevertheless, his results indicate the importance of incorporating heterogeneity assumptions into the learning models in optimal teaching situations.

CHAPTER II

# EFFECTS ON LEARNING PROPERTIES OF HAVING CONTINUOUS DISTRIBUTIONS OVER THE LEARNING RATES

## II.1 INTRODUCTION

The OEM with parameters $c$ and $g$ was introduced in Chapter I. The LM with parameters $\alpha$ and $q$ also was presented. In this chapter we consider the effect on learning properties, e.g., expected total errors $E(T)$ or response n-tuples probabilities $\{x_j, x_{j+1}, \ldots, x_{n+j-1}\}$† when the learning parameters are no longer exact numbers but rather they have now become random variables.

The effect of such modification introduces heterogeneity of individuals and curriculum items into the models expressed in terms of the distribution of the individuals or items population. An individual or an item may then have learning rate parameters which are random variables from this distribution. Mathematically, the population's learning properties are no longer conditional on given $c$ or $g$ ($\alpha$ or $q$). Thus if $E(t|g,c)$ denotes the conditional expectation then $E_B(E(T|g,c))$, with respect to the distribution $B$ of $g$ and $c$, is the expectation of $T$ with the effects due to the parameter differences integrated in.

In the remainder of this introductory section (II.1), we review the existing literature on stochastic models with prior distribution assumption on the parameters (II.1.1) and hence the reasons that compelled us to choose independent

†The notation $\{x_j, x_{j+1}, \ldots, x_{n+1-1}\}$ represents the joint probability distribution of the random variables $(x_j, x_{j+1}, \ldots, x_{n-j+1})$

bivariate beta density as a prior for the learning parameters, (II.1.1).

In Section II.2, we describe the effects on total error statistics for the OEM and the LM of having an independent bivariate beta distribution over the learning rates.

The effects on Response 4-Tuple probabilities under this prior is examined in Section II.3. In Section II.3.1 we derive the probabilities of response sequences over trials 2 to 5 for the OEM and the LM and propose the minimum $\chi^2$ procedure for estimating the four prior parameters using 16 response probabilities. The experimental data and the results are tabulated in II.3.2.

Finally the discussion and conclusions are presented in Section II.4.

## II.1.1 General Remarks

Very little work has been done in which variation in the learning rate parameters is allowed. One example appears in Bush and Mosteller's (1959) analysis of the Solomon-Wynne data: the LM was used with a beta distribution of $\alpha$ values. In certain respects this generalization improved the agreement between the model and the data.

Another example is Gregg and Simon's (1967) analysis of the Bower-Trabasso data: the Concept Identification model was used with a uniform prior distribution of $c$ values in a certain range $[c_1, c_2]$, $0 < c_1 \leq c \leq c_2 < 1$. Their conclusion was that for large individual differences expressed by the range size $[c_1, c_2]$ the increase in the variance of total number of errors is barely detectable. They go on further to say: "By similar arguments we can show that almost all the 'fine grain' statistics reflect mainly a random component ... Hence the statistics are insensitive to individual differences, or, for that matter, to any other psychological aspects of the subjects' behavior that might be expected to effect the statistics."

Birnbaum (1969) modified his previous work on a Logistic Model for Mental Test (1968) by further assuming a logistic prior distribution on the ability parameter $\theta$ . Thus if $\underset{\sim}{x} = \langle x_1, x_2, \ldots x_m \rangle$ denotes the examinee's response pattern where $x_k = 1$ or $0$ indicating respectively, correct or incorrect response to item $k$ , the probability of a correct response on item $k$ is $\Psi(Da_k(\theta - b_k))$ , for an examinee with ability level $\theta$ ; where $\Psi(D\theta) = [1 + e^{-D\theta}]^{-1}$ , $b_k$ is a parameter indicating a difficulty-level of test item $k$, $a_k$ is a parameter indicating the item's sensitivity or power of discrimination among ability levels not far from $b_k$ and $D$ is a constant. The general Logistic Model is represented by

$$\{X = \underset{\sim}{x} | \theta\} = \prod_{k=1}^{n} \Psi[Da_k(\theta - b_k)]^{x_k} \, \Psi[-Da_k(\theta - b_k)]^{1-x_k} \qquad -\infty < \theta < \infty . \tag{1.1}$$

Under a logistic prior assumption on $\theta$ , (1.1) is interpreted as the conditional probability of the response pattern $\underset{\sim}{x}$ , given that an examinee, randomly selected from a population with abilities distributed as indicated, has ability $\theta$ .

The unconditional probability of response pattern $\underset{\sim}{x}$ is

$$\{X = \underset{\sim}{x}\} = D \int_{-\infty}^{\infty} \{X = \underset{\sim}{x} | \theta\} \psi(D\theta) d\theta \ . \qquad (1.2)$$

The conditional density function of $\theta$ , given $X = \underset{\sim}{x}$ , $f(\theta | \underset{\sim}{x})$ is easily calculated and corresponding statistical inference methods are developed (Birnbaum, 1969).

Finally, Silver (1963) considered general Markov Chains (MC) situations with observable states where the transition probabilities are r.v.'s themselves and are Dirichlet distributed (III.2.2). Thus, for example, in a general 3-state MC with transition probabilities $(p_{ij})$ the Dirichlet prior on the $i^{th}$ state transition probabilities can be written as

$$f_{p_{i1}p_{i2}p_{i3}}(x_{i1}, x_{i2}, x_{i3}) = [B(r_i, s_i, t_i)]^{-1} \, x_{i1}^{r_i - 1} \, x_{i2}^{s_i - 1} \, x_{i3}^{t_i - 1}$$

where

$$B(r_i, s_i, t_i) = \frac{\Gamma(r_i)\Gamma(s_i)\ \Gamma(t_i)}{\Gamma(r_i + s_i + t_i)}$$

and

$$\sum_{j=1}^{3} x_{ij} = 1 \qquad 0 < x_{ij} < 1 \ . \qquad (1.3)$$

Silver considered under this setup the effect of the Dirichlet prior on MC properties such as steady state probabilities, first passage times and occupancy times. For example, consider the two-state situation where one probability is known exactly while the other is beta distributed, the beta density is the marginal of the Dirichlet distribution. We are interested in the expected values of the steady state probability for the MC with the following structure

$$P = \begin{matrix} 1 \\ \\ 2 \end{matrix}\begin{bmatrix} 1-a & a \\ & \\ b & 1-b \end{bmatrix} \tag{1.4}$$

where $a$ is assumed exactly known but $b$ has the beta density $f_b(x) = f_\beta(x|m,n)$. For a given pair $(a,b)$ the steady state probability of being in state 2 is $\pi_2 = \frac{a}{a+b}$; however, since $b$ is beta distributed, then

$$E(\pi_2) = E\left(\frac{a}{a+b}\right) = \int_0^1 \frac{a}{a+x} f_b(x)dx = \frac{1}{B(m,n)} \int_0^1 \frac{a}{a+x} x^{m-1}(1-x)^{n-1}dx . \tag{1.5}$$

Only in special cases can $E(\pi_2)$ be exactly evaluated.

II.1.2 The Evolution of the Method

All of the studies mentioned in II.1.1, except for Silver's, considered only univariate situations where only one parameter was allowed to vary. Gregg and Simon's approach is a special case of the Bush and Mosteller one in the sense that the uniform density is a special case of the univariate Beta density $f_c(x|m,n)$ with $m = n = 1$. The general statements made therefore by Gregg and Simon on the basis of a uniform prior are unwarranted. Were they to choose a "richer" prior density and a different range of $c$ values the results would have probably been markedly different. We shall later demonstrate a similar situation to the one discussed in their study where the prior does change the variance considerably.

Birnbaum's method is unsuitable in our context for several reasons. It lacks the classical psychological description of the learning process. His prior on ability is distributed on the whole real line whereas our parameters are distributed on the unit square. Finally it is computationally quite difficult.

In contrast, Silver's approach possesses a multivariate prior distribution but it is restricted to MC situations with observable states only. The LM is not a Markov Chain and Silver's estimation procedures for the prior parameters are inapplicable for the case of non-observable transition probabilities as is the case with the OEM.

Our research goals included finding a general family of bivariate distributions rich enough in parameters. Such a family had to assume a variety of shapes and provide us with posterior distribution of $c$ and $g$ and also a measure of association between $c$ and $g$.

Our first inclination was to consider transformations from existing distributions on $R^2$ to the unit square. Thus if $X$ and $Y$ are r.v.'s from a Bivariate Normal BVN(X,Y) with five parameters we may let $c = \frac{e^X}{1+e^X}$ and

$g = \frac{e^y}{1+e^y}$ , i.e., $c = f(x)$ and $g = f(y)$ . For any statistics $s = s(g,c)$ of the OEM or the LM, the integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{s \mid f(x)f(y)\} f_{BVN}(x,y)\,dx\,dy$$

could not be evaluated in a closed form and a fortiori estimation procedures for the prior parameters would be impossible.

If X and Y are Bivariate Logistic the same problem exists but now c and g are c.d.f.'s and as such are uniformly distributed.

If c and g are Dirichlet distributed then they are defined only on the simplex $c + g \leq 1$ and we have inadequate domain for both parameters. Finally, since the OEM can be represented as a three-state Absorbing Markov Chain

$$\begin{array}{c|ccc} & L & S & E \\ \hline L & 1 & 0 & 0 \\ S & c & \bar{c}g & \bar{c}\bar{g} \\ E & c & \bar{c}g & \bar{c}\bar{g} \end{array} \qquad \text{where} \qquad \begin{array}{l} \bar{c} = 1 - c \\ \\ f = 1 - g \end{array} \tag{1.6}$$

we may suppose that c, $(\bar{c}g)$ and $(\overline{c\underline{g}})$ are Dirichlet distributed

$$f_{c,\overline{c}g,\overline{cg}} = [B(r,s,t)]^{-1} x_1^{r-1} x_2^{s-1} (1-x_1-x_2)^{t-1} \; ; \; \sum_{j=1}^{3} x_i = 1 \; . \tag{1.7}$$

But under this assumption it becomes immediately clear that c and g are independently distributed with beta densities $f_c(x \mid r, s+t)$ and $f_g(x \mid s,t)$ respectively.

To establish this, make the transformation $x = x_1$, $y = x_2/(1-x_1)$. The Jacobian is $\frac{D(x_1 x_2)}{D(x,y)} = 1/(1-x_1)$.

Considerable effort was made by the present author and others to find a more adequate prior bivariate distribution with sufficient number of parameters. Unfortunately all efforts were unsuccessful. Moreover, even for the simple case of 2-state MC with one parameter, beta distributed, the integral 1.5 is not evaluated in a closed form.

## II.2 EFFECTS ON TOTAL ERROR STATISTICS HAVING INDEPENDENT BETA DISTRIBUTIONS OVER THE LEARNING RATES

For the reasons enumerated in the preceding section we will consider for the remainder of this chapter only the case where $c$ and $g$ are independent r.v.'s from beta densities $f_c(x|m,n)$ and $f_g(y|r,s)$ .

Let $T$ be the total number of errors, in n learning trials, where $n \to \infty$. Atkinson, et al. (1965, ch. 3), derived the following total error properties for given $c$ and $g$.

| | | OEM | LM | |
|---|---|---|---|---|
| Distribution | $\{T=0\|g,c\}$ | $bg$ | -- | |
| | $\{T=k\|g,c\}$ | $(1-b)^k b(1-c)^{-1}$ | -- | |
| Mean | $E(T\|g,c)$ | $(1-g)\|c$ | $q\|1-\alpha$ | (2.1) |
| Variance | $V(T\|g,c)$ | $E(T\|g,c)[E(T\|g,c)(1-2c)+1]$ | $E(T)-\frac{q^2}{1-\alpha^2}$ | |

$$b = [1-(1-c)g]^{-1} .$$

We now calculate the unconditional properties for the OEM:

$$E^*(T) = E_\beta((E(T|g,c))$$

$$= E_\beta\left(\frac{(1-g)}{c}\right)$$

$$= D\int_0^1 \int_0^1 \left(\frac{1-g}{c}\right) g^{r-1}(1-g)^{s-1} c^{m-1}(1-c)^{n-1} dg dc$$

where

$$D = [B(m,n)B(r,s)]^{-1}$$

It is readily found that

$$E^{*}(T) = \frac{B(r,s+1)B(m-1,n)}{B(r,s)\,B(m,n)} = \left(\frac{s}{r+s}\right)\left(\frac{m+n-1}{m-1}\right) \tag{2.3}$$

To calculate the variance,

$$V^{*}(T) = E^{*}(T^{2}) - E^{*2}(T) = E_{\beta}(E(T^{2}\,|g,c)) - E^{*2}(T) , \tag{2.4}$$

we need to derive $E_{\beta}(E(T^{2}|g,c))$ :

$$E_{\beta}(E(T^{2}\,|g,c)) = E_{\beta}\left(E(T\,|g,c)\left(\frac{2-b}{b}\right)\right)$$

$$= E_{\beta}\left(\frac{2(1-g)^{2}}{c^{2}} + \frac{2g(1-g)}{c} - \frac{1-g}{c}\right)$$

$$= D[2B(r,s+2)B(m-2,n)$$

$$+ B(m-1,n)[2B(r+1,s+1) - B(r,s+1)]] \tag{2.5}$$

where D is as above.

The unconditional mean for the LM is the same as the OEM mean. Unfortunately the unconditional variance for the LM cannot be evaluated in a closed form because of the $(1-\alpha)^{2}$ term in the conditional variance.

For Atkinson and Crothers' data (1964) and our estimates of the prior parameters, to be described in the next section, we calculated the expected value of the total number of errors, $E^{*}(T)$ , for experiments Ia and Ib and the variance of the total number of errors, $V^{*}(T)$ , for experiments Ia, Ib, Vc, and Ve using Eqs. 2.3 and 2.4.

Table 2.7 presents our prior estimates for all four experiments for the unconditional OEM (OEM*) and the "c" estimates derived by Atkinson and Crothers for the conditional model (OEM).

In Table 2.8 we report the $E^*(T)$ values for OEM* calculated by using Eq. 2.3. Also listed are $E(T)$ values for OEM calculated by using the equation $E(T) = \frac{1-g}{c}$ for $g = \frac{1}{3}$ and "c" values as reported in Table 2.7. Atkinson's predictions using the LS-3 Model are presented in the right-hand column. These predictions may be compared with the observed values listed in the left-hand column. Our estimate for Ib is closer to the observed value than is the LS-3's prediction; for Ia our prediction falls farther afield. The conditional estimates, $E(T)$, deviate the most from the observed values. The expected value of the total number of errors for the LM* is calculated from Eq. 2.3 as is the value for OEM*, but generally for different estimates of the prior parameters. Using these estimates, the LM* gave the poorest predictions of the expected value: 2.015 for experiment Ia and 1.0433 for experiment Ib. These predictions were not included therefore in Table 2.8.

The variances for experiments Ia, Ib, Vc, and Ve were calculated by using Eq. 2.4 and are presented in Table 2.9; the conditional variances are calculated from the equation $V(T) = E(T)[E(T)(1-2c)+1]$.

Table 2.9 demonstrates that the variance of total errors is very sensitive indeed to individual differences. For

TABLE 2.7

| EXPERIMENT | PARAMETER | | | | |
|---|---|---|---|---|---|
| | r | s | m | n | c |
| Ia | 53.938 | 53.595 | 15.444 | 27.094 | .328 |
| Ib | 54.128 | 63.123 | 34.124 | 76.625 | .328 |
| Vc | 3.0 | 3.0 | 2.0 | 12.250 | .172 |
| Ve | 10.5 | 10.5 | 3.0 | 8.0 | .289 |

Parameter Estimates for $OEM^*$ and OEM.

TABLE 2.8

| EXPERIMENT | obs | Pred(OEM ) | Pred($OEM^*$) | Pred(LS-3) |
|---|---|---|---|---|
| Ia | 1.52 | 1.74 | 1.44 | 1.54 |
| Ib | 1.65 | 2.03 | 1.78 | 1.79 |

Observed and Predicted Expectations for Experiments Ia and Ib

TABLE 2.9

| EXPERIMENT | VARIANCE | |
|---|---|---|
| | V(T) | $V^*$(T) |
| Ia | 2.45 | 2.22 |
| Ib | 3.45 | 3.22 |
| Vc | 16.83 | 47.91 |
| Ve | 5.44 | 17.16 |

Predicted Conditional and Unconditional Variances small differences it can usually be expected that the unconditional variance $V^*(T)$ will be slightly larger than $V(T)$. Experiments Ia and Ib were run with college students and almost no errors were committed after the second trial as can be seen from Table 3.5. In these two experiments the $V^*(T)$ variances are actually slightly below the conditional ones.[†]

On the other hand, experiments Vc and Ve were run with four and five year old children and there was a large difference in their performance. This difference is expressed overwhelmingly in the magnitude of the difference between the variances, i.e., $V^*(T) \gg V(T)$. It is clear therefore that the model is sensitive enough to detect individual differences if there are any. The reason that Gregg and Simon detect only a slight difference may be attributed to their choice of a uniform prior with a restricted range which may not describe the differences in their data.

As indicated above in the case of integral (1.5), it is not clear how to evaluate the unconditional quantity in a closed form under a beta prior assumption when the quantity of interest is a function of the steady state probabilities. The conditional distributions of the total errors and of the trial of last error depend on the probability of entering

---

[†]For similar results for the Solomon-Wynne data see Bush and Mosteller (Ibid).

the learned state. As such these conditional distributions possess a denominator which is a function of $(1-(1-c)g)$. Thus the distribution of total errors is

$$P(T=k|g,c) = ((1-g)(1-c))^k c(1-g(1-c))^{-k-1}(1-c)^{-1} \qquad k \geq 1 \qquad (2.9)$$

and the distribution of the trial of last error is

$$P(L=k|g,c) = (1-c)^{k-1}(1-g)c(1-g(1-c))^{-1} \qquad k \geq 1 \qquad (2.10)$$

The magnitude of the above problem is described in the following special case where it is possible to get a closed form result.

Theorem. If $c$ and $g$ are independently beta variables with parameters $(m,n)$ and $(r,s)$ respectively and $s+r=1$, then the unconditional distribution

$$P(T=k) = (B(m,n)B(r,s))^{-1}B(k+s,r)B(m+s,n+k-1), \qquad (2.11)$$

where $k$ is a positive real number.

Proof of the Theorem.

Given

$$\{T=k|g,c\}^{\dagger} = [(1-g)(1-c)]^k c(1-c)^{-1}(1-g\bar{c})^{-k-1}$$

† The notation $\{X|\Theta\}$ represents the probability distribution of a random variable $X$ given the state of information $\Theta$.

$$f_c(x) = [B(m,n)]^{-1} x^{m-1}(1-x)^{n-1} \qquad m,n > 0$$

$$f_g(y) = [B(r,s)]^{-1} y^{r-1}(1-y)^{s-1} \qquad r,s > 0$$

then

$$\{T=k\} = \int_0^1 \int_0^1 \{T=k \mid x,y\} f_c(x) f_g(y) d_x d_y$$

$$\{T=k\} = D\int_0^1 x^{m+1-1}(1-x)^{k+n-1-1}\left[\int_0^1 y^{r-1}(1-y)^{k+s-1} [(1-x(1-y)]^{-k-1} dy\right] \qquad (2.12)$$

where

$$D = [B(m,n)B(r,s)]^{-1} \quad \text{and} \quad 0 < [1-x(1-y)] < 1$$

Consider the first integration with respect to $y$ :

$$I_g = \int_0^1 y^{r-1}(1-y)^{k+s-1}[1-x(1-y)]^{-k-1} dy \qquad (2.13)$$

$I_g$ is known as the Euler-Integral and is defined in terms of the Hypergeometric Function $F(a,b;c;z)$ [see Erdélyi, 1953, Vol. 1].

$$F(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}$$

$$(Rc > Rb > 0) \qquad (2.14)$$

$F(a,b;c;z)$ itself is defined in terms of infinite series. Here it suffices to note the following recursive relation (Ibid):

$$F(a,b;c;z) = (1-z)^{c-a-b}F(c-a,c-b;c;z) \quad . \qquad (2.15)$$

From (2.14),

$$I_g = \frac{\Gamma(r)\Gamma(k+s)}{\Gamma(k+s+r)} \; F(k+1,r;k+s+r;(1-x)) \quad .$$

In our case $s+r=1$ and using (2.15) we now have

$$I_g = \frac{\Gamma(r)\Gamma(k+s)}{\Gamma(k+1)} \; [1-(1-x)]^{s-1}F(0,k+s;k+1;(1-x)) . \qquad (2.16)$$

Again from (2.14),

$$I_g = X^{s-1}\int_0^1 Y^{k+s-1}(1-y)^{r-1}dy$$

$$I_g = X^{s-1}B(k+s,r) \quad .$$

Now substituting $I_g$ in (2.12)

$$\{T=k\} = DB(k+s,r)\int_0^1 x^{m+s-1}(1-x)^{n+k-1-1}dx \; ,$$

from which

$$\{T=k\} = DB(k+s,r)B(m+s,n+k-1) \; , \qquad (2.17)$$

and this is the desired equation (2.11).

A closed form integration is possible for a similar restriction on $m,n$ , i.e., $m+n=1$ .

The posterior probabilities of $\theta=(c,g)$ given $T=k$ can now be derived using Bayes' theorem

$$\{\theta\,|T=k\} = \frac{\{T=k\,|\theta\}\{\theta\}}{\{T=k\}} \qquad (2.18)$$

From our estimation results for the four prior parameters $m$, $n$, $r$ and $s$ calculated for the Atkinson and Crothers data to be described in the next section, it became clear that a restriction $r+s=1$ or $m+n=1$ does not in fact hold for the data. The reason for this is obvious from the expressions for the prior variances of $c$ or $g$; under such restriction these variances must be quite large. Our results show that these variances are very small indeed, which is typical for Paired-Associate Learning data.

In order to demonstrate the effect of any statistics introduced by the prior assumption we would need an estimate of the four prior parameters $m$, $n$, $r$ and $s$. The mean $E^*(T)$ and the variance $V^*(T)$ are clearly not enough to estimate these four parameters. On the other hand, moments for the other statistics, under an independent bivariate beta prior, could not be derived. We could, however, estimate the parameters by considering Response n-Tuple probabilities and that we do in the following section.

## II.3 EFFECTS ON RESPONSE 4-TUPLE OF HAVING INDEPENDENT BETA DISTRIBUTIONS OVER THE LEARNING RATES

### II.3.1 Probabilities of Response Sequences Over Trials 2 to 5

Response 4-tuple is the sequence

$$O_{i,n} = \langle x_n = j_n, x_{n+1} = j_{n+1}, \ldots, x_{n+3} = j_{n+3} \rangle \qquad (3.1)$$

where $i = 1, 2, \ldots, 16$ and $j_i = 0$ or $1$ denoting a correct or an incorrect response on trial $i$, respectively. Here we use only the response 4-tuple and only over trials 2 to 5; these quantities are particularly useful in making comparisons among the two models -OEM and the LM - with or without the prior assumption. They are also useful in comparing the unconditional models, i.e., with priors, with more elaborate conditional models, i.e., without priors. In our case $n = 2$ in Eq. (3.1).

We now present the arrays of prediction probabilities over trials 2 to 5. We do not present here the derivations for $\Pr(O_{i,2})$ since they are straightforward and involve only elementary probability theory. (Readers not familiar with the methods involved in such derivations can consult Atkinson, et al., 1965.) Notation-wise we present the probability of the 16 sequences as $\{j_2, j_3, j_4, j_5\}$ where $j_i = 0$ or $1$ indicating correct or incorrect response on trial $i$ . Thus $\{1,0,0,1\}$ is the probability of errors on trials 2 and 5 and correct responses on trials 3 and 4. To derive our equations in the form of Tables 3.1 to 3.4 we use some elementary probability identities, for example, $\{1,1,1,0\} = \{1,1,1\} - \{1,1,1,1\}$ or $\{1,1,0,0\} = \{1,1\} - \{1,1,0,1\} - \{1,1,1,0\} - \{1,1,1,1\}$ . When this procedure is used starting with the sequence $O_{16} = \langle 1,1,1,1 \rangle$ of four errors only one new term involving $c$ and $g$ is introduced in each subsequent equation. For example: $O_{15} = (1-c)^3(1-g)^3 - \{O_{16}\}$ as seen from the first identity above. The derivations of response 4-tuple for the LM are just as simple.

The next step is to find the unconditional probabilities for the two models. The derivation here is straightforward. Let $D = [B(r,s)B(m,n)]^{-1}$ . For the OEM probabilities we integrate the conditional probabilities listed in Table 3.1. For example

$$\{O_{16}\} = D\int_0^1\int_0^1 (1-x)^4(1-y)^4 x^{m-1}(1-x)^{n-1}y^{r-1}(1-y)^{s-1}dxdy$$

and we get

$$\{O_{16}\} = DB(m,n+4)B(r,s+4) .$$

The next sequence is $O_{15}$ for which

$$\{O_{15}\} = D\int_0^1\int_0^1 (1-x)^3(1-y)^3 x^{m-1}(1-x)^{n-1}y^{r-1}(1-y)^{s-1}dxdy - \{O_{16}\}$$

and the result

$$\{O_{15}\} = DB(m,n+3)B(r,s+3) - \{O_{16}\}$$

and so on. The complete array for the OEM is given in Table 3.3.

The derivations for the unconditional probabilities of the LM are the same, using Table 3.2. Here for later comparison purposes we let $\alpha = 1 - c$ and $q = 1 - g$ .

In order to make predictions from Tables 3.3 and 3.4 estimates of the prior parameters are needed. Toward this end we minimize the $\chi^2$ associated with the $O_i$ events Let $\{O_i; m, n, r, s\}$ denote the probability of the event $O_i$ where $m$ , $n$ , $r$ and $s$ have been listed to make explicit the fact that the expression is a function of the four prior parameters. Further, let $N(O_i)$ denote the observed frequency of outcome $O_i$ over trials 2 to 5. Finally, let $T = N(O_1) + N(O_2) + \ldots + N(O_{16})$ . Then we define the function

$$\chi^2(m,n,r,s) = \sum_{i=1}^{16} \frac{[T\{O_i;m,n,r,s\} - N(O_i)]^2}{T\{O_i;m,n,r,s\}} \qquad (3.2)$$

and select our estimates of $r$ , $s$ , $m$ , and $n$ so they jointly minimize the function (3.2). It is difficult to carry out this minimization analytically and consequently we programmed a high-speed computer to carry out a numerical search over all possible parameters until a minimum is obtained that is accurate up to one decimal place. If we assume that all stimulus items are independent and identical, then under the null hypothesis it can be shown that this minimum $\chi^2$ has the usual limiting distribution with

TABLE 3.1

$$\{O_{16}\} = \{1,1,1,1\} = (1-c)^4(1-g)^4$$

$$\{O_{15}\} = \{1,1,1,0\} = (1-c)^3(1-g)^3 - \{O_{16}\}$$

$$\{O_{14}\} = \{1,1,0,1\} = (1-c)^4(1-g)^3 g$$

$$\{O_{13}\} = \{0,1,0,0\} = (1-c)^2(1-g)^2 - \{O_{14}\} - \{O_{15} - \{O_{16}\}$$

$$\{O_{12}\} = \{1,0,1,1\} = (1-c)^4(1-g)^3 g$$

$$\{O_{11}\} = \{1,0,1,0\} = (1-c)^3(1-g)^2 g - \{O_{12}\}$$

$$\{O_{10}\} = \{1,0,0,1\} = (1-c)^4(1-g)^2 g^2$$

$$\{O_9\} = \{1,0,0,0\} = (1-c)\ (1-g) - \sum_{i=10}^{16} \{O_i\}$$

$$\{O_8\} = \{0,1,1,1\} = \{O_{14}\}$$

$$\{O_7\} = \{0,1,1,0\} = \{O_{11}\}$$

$$\{O_6\} = \{0,1,0,1\} = \{O_{10}\}$$

$$\{O_5\} = \{0,1,0,0\} = (1-c)^2(1-g)g - \{O_6\} - \{O_7\} - \{O_8\}$$

$$\{O_4\} = \{0,0,1,1\} = \{O_{10}\}$$

$$\{O_3\} = \{0,0,1,0\} = (1-c)^3(1-g)g^2 - \{O_4\}$$

$$\{O_2\} = \{0,0,0,1\} = (1-c)^4(1-g)g^3$$

$$\{O_1\} = \{0,0,0,0\} = 1 - \sum_{i=2}^{16} \{O_i\}$$

OEM Probabilities of Response Sequences
Over Trials 2 to 5 given g and c .

TABLE 3.2

$$\{O_{16}\} = \{1,1,1,1\} = \alpha^{10}q^{4}$$

$$\{O_{15}\} = \{1,1,1,0\} = \alpha^{6}q^{3} - \{O_{16}\}$$

$$\{O_{14}\} = \{1,1,0,1\} = \alpha^{7}q^{3} - \{O_{16}\}$$

$$\{O_{13}\} = \{1,1,0,0\} = \alpha^{3}q^{2} - \{O_{14}\} - \{O_{15}\} - \{O_{16}\}$$

$$\{O_{12}\} = \{1,0,1,1\} = \alpha^{8}q^{3} - \{O_{16}\}$$

$$\{O_{11}\} = \{1,0,1,0\} = \alpha^{4}q^{2} - \{O_{12}\} - \{O_{15}\} - \{O_{16}\}$$

$$\{O_{10}\} = \{1,0,0,1\} = \alpha^{5}q^{2} - \{O_{12}\} - \{O_{14}\} - \{O_{16}\}$$

$$\{O_{9}\} = \{1,0,0,0\} = \alpha\ q - \sum_{i=10}^{16} \{O_{i}\}$$

$$\{O_{8}\} = \{0,1,1,1\} = \alpha^{9}q^{3} - \{O_{16}\}$$

$$\{O_{7}\} = \{0,1,1,0\} = \alpha^{5}q^{2} - \{O_{8}\} - \{O_{15}\} - \{O_{16}\}$$

$$\{O_{6}\} = \{0,1,0,1\} = \alpha^{6}q^{2} - \{O_{8}\} - \{O_{14}\} - \{O_{16}\}$$

$$\{O_{5}\} = \{0,1,0,0\} = \alpha^{2}q - \{O_{6}\} - \{O_{7}\} - \{O_{8}\} - \{O_{13}\} - \{O_{14}\} - \{O_{15}\} - \{O_{16}\}$$

$$\{O_{4}\} = \{0,0,1,1\} = \alpha^{7}q^{2} - \{O_{8}\} - \{O_{12}\} - \{O_{16}\}$$

$$\{O_{3}\} = \{0,0,1,0\} = \alpha^{3}q - \{O_{4}\} - \{O_{7}\} - \{O_{8}\} - \{O_{11}\} - \{O_{12}\} - \{O_{15}\} - \{O_{16}\}$$

$$\{O_{2}\} = \{0,0,0,1\} = \alpha^{4}q - \{O_{4}\} - \{O_{6}\} - \{O_{8}\} - \{O_{10}\} - \{O_{12}\} - \{O_{14}\} - \{O_{16}\}$$

$$\{O_{1}\} = \{0,0,0,0\} = 1 - \sum_{i=2}^{16} \{O_{i}\}$$

LM Probabilities of Response Sequences
Over Trials 2 to 5 Given $\alpha$ and $q$.

TABLE 3.3

$$\{O_{16}\} = \{1,1,1,1\} = D[B(r,s+4)B(m,n+4)]$$

$$\{O_{15}\} = \{1,1,1,0\} = D[B(r,s+3)B(m,n+3)] - \{O_{16}\}$$

$$\{O_{14}\} = \{1,1,0,1\} = D[B(r+1,s+3)B(m,n+4)]$$

$$\{O_{13}\} = \{1,1,0,0\} = D[B(r,s+2)B(m,n+2)] - \{O_{14}\} - \{O_{15}\} - \{O_{16}\}$$

$$\{O_{12}\} = \{1,0,1,1\} = D[B(r+1,s+3)B(m,n+4)] = \{O_{14}\}$$

$$\{O_{11}\} = \{1,0,1,0\} = D[B(r+1,s+2)B(m,n+3)] - \{O_{12}\}$$

$$\{O_{10}\} = \{1,0,0,1\} = D[B(r+2,s+2)B(m,n+4)]$$

$$\{O_{9}\} = \{1,0,0,0\} = D[B(r,s+1)B(m,n+1)] - \sum_{i=10}^{16} \{O_i\}$$

$$\{O_{8}\} = \{0,1,1,1\} = \{O_{14}\}$$

$$\{O_{7}\} = \{0,1,1,0\} = \{O_{11}\}$$

$$\{O_{6}\} = \{0,1,0,1\} = \{O_{10}\}$$

$$\{O_{5}\} = \{0,1,0,0\} = D[B(r+1,s+1)B(m,n+2)] - \{O_{6}\} - \{O_{7}\} - \{O_{8}\}$$

$$\{O_{4}\} = \{0,0,1,1\} = \{O_{10}\}$$

$$\{O_{3}\} = \{0,0,1,0\} = D[B(r+2,s+1)B(m,n+3)] - \{O_{4}\}$$

$$\{O_{2}\} = \{0,0,0,1\} = D[B(r+3,s+1)B(m,n+4)]$$

$$\{O_{1}\} = \{0,0,0,0\} = 1 - \sum_{i=2}^{16} \{O_i\}$$

OEM Probabilities of Response Sequences
Over Trials 2 to 5 in Terms of the Prior Parameters (r,s) and (m,n)

TABLE 3.4

$$\{O_{16}\} = \{1,1,1,1\} = D[B(r,s+4)B(m,n+10)]$$

$$\{O_{15}\} = \{1,1,1,0\} = D[B(r,s+3)B(m,n+6)]-\{O_{16}\}$$

$$\{O_{14}\} = \{1,1,0,1\} = D[B(r,s+3)B(m,n+7)]-\{O_{16}\}$$

$$\{O_{13}\} = \{1,1,0,0\} = D[B(r,s+2)B(m,n+3)]-\{O_{14}\}-\{O_{15}\}-\{O_{16}\}$$

$$\{O_{12}\} = \{1,0,1,1\} = D[B(r,s+3)B(m,n+8)]-\{O_{16}\}$$

$$\{O_{11}\} = \{1,0,1,0\} = D[B(r,s+2)B(m,n+4)]-\{O_{12}\}-\{O_{15}\}-\{O_{16}\}$$

$$\{O_{10}\} = \{1,0,0,1\} = D[B(r,s+2)B(m,n+5)]-\{O_{12}\}-\{O_{14}\}-\{O_{16}\}$$

$$\{O_{9}\} = \{1,0,0,0\} = D[B(r,s+1)B(m,n+1)]-\sum_{i=10}^{16}\{O_i\}$$

$$\{O_{8}\} = \{0,1,1,1\} = D[B(r,s+3)B(m,n+9)]-\{O_{16}\}$$

$$\{O_{7}\} = \{0,1,1,0\} = D[B(r,s+2)B(m,n+5)]-\{O_{8}\}-\{O_{15}\}-\{O_{16}\}$$

$$\{O_{6}\} = \{0,1,0,1\} = D[B(r,s+2)B(m,n+6)]-\{O_{8}\}-\{O_{14}\}-\{O_{16}\}$$

$$\{O_{5}\} = \{0,1,0,0\} = D[B(r,s+1)B(m,n+2)]-\{O_{6}\}-\{O_{7}\}-\{O_{8}\}-\{O_{13}\}-\{O_{14}\}-\{O_{15}\}-\{O_{16}\}$$

$$\{O_{4}\} = \{0,0,1,0\} = D[B(r,s+2)B(m,n+7)]-\{O_{8}\}-\{O_{12}\}-\{O_{16}\}$$

$$\{O_{3}\} = \{0,0,1,0\} = D[B(r,s+1)B(m,n+3)]-\{O_{4}\}-\{O_{7}\}-\{O_{8}\}-\{O_{11}\}-\{O_{12}\}-\{O_{15}\}-\{O_{16}\}$$

$$\{O_{2}\} = \{0,0,0,1\} = D[B(r,s+1)B(m,n+4)]-\{O_{4}\}-\{O_{6}\}-\{O_{8}\}-\{O_{10}\}-\{O_{12}\}-\{O_{14}\}-\{O_{16}\}$$

$$\{O_{1}\} = 1 - \sum_{i=2}^{16}\{O_i\}$$

LM Probabilities of Response Sequences
Over Trials 2 to 5 in Terms of the Prior Parameters (r,s) and (m,n)

$16 - 4 - 1 = 11$ degrees of freedom. In addition to having desirable estimation properties[†] the minimum $\chi^2$ also provides a measure of adequacy of any single model and a method for comparing the fit of several models, if the degrees of freedom are equal. If several models are being analyzed, each involving a different number of free parameters then the probability levels of the $\chi^2$ may be compared. The degrees of freedom associated with a model that requires $k$ parameters to be estimated from the data are $df = 16 - k - 1$. The one is subtracted because of the restriction that the 16 probabilities sum to 1. There are other numerical estimation procedures available, e.g., numerical maximum likelihood or least-squares procedures, but since the data described in this chapter was analyzed originally by means of minimum $\chi^2$ procedures, we prefer this method in order to facilitate later comparisons between the original analysis and ours.

II.3.2 Data Analysis

A summary and analysis of the data using seven different conditional models is presented by Atkinson and Crothers (1964). For the convenience of the reader we restate the main features of the experimental procedure and data.

---

[†] See Cramér (1951, pp. 424-441) for example.

"The data was collected from eight paired-associate learning experiments that all utilize the same general experimental procedure. At the start of the experiment the subject is told the responses available to him; each alternative occurs equally often as the to-be-learned response. A response is obtained from the subject on each presentation of an item and he is informed of the correct answer following his response.

TABLE 3.5

ATKINSON AND CROTHERS

FEATURES OF THE EXPERIMENTAL PROCEDURE

| Experiment | Number of stimuli | Number of responses | Number of subjects | Pr ($c_s$) |
|---|---|---|---|---|
| Ia | 9 | 3 | 26 | .95 |
| Ib | 18 | 3 | 16 | .91 |
| II | 12 | 3 | 65 | .83 |
| III | 12 | 4 | 40 | .75 |
| IV | 16 | 4 | 20 | .84 |
| Va | 12 | 4 | 40 | .60 |
| Vc | 12 | 4 | 40 | .71 |
| Ve | 12 | 4 | 40 | .85 |

"Relevant details of each experiment are given in Table 3.5. Experiments Ia and Ib were run with college students. For both experiments the stimuli were Greek letters and the responses were the low association trigrams RIX, FUB, and GED; the experiments differed in that one used a 9 item stimulus list and the other 18 item list. Experiment II was also run with college students using 12 Greek letters as stimuli and the numbers 3, 4, 5 and 6 as the responses. Experiment III was run with 3rd and 4th grade students using 12 Greek letters as stimuli and the numbers 2, 3, 4 and 5 as the responses. Experiment IV was run with college students using double digit numbers as stimuli and the letters A, B, C and D as responses. For Experiment I-IV the experimental procedure (method of stimulus display, presentation rate, etc.) was the same as described by Bower (1961). In Experiment V, a group of four and five year old children learned a list of paired-associates each day for five consecutive days. The lists were composed of double digit numbers as stimuli and letters as reponses but the stimuli and responses were different for each list. To simplify the discussion, only results for days 1, 3, and 5 are presented (labeled Experiments Va, Vc, and Ve respectively); however these data are representative of the results for the full experiment."

Atkinson and Crothers carried the original analysis of these eight experiments for seven different conditional models, i.e., models for which the learning parameters are fixed constants for the population of subject-items. The reason for considering response sequences over trials 2 to 5 only is provided by the fact that a major portion of the learning occurred during the first five trials. This fact is indicated in the last column of Table 3.5 where $Pr(x_5 = 0)$ is presented; in five of the eight experiments the subjects have reached a correct response level of 0.83 or better on trial 5.

For the convenience of the reader Tables 3.6, 3.7, 3.8, and 3.9 are reproduced directly from Atkinson and Crothers' study.

The $\chi^2$ minimization procedure described in Eq. (3.2) was applied to the data of observed frequencies presented in Table 3.6.

Table 3.7 presents the parameter estimates associated with the minimum $\chi^2$ values for the conditional models. Table 3.7* presents on the other hand the estimates of the four prior parameters $r$, $s$, $m$, and $n$ that minimize the $\chi^2$ function for the unconditional models OEM* and LM*. This table summarizes some of the data presented in the appendix to this chapter which describes two or usually three sets of the best estimates for both models and for all eight experiments. The estimates were calculated by the computer minimization program mentioned before.†

Table 3.7** summarizes the estimated values of the prior means and variances of the beta densities of $g$ and $c$. The prior means and variances are calculated by substituting the estimates of Table 3.7* in the following equations for the mean and variance of the Beta density:

The (prior) mean of $g$ is given by $\frac{r}{r+s}$ (3.10)

The (prior) variance of $g$ is given by $\frac{rs}{(r+s)^2(r+s+1)}$ (3.11)

† See appendix to this chapter.

TABLE 3.6

OBSERVED FREQUENCIES FOR THE $O_{i,2}$ EVENTS

| | Experiment | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Ia | Ib | II | III | IV | Va | Vc | Ve |
| $N(O_{1,2})$ | 123 | 125 | 303 | 160 | 117 | 82 | 144 | 216 |
| $N(O_{2,2})$ | 3 | 3 | 14 | 13 | 3 | 11 | 18 | 4 |
| $N(O_{3,2})$ | 6 | 10 | 19 | 16 | 10 | 14 | 23 | 17 |
| $N(O_{4,2})$ | 1 | 4 | 12 | 11 | 1 | 13 | 9 | 6 |
| $N(O_{5,2})$ | 16 | 21 | 54 | 24 | 15 | 22 | 28 | 34 |
| $N(O_{6,2})$ | 3 | 0 | 17 | 6 | 3 | 21 | 14 | 16 |
| $N(O_{7,2})$ | 5 | 6 | 32 | 18 | 9 | 20 | 12 | 12 |
| $N(O_{8,2})$ | 2 | 3 | 18 | 7 | 6 | 31 | 13 | 12 |
| $N(O_{9,2})$ | 43 | 55 | 125 | 57 | 54 | 58 | 62 | 66 |
| $N(O_{10,2})$ | 1 | 5 | 15 | 9 | 7 | 13 | 14 | 4 |
| $N(O_{11,2})$ | 7 | 10 | 25 | 27 | 9 | 34 | 25 | 17 |
| $N(O_{12,2})$ | 2 | 2 | 17 | 14 | 10 | 18 | 14 | 7 |
| $N(O_{13,2})$ | 15 | 30 | 61 | 33 | 34 | 34 | 28 | 29 |
| $N(O_{14,2})$ | 0 | 1 | 19 | 25 | 8 | 21 | 20 | 8 |
| $N(O_{15,2})$ | 6 | 6 | 30 | 24 | 22 | 26 | 21 | 19 |
| $N(O_{16,2})$ | 1 | 7 | 19 | 36 | 12 | 62 | 35 | 13 |
| $T$ | 234 | 288 | 780 | 480 | 320 | 480 | 480 | 480 |

TABLE 3.7

PARAMETER ESTIMATES FOR THE VARIOUS MODELS

| Model | Parameter | Experiment | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Ia | Ib | II | III | IV | Va | Vc | Ve |
| One-element | $c$ | .383 | .328 | .273 | .203 | .281 | .125 | .172 | .289 |
| Linear | $\theta$ | .414 | .328 | .289 | .258 | .297 | .164 | .250 | .336 |
| Two-phase | $c$ | .563 | .484 | .352 | .359 | .398 | .227 | .406 | .422 |
| | $\theta$ | .664 | .633 | .695 | .563 | .648 | .500 | .477 | .656 |
| RTI | $c$ | .531 | .461 | .344 | .328 | .367 | .219 | .359 | .438 |
| | $\theta$ | .820 | .805 | .867 | .797 | .859 | .727 | .711 | .789 |
| LS-2 | $a$ | .352 | .305 | .250 | .188 | .266 | .109 | .156 | .258 |
| | $f$ | .719 | .805 | .805 | .789 | .836 | .844 | .727 | .680 |
| LS-3 | $a$ | .367 | .352 | .250 | .188 | .289 | .109 | .156 | .266 |
| | $f$ | .648 | .375 | .805 | .789 | .789 | .844 | .727 | .688 |
| | $c$ | .844 | .500 | 1.000 | 1.000 | .789 | 1.000 | 1.000 | .992 |
| Two-element | $g'$ | .883 | .852 | .922 | .891 | .922 | .797 | .859 | .844 |
| | $b$ | .391 | .398 | .227 | .078 | .195 | .133 | .016 | .227 |
| | $a$ | .539 | .477 | .344 | .320 | .359 | .219 | .352 | .477 |

TABLE 3.7*

| model | parm | Ia+ | Ib | II | III | IV | Va | Vc | Ve |
|---|---|---|---|---|---|---|---|---|---|
| OFM* | r | 55.00 | 54.13 | 46.75 | 2.69 | 12.69 | 6.500 | 3.00 | 10.50 |
| | s | 53.62 | 63.12 | 54.04 | 3.00 | 20.75 | 10.500 | 3.00 | 10.50 |
| | m | 15.59 | 34.12 | 14.00 | 3.31 | 24.00 | 2.500 | 2.00 | 3.00 |
| | n | 27.12 | 76.62 | 41.99 | 16.75 | 67.69 | 21.500 | 12.25 | 8.00 |
| LM* | r | 13.59 | 2.69 | 1.82 | 1.13 | 1.75 | 2.22 | 1.00 | 1.48 |
| | s | 31.48 | 19.25 | 1.79 | 1.14 | 51.75 | 2.20 | 1.75 | 1.37 |
| | m | 1.93 | 3.00 | 2.46 | 2.06 | 3.31 | 1.29 | 1.00 | 2.42 |
| | n | 1.75 | 3.00 | 8.84 | 24.30 | 4.25 | 51.29 | 2.75 | 7.07 |

Experiment (columns Ia+ through Ve)

PARAMETER ESTIMATES FOR ALL EIGHT EXPERIMENTS

+ For explanation see appendix

TABLE 3.7**

| model | stat. | Experiment | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Ia+ | Ib | II | III | IV | Va | Vc | Ve |
| OEM* | E(g) | .502 (.516) | .461 | .463 | .472 | .379 | .382 | .500 | .500 |
| | E(c) | .363 (.356) | .308 | .250 | .165 | .262 | .104 | .140 | .273 |
| | 100V(g) | .230 (.206) | .210 | .244 | 3.730 | .684 | 1.310 | 3.57 | 1.130 |
| | 100V(c) | .531 (.237) | .190 | .329 | .654 | .208 | .373 | .791 | 1.650 |
| LM* | E(g) | .301 | .122 | .503 | .498 | .032 | .502 | .363 | .521 |
| | E(c) | .523 | .500 | .218 | .078 | .438 | .024 | .267 | .255 |
| | 10V(g) | .046 | .046 | .542 | .764 | .005 | .460 | .617 | .647 |
| | 10V(c) | .532 | .357 | .139 | .026 | .287 | .004 | .411 | .181 |

ESTIMATES OF PRIOR MEANS AND VARIANCES

+For explanation see appendix to this chapter.

TABLE 3.8

MINIMUM $\chi^2$ VALUES

| Experiment | One-element | Linear model | Two-phase | RTI | LS-2 | LS-3 | Two-element |
|---|---|---|---|---|---|---|---|
| Ia | 30.30 | 50.92 | 17.51[a] | 9.74[a] | 6.75[a] | 5.67[a] | 9.30[a] |
| Ib | 39.31 | 95.86 | 18.25[a] | 13.09[a] | 19.69[a] | 12.42[a] | 12.74[a] |
| II | 62.13 | 251.30 | 54.78 | 29.11 | 3.73[a] | 3.73[a] | 28.46 |
| III | 150.66 | 296.30 | 95.44 | 51.12 | 33.02 | 33.02 | 47.13 |
| IV | 44.48 | 146.95 | 22.39[a] | 10.66[a] | 12.32[a] | 10.77[a] | 10.32[a] |
| Va | 102.02 | 201.98 | 59.20 | 40.17 | 24.41[a] | 24.41[a] | 39.47 |
| Vc | 246.96 | 236.15 | 99.97 | 46.43 | 27.12[a] | 27.12 | 34.75 |
| Ve | 161.03 | 262.56 | 126.05 | 84.07 | 20.12[a] | 20.12[a] | 77.39 |
| Total $\chi^2$ | 836.89 | 1542.02 | 493.59 | 284.39 | 147.16 | 137.26 | 259.56 |
| *df* | 14 | 14 | 13 | 13 | 13 | 12 | 12 |

[a] Not significant at .01 level.

TABLE 3.9

OBSERVED AND PREDICTED RESPONSE SEQUENCE PROPORTIONS FOR EXPERIMENT II

| Outcomes | Observed proportion | One-element | Linear model | Two-phase | RTI | Long-short | Two-element |
|---|---|---|---|---|---|---|---|
| $O_1$ | .389 | .362 | .220 | .328 | .354 | .390 | .357 |
| $O_2$ | .018 | .007 | .045 | .008 | .017 | .017 | .018 |
| $O_3$ | .024 | .015 | .069 | .022 | .028 | .029 | .029 |
| $O_4$ | .015 | .014 | .014 | .010 | .011 | .020 | .011 |
| $O_5$ | .069 | .047 | .112 | .066 | .063 | .064 | .062 |
| $O_6$ | .022 | .014 | .023 | .012 | .013 | .020 | .013 |
| $O_7$ | .041 | .029 | .035 | .028 | .026 | .034 | .026 |
| $O_8$ | .023 | .028 | .007 | .021 | .020 | .023 | .020 |
| $O_9$ | .161 | .178 | .198 | .210 | .189 | .164 | .188 |
| $O_{10}$ | .019 | .014 | .041 | .014 | .018 | .020 | .018 |
| $O_{11}$ | .032 | .029 | .062 | .035 | .034 | .034 | .034 |
| $O_{12}$ | .022 | .028 | .013 | .021 | .020 | .023 | .020 |
| $O_{13}$ | .079 | .093 | .101 | .102 | .092 | .074 | .091 |
| $O_{14}$ | .024 | .028 | .021 | .024 | .024 | .023 | .024 |
| $O_{15}$ | .038 | .059 | .032 | .055 | .051 | .039 | .050 |
| $O_{16}$ | .024 | .055 | .007 | .042 | .040 | .026 | .039 |

TABLE 3.8*

| EXPERIMENT | OEM* | LM* |
|---|---|---|
| Ia | 5.10[a] (7.15)[+] | 10.31[a] |
| Ib | 20.21[a] | 21.75[a] |
| II | 4.45[a] | 66.26 |
| III | 22.96[a] | 66.56 |
| IV | 12.54[a] | 42.45 |
| Va | 21.36[a] | 65.88 |
| Vc | 9.92[a] | 6.80[a] |
| Ve | 17.69[a] | 47.05 |
| Total | 114.23 | 327.06 |
| df | 11 | 11 |

a Not significant at .01 level

MINIMUM $\chi^2$ VALUES

[+] For explanation see appendix to this chapter.

TABLE 3.9*

| Outcomes | observed | OEM* | LM* |
|---|---|---|---|
| $O_1$ | .389 | .391 | .327 |
| $O_2$ | .018 | .018 | .047 |
| $O_3$ | .024 | .029 | .063 |
| $O_4$ | .015 | .020 | .018 |
| $O_5$ | .069 | .063 | .088 |
| $O_6$ | .022 | .020 | .023 |
| $O_7$ | .041 | .033 | .031 |
| $O_8$ | .023 | .023 | .014 |
| $O_9$ | .161 | .161 | .135 |
| $O_{10}$ | .019 | .020 | .034 |
| $O_{11}$ | .032 | .033 | .045 |
| $O_{12}$ | .022 | .023 | .020 |
| $O_{13}$ | .079 | .073 | .065 |
| $O_{14}$ | .024 | .023 | .027 |
| $O_{15}$ | .038 | .039 | .036 |
| $O_{16}$ | .024 | .028 | .025 |

OBSERVED AND PREDICTED RESPONSE SEQUENCE PROPORTIONS

EXPERIMENT II

The mean and variance of the prior density of c are calculated by replacing the values for r by the values of m and the values of s by those of n .

Table 3.8* presents the minimum $\chi^2$ values for the OEM* and the LM* ; i.e., the values obtained by using the parameter estimates of Table 3.7* in Eq. (3.2). The $\chi^2$ value needed for significance at the 0.01 level is 24.7 for 11 degrees of freedom. All of the $\chi^2$ values for the OEM* are not significant at this level. For the LM* the $\chi^2$ values for experiments Ia, Ib, and Vc are not significant.

Finally, Table 3.9* gives the observed and predicted response sequence probabilities for experiment II and may be compared to Atkinson and Crothers' Table 3.9 of the same proportions calculated for the conditional models.

The LM* parameter estimates in Table 3.7* tend to be much smaller when compared with the same estimates for the OEM*. This fact is reflected more clearly in Table 3.7** where the prior variances for both g and c assume larger magnitude of order greater than 10 for the LM* than for the OEM*.

When comparing Tables 3.7 and 3.7**, it becomes apparent that the between-experiment values for the prior mean of c have the same relative magnitudes as the values estimated for c in Table 3.7, with the exception of the LM* value for experiment Ve. The monotonicity over the sets of experiment V data, which is described by Atkinson and Crothers with respect to Table 3.7 and is inferable from the nature of the experiments, is maintained in Table 3.7**, again with the exception noted above. It seems therefore that the parameter estimates remain relatively invariant under our prior assumption.

We next observe in Table 3.7** that the variances of c for the OEM* are larger for experiments III, Va, Vc and Ve as compared to the same variances for the other four experiments. And indeed we would have expected them to be larger because the experiments noted were run with young children; the other four experiments were run with college students whose conditioning variances are expected to be smaller. In addition, it seems that the accuracy of the predictions, especially when compared to the LS-3 model, is inversely related to the magnitude of the estimated prior mean of c .

The LM* procedure tends to ascribe higher values for both the variances of c and g . The over-estimated variances may be a consequence of the model inadequacy to account for the data. Interestingly, the highest variances in the LM* setup are for experiments Ia, Vc and Ib which are, with the exception of Vc, unlike the results for the OEM*. The accuracy of the predictions for the LM*, as may be noted from Table 3.8*, is much better for experiments Ia, Ib and Vc. A regression analysis indicated that the variance of c was the influential factor in the predictive power of the model -- the $\chi^2$ value being the dependent variable--contributing a multiple $R^2$ of .95. By adding the prior mean of c to the regression equation, the $R^2$ value improved by 15 per cent. In general, the $\chi^2$ values were highly and negatively correlated with the variance of c and the prior mean of c.

Tables 3.8* and 3.9* compared with 3.8 and 3.9 demonstrate the following facts. The OEM* is a better model than the LS-3 model. This conclusion is further supported by the pseudo-F statistics (Holland, 1965)†. The F value in this case is the ratio of total $\chi^2/88$ of the OEM* divided by the total $\chi^2/96$ of the LS-3 model. The resulting F value is .90787 which is less than 1.

The best improvements in prediction for both the OEM* and the LM* appeared for experiments possessing high prior variances of c, as was noted in the preceding paragraphs. The most remarkable improvement was noted for experiment Vc where the $\chi^2$ values dropped from 246.96 to 9.92 for the OEM* and from 236.15 to 6.80 for the LM*. The LM* value has kept its relative lower magnitude with respect to the OEM* as was the case with the conditional results.

On the whole, between models invariance does in fact hold. In other words, the $\chi^2$ values for the OEM* and the LM* do maintain relative magnitudes which correspond to the conditional models relative magnitudes. Thus the $\chi^2$ for

†The precise significant levels for the pseudo-F could not be ascertained.

all experiments with the exception of Vc are smaller for the OEM* compared with the LM* as well as the OEM values compared with the LM values.

## II.4 DISCUSSION AND CONCLUSIONS

### II.4.1. General Remarks: Mathematical Methods for the Analysis and Evaluation of Models.

Before we draw our final conclusions from the results of the previous sections we present some of the prevailing views on the mathematical methods used in the analysis and the evaluation of stochastic learning models (e.g., Sternberg, 1963). These remarks should put our conclusions in a proper perspective on the one hand and imply further areas of investigation on the other.

Many objections have been raised as to the statistical soundness of the methods involved in the analysis and the evaluation of stochastic learning models (Gregg and Simon, 1967). Unlike classical statistical inference the evaluation of stochastic learning models is not a simple acceptance-rejection problem. Neither do we satisfy the formal data requirements needed by formal statistical decision making. So, if we accept the unavailability or even the undesirability of a formal evaluation procedure, we still need some tools for informal evaluation or "plausible inference" (Pólya, 1954).

One approach of plausible inference concerns itself with the assumptions that give rise to the model which is capable of representing a theory about the learning process at hand, another with providing descriptive statistics of the data. Both approaches require critical experiments or discriminating statistics to be used in the model's evaluation. Unfortunately again, no unified method of constructing crucial experiments or analyzing discriminating statistics exists. In principle, only the investigator's imagination limits the number of different statistics that can be used to evaluate the model. Examples of such statistics are the mean learning curve, the mean trial of last error, the number of runs of a particular length and the frequencies of particular response n-tuples. Which statistics are more pertinent and how many of them are needed in order to prefer one model over another is an open question.

Following Sternberg (Ibid), consider the n-dimensional "property space" consisting of all values of the vector $(s_1, s_2, \ldots, s_n)$ where $s_j$ denotes a property (the expectation or variance of a statistic) of the model. Denote by $\bar{s}_j$ the corresponding statistics for some observed data sequences. In general, the properties depend on the parameter values, and therefore $s_j = s_j(\Theta)$ , where $\Theta$ is a vector of parameters corresponding to a point in the parameter space.

Using this terminology, most work that has been done on fitting and testing models can be thought of as a two stage process. First, estimation, in which the parameter values are selected so that a subset of the $\bar{s}_j$ agrees with the theoretical values, and, the second, testing, in which the remaining $\bar{s}_j$ are compared to their corresponding $s_j(\Theta)$ .

Clearly, conclusions from this method are conditional on the choice of properties used in each of the two stages.

The estimation procedures of the models' free parameters can be classified into two categories. Global estimation, such as maximum likelihood or minimum chi-square, usually satisfies some overall optimal criteria and cannot usually be obtained explicitly in terms of statistics of the data; and fine-grain estimation, such as the distribution of error-run lengths. Objections may be raised sometimes as to the order of which property is used for estimation and which property is used for testing. Occasionally, as in our study, the choice of which property is available for what is restricted because of the small number of statistics with analytic expressions. It has been noted also (Ibid) that using the same estimating statistics for all models to be compared does not ensure equal "fairness" to them.

Up to this point we have made some cautious statements concerning the applicability of certain methods for comparing the model and the data, and other statements concerning comparative studies of models. These points were made to warn the reader to consider past and future inferential remarks in a proper perspective, especially with respect to model comparisons. Our intention has not been to compare or select models but rather to amend the inadequacies introduced into simple learning models by ignoring the essential features of individual learning rates. Just as important was our intention to use simple learning models as baselines and aids to inference, i.e., to test whether or not the homogeneity assumption has in fact a sizable effect on the learning properties. Our method succeeded where a model-free analysis might have failed.

Before turning to discussion of our results consider a final evaluation remark. It has long been held (e.g., Galanter and Bush, 1959) that when a model predicts how behavior depends upon some experimental variable, the model parameters should be invariant to changes in that variable. This criterion when satisfied should indicate some general descriptive ability of the model. This criterion is indeed satisfied by our models' parameters as well as by many of the models' properties.

II.4.2. The Important Features of the Results

The results of this study demonstrate unequivocally that the OEM with the heterogeneity provision is still a fairly accurate model, at least for the type of data considered. More significant is the observation that individual differences have a first order effect on the predictive power of simple stochastic models. These facts are demonstrated by the large improvement in the $\chi^2$ values as well as by the accuracy of the prediction of the mean learning curve for Experiments Ia and Ib.

Properties of the models become sensitive to individual differences -- to the degree that such differences exist. This fact is demonstrated by change in magnitude of the variances of total number of errors. It can be said therefore that the first goal of plausible inference, which is having a model capable of representing the theory about the learning process, is satisfied.

The second goal of having a model which can provide descriptive statistics of the data is also fulfilled by satisfying many of the criteria partially described in the last section.

Parameter estimates remain relatively invariant under the prior assumption as does the descriptive power of the models. In addition, it can be shown that some important properties of the models remain invariant under the prior assumption, e.g., the stationarity property of presolution trials in the OEM case remains invariant as exemplified by Vincent curves or other tests.

The statistics of the prior mean and variance of the conditioning and guessing parameters of the OEM*, presented in Table 3.7** are most descriptive of the experimental data. Higher means and smaller variances of conditioning characterize the experiments run with college students. Smaller means and larger variances describe the experiments run with young children. The discrepancies between the results of experiments Ia and Ib have to be explained, again, as in Atkinson's study, in terms of the different experimental procedures used in the two experiments. This latter fact, however, may be now partially accounted for by the guessing prior mean for experiment Ib which was lower than for experiment Ia.

The last point leads us to consider next the guessing parameters and their relation to the conditioning parameters. In the OEM* situation the prior guessing means assumed higher values than are usually ascribed to them -- one over the number of response alternatives. Moreover, in spite of the independence assumption for the two prior densities, there seems to be a definite relation between the guessing and the conditioning parameters. Higher guessing parameters are associated with lower conditioning parameters and vice versa. This association is particularly strong between the mean of the one parameter and the variance of the other, i.e., a lower conditioning mean is associated with a higher guessing variance. These observations, in addition to being intuitively appealing, are supported by a large body of data on "short-term" recall (e.g., Murdock, 1961, 1963).

The studies referred to differentiate between short- and long-term memory. Items in short-term memory can be retrieved for immediate recall, but since the short-term store is of limited capacity the probability of guessing depends on the number of intervening items from one presentation of an item to its next presentation. The limited buffer capacity may be described by a forgetting parameter, the same parameter f of the LS model introduced in Chapter I. We also described in Chapter I the LS-2 model which says that at the moment an S-R pair is studied, with probability $a$ it goes into a long-term memory storage system and with probability $1-a$ the S-R pair goes into a short-term store, where it is vulnerable to interference from intervening items. When we compare the estimates of $a$ for the LS-2 in Table 3.7 and our estimates of $c$ for the OEM* in Table 3.7**, the similarity of the results is more than striking. Furthermore comparison of the $\chi^2$ values for the two models, LS-2 and OEM$^*$, between Tables 3.8 and 3.8* demonstrate again extreme closeness of the corresponding values. We have yet to account for the high guessing probabilities. We do that by rewriting the LS-2 model as a 3-state process: collapse states S and F and make the response probability in the single intermediate state (SF) a function of the forgetting parameter. We now have the following transition matrix and response probability vector:

$$\begin{array}{c} \\ L \\ SF \\ U \end{array}\begin{array}{c} \begin{array}{ccc} L & SF & U \end{array} \\ \left[\begin{array}{ccc} 1 & 0 & 0 \\ a & 1-a & 0 \\ a & 1-a & 0 \end{array}\right] \end{array} \begin{array}{c} Pr(\text{correct}) \\ \left[\begin{array}{c} 1 \\ 1-f+fg \\ g \end{array}\right] \end{array} \qquad (4.1)$$

The guessing probability for state SF is $1-f+fg$ which is larger than the guessing probability of $g$ alone and may explain the high guessing estimates that we calculated.

Atkinson and Crothers actually tried this collapsing procedure for the LS-3 model (Ibid, Eq. 25), but had allowed the additional parameter $c$ of the LS-3 model to be different from 1, i.e., there was a positive probability $1-c$ of staying in the unlearned state U. When they applied this model to the four-tuple response data, Atkinson and Crothers reached the smallest $\chi^2$ of all the models described in their paper. The estimates for $c$ under this setup were all close to 1 which may indicate that the model described by Eq. (4.1) is the most plausible model yet.

### II.4.3. Further Research and Conclusions

The empirical results confirm the hypothesis that the heterogeneity assumption increases the predictive power of simple learning models and has a sizable effect on their learning properties.

Further theoretical research should be directed toward finding more satisfying prior bivariate (multivariate) distributions on the unit square (n-dimensional space). These distributions should be able to describe the relationship between the learning, or performance, parameters. They should provide fast and easy estimates for the prior parameters of a variety of models and easily calculable estimates for a variety of learning properties.

When it is done, posterior probabilities could be then simply derived and would enable us to characterize the ability of individual students, the difficulty of individual curriculum items and the interaction between ability and difficulty with respect to the particular educational task.

## CHAPTER III

## PERFORMANCE MODELS FOR SIMPLE ARITHMETIC PROBLEMS

### III.1 Introduction

In Chapter II, we confirmed the hypothesis that the heterogeneity assumption increases the predictive power of simple learning models and has a sizable effect on their learning properties. In the present chapter, we consider simple performance models for addition problems and propose a method for describing the distribution of different performance rates.

A performance model for simple addition was introduced in Section I.1.2. In Section III.2 we give some basic results relating to the bivariate Dirichlet distribution. In addition, maximum likelihood procedures are suggested for estimating the models' parameters and a Dirichlet distribution is assumed for the performance rates.

Total error statistics are considered in III.3; we derive the conditional and unconditional expectations and variances of the total error statistic.

The empirical data are presented in Section III.4, along with a method for evaluating the exact distribution of item performance rates with homogeneous individuals.

Finally, the discussion and conclusions are presented in III.5.

## III.2 SOME BASIC RESULTS

### III.2.1 The Likelihood Function and Maximum Liklihood Estimates

Let IS denote the internal state of the two-state automaton introduced in Section I.1.2. Then IS = 0 or 1 indicating no carry or carry respectively. We consider three alternatives:

a) digit $i$ is a ones' column digit

b) not (a) and IS = 0

c) not (a) and IS = 1

Let $\bar{c}, \bar{c}\bar{a}$, and $\bar{c}\bar{b}$ denote the probabilities of a correct response to digit $i$ for (a), (b) and (c) respectively. If, in addition, $n_1$, $n_2$, and $n_3$ denote the number of digits under the three alternatives above, then the likelihood of an $n$ digit response is given by

$$L = \bar{c}^{t}(1-\bar{c})^{n_1-t_1}\bar{a}^{t_2}\bar{b}^{t_3}(1-\bar{c}\bar{a})^{n_2-t_2}(1-\bar{c}\bar{b})^{n_3-t_3} \qquad (2.1)$$

where $t_1$, $t_2$ and $t_3$ are the number of correct responses under (a), (b), and (c) respectively and $t = t_1 + t_2 + t_3$.

The maximum likelihood estimates of $\bar{c}$, $\bar{a}$, and $\bar{b}$ were derived by Suppes (1968) and are given by

$$1 - \hat{c} = t_1/n_1$$

$$1 - \hat{a} = \frac{t_2/n_2}{t_1/n_1} \qquad (2.2)$$

$$1 - \hat{b} = \frac{t_3/n_3}{t_1/n_1}$$

The estimates in Eq. (2.2) hold only if the proportion of correct responses to the ones' column digit is greater than the proportion of correct responses to the other digits. The model presented in Eq.'s 2.1 and 2.2 will henceforth be referred to as Performance Model I.

After analyzing the data presented in Tables 4.1 and 4.2 it became clear that the carry parameter, $\overline{a}$, did not contribute to improvement in the prediction of expected total number of errors. Consequently, we tried a two-parameter model by letting $\overline{a} = 1$ this situation is designated as Performance Model II. The predictions for this model (in Table 4.2) improved the error predictions of certain items and had the opposite effect on other error predictions. Since we did not improve the error predictions very much using Performance Model II, the only predictions listed are those for Test 2 in Table 4.2.

It was finally decided to redefine the no-carry state. With this new definition, a transition to a no-carry state is possible only if the automaton was already in a carry state. Equations 2.1 and 2.2 remain the same, but the number of digits $n_1$ and $n_2$ and the number of correct responses $t_1$ and $t_2$ in the corresponding columns are now different. More explicitly, only the problems 639 + 212 and 5267 + 283 have a no-carry column in the third and fourth columns, respectively. This last situation is referred to as Performance Model III.

### III.2.2 The Bivariate Dirichlet Distribution

The bivariate Dirichlet probability density function of two r.v.'s, $p_1$ and $p_2$, is defined by the following equation f

$$f_d(p_1, p_2 \mid \alpha_1, \alpha_2, \alpha_3) = K p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} p_3^{\alpha_3 - 1} \qquad (2.3)$$

where $0 \le P_i \le 1 \quad i = 1,2,3 \,, \quad \sum_{i=1}^{3} p_i = 1$

and $K = B^{-1}(\alpha_1, \alpha_2, \alpha_3) = \dfrac{\Gamma(\alpha_1+\alpha_3+\alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)}$.

The following properties can be established (Silver, 1963):

i) The marginal p.d.f. of a specific $p_i$ is a beta density given by

$$f_\beta(p_j) = \frac{1}{B(\alpha_j, \sum_{i \ne j} \alpha_i)} p_j^{\alpha_j - 1} (1-p_j)^{\sum_{i \ne j} \alpha_i - 1} \qquad 0 \le p_j \le 1 \qquad (2.4)$$

ii) The expected value of $p_j$ is

$$E(p_j) = \frac{\alpha_j}{\sum_{i=1}^{3} \alpha_i} = \frac{\alpha_j}{A} \qquad \text{where} \quad A = \sum_{i=1}^{3} \alpha_i \qquad (2.5)$$

iii) The variance of $p_j$ is

$$V(P_j) = \frac{\alpha_j \sum'_{i \ne j} \alpha_i}{\left(\sum_{i=1}^{3} \alpha_i\right)^2 \left(1 + \sum_{i=1}^{3} \alpha_i\right)} = \frac{\alpha_j \sum_{i \ne j} \alpha_i}{A^2 (A+1)} \qquad (2.6)$$

iv) The covariance of $p_j$ and $p_m$, $j \neq m$, is

$$Cov(P_j, P_m) = - \frac{\alpha_j \alpha_m}{\left(\sum_{i=1}^{3} \alpha_i\right)^2 \left(1 + \sum_{i=1}^{3} \alpha_i\right)} = - \frac{\alpha_j \alpha_m}{A^2(A+1)} \quad (2.7)$$

III.2.3 The Distribution of Item Performance Rates with Homogeneous Individuals

In the following analysis we assume that individual students perform equally well, but the items are heterogeneous and of different difficulty.

Since the error frequencies are very small in performance data, the sum of output and carry error rates is considerably smaller than unity. We may assume, therefore, that the output error rate, $c \equiv 1 - \bar{c}$, and the product $\bar{c}b$ are Dirichlet distributed. For convenience, let $p_1 \equiv c$, $p_2 \equiv \bar{c}b$ and $p_3 \equiv \bar{c}\bar{b}$; $b \equiv 1 - \bar{b}$

With the assumption of a Dirichlet prior on $p_1$ and $p_2$ and using equations 2.4 to 2.7 we have the following properties:

v) The error rates $c$, i.e., $p_1$, and $b$ are independent beta r.v.'s with parameters $(\alpha_1, \alpha_2+\alpha_3)$ and $(\alpha_2, \alpha_3)$ respectively. (2.8)

Conversely, the correct rates $\bar{c}$ and $\bar{b}$ are independent beta r.v.'s with parameters $(\alpha_2+\alpha_3, \alpha_1)$ and $(\alpha_3, \alpha_2)$ respectively.

vi) The carry-output correct rate $(\bar{c}\bar{b})$, i.e., $p_3$, is a beta r.v. with parameters $(\alpha_3, \alpha_1+\alpha_2)$. (2.9)

Conversely, the carry-output error rate $1 - \bar{c}\bar{b}$, i.e., $1 - p_3$, is a beta r.v. with parameters $(\alpha_1+\alpha_2, \alpha_3)$.

Properties (v) and (vi) are intuitively appealing. First, they conform to our model assumption that carry and output errors are independent. Secondly, each difficulty may be indicated by the size of the prior parameters. Thus the output error rate increases as a function of $\alpha_1$, the carry error rate increases as a function of $\alpha_2$, and finally, the carry-output error rate increases as a function of $\alpha_1 + \alpha_2$.

From Eq.'s (2.8) and (2.9) we have the following:

vii) The mean and the variance for the output error rate are

$$E(p_1) = \frac{\alpha_1}{A} \quad \text{and} \quad V(p_1) = \frac{\alpha_1(\alpha_2+\alpha_3)}{A^2(A+1)} = \frac{\alpha_1(A-\alpha_1)}{A^2(A+1)} \tag{2.10}$$

viii) The mean and variance of the carry-output error rate are

$$E(\overline{p}_3) \equiv E(1-p_3) = \frac{\alpha_1 + \alpha_2}{A} = \frac{A - \alpha_3}{A}$$

and (2.11)

$$V(1-p_3) = \frac{(\alpha_1+\alpha_2)\alpha_3}{A^2(A+1)} = \frac{\alpha_3(A-\alpha_3)}{A^2(A+1)}$$

Obviously, $V(\overline{p}_3) = V(p_3)$.

ix) The covariance between output and carry-output errors is

$$\text{Cov}(p_1, 1-p_3) = -\text{Cov}(p_1, p_3) = \frac{\alpha_1\alpha_3}{A^2(A+1)}$$

In order to estimate the prior parameters $\alpha_1, \alpha_2$, and $\alpha_3$ of the Dirichlet distribution we used the method of moments.

It is also possible to integrate the likelihood (2.1) with respect to the Dirchlet prior and determine numerically the estimators $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ which maximize the resulting function.

In Section III.3 we calculate the estimates of the prior means $E(p_1)$ and $E(p_3)$ and the prior variances $v(p_1)$ and $v(p_3)$. Using these estimates we proceed to determine $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ by the following procedure:

$$\hat{\alpha}_1 = \hat{A}\,\hat{E}\,(p_1) \qquad (2.12)$$

$$\hat{\alpha}_2 = \hat{A}(1-\hat{E}(p_1) - \hat{E}(p_3))$$

$$\hat{\alpha}_3 = \hat{A}\,\hat{E}\,(p_3) \qquad (2.13)$$

$A$ itself is determined from (2.10) and (2.11), substituting $A = \alpha_i E(p_i)$ in the expression for the prior variance

$$V(p_i) = \frac{A^2 E(p_i)[1-E(p_i)]}{A^2(A+1)} \qquad (2.14)$$

$$A = \frac{E(p_i)[1-E(p_i)]}{V(p_i)} - 1 \qquad (2.15)$$

As $E(p_i)[1-E(p_i)] \rightarrow v(p_i)$, $A \rightarrow 0$, which implies in return that $\alpha_i \rightarrow 0$.

This situation is noted in Raiffa and Schlaifer (1961, pp. 263-264). In their section "Limiting Behavior of the Prior Distribution", they prove that as the parameters $\alpha_i$ and $A$ both approach zero in such a way that the ratio $\frac{\alpha_i}{A}$

remains fixed, a fraction $E(p_i)$ of the total probability becomes more and more concentrated toward $p_i = 1$, the remainder toward $p_i = 0$; the variance $v(p_i)$ approaches $E(p_i)[1-E(p_i)]$. It is interesting to note that the graph of the beta density with parameters $(r,s)$ is U-shaped when $r + s = 1$. If only one of the parameters is smaller than unity the density concentrates on one side. We also recall that for our learning data, in Chapter II, the parameters were much larger than one and the graph was bell-shaped.

## III.3 TOTAL ERROR STATISTICS

Let $X_1$, $X_2$ and $X_3$ be independently distributed, each having a binomial distribution with parameters $(n_1', q_1)$, $(n_2', q_2)$ and $(n_3', q_3)$ respectively. $n_1'$, $n_2'$ and $n_3'$ are the total output, no carry-output and carry-output digits. Also $q_1 \equiv p_1 \equiv c$ denotes the output error rate, $q_2 \equiv 1-\bar{c}\bar{a}$ denotes the no carry-output error rate, and $q_3 \equiv \bar{p}_3 \equiv 1-\bar{c}\bar{b}$ denotes the carry-output error rate. Then, the r.v. designating the total number of errors in $n_1' + n_2' + n_3'$ digits is $T' = X_1 + X_2 + X_3$. The conditional expectation of $T'$ given $Q \equiv (q_1, q_2, q_3)$ is

$$E(T'|Q) = \sum_{i=1}^{3} n_i' q_i (1-q_i) \qquad (3.1)'$$

and

$$V(T'|Q) = \sum_{i=1}^{3} n_i' q_i (1-q_i) \qquad (3.2)'$$

Let $I_i$ denote the number of <u>items</u> having type i digit (i=1,2,3) and $n_i = n_i'/I_i$, i.e., the average digits per item of type i. Then, the mean total errors per item is

$$E(T|Q) = \sum_{i=1}^{3} n_i q_i \tag{3.1}$$

$$V(T|Q) = \sum_{i=1}^{3} n_i q_i (1-q_i) \tag{3.2}$$

In order to simplify the expressions for the unconditional properties we consider only the situation described by the two-parameter Performance Model II. In this case $T = X_1^* + X_3$ where $X_1^*$ and $X_2$ are binomial random variables with parameters $(n_1^*, q_1)$ and $(n_3, q_3)$ respectively; $n_1^* = n_1 + n_2$.

Then, the conditional mean total errors per item is

$$E(T|q_1 q_3) = n_1^* q_1 + n_3 q_3 \tag{3.3}$$

and the conditional variance is

$$V(T|q_1 q_3) = n_1^* q_1 (1-q_1) + n_3 q_3 (1-q_3) \tag{3.4}$$

The unconditional mean is given by integrating $E(T|q_1 q_3)$ with respect to the Dirichlet prior density of $q_1 \equiv p_1 \equiv c$ and $1 - q_3 \equiv p_3 \equiv \overline{(c\bar{b})}$

$$E^*(T) \equiv E_d[E(T|q_1, q_3)] = n_1^* E(q_1) + n_3 E(q_3) \tag{3.5}$$

The unconditional variance is given by

$$V^*(T) \equiv E^*(T^2) - E^*(T) \doteq E_d[T^2 | q_1, q_3)] - E^{*2}(T)$$

which reduces to

$$V^*(T) = n_1^* E(q_1)[1-E(q_1)] + n_3 E(q_3)[1-E(q_3)]$$

$$+ n_1^*(n_1^*-1)V(q_1) + n_3(n_3-1)\ V(q_3) \qquad (3.6)$$

It is clear that when the output error rate $q_1$ and the carry-output error rate $q_3$ are exact numbers the unconditional variance $V^*(T)$ in (3.6) becomes the conditional variance $V(T|q_1, q_3)$ in (3.4).

## III.4 DATA ANALYSIS

### III.4.1 Description of the Data

The data described in Tables 4.1 and 4.2 were collected as part of the computer-assisted instruction program in elementary mathematics at the Institute for Mathematical Studies in the Social Sciences, Stanford University.

Two row addition problems were given to 80 third graders in local California schools as a pretest before five drill-and-practice sessions; the data for this group are presented in Table 4.1. The same problems were given to a different grou: of 62 third graders after five drill-and-practice sessions; these data are presented in Table 4.2. (Although the groups were not the same, one may infer that some learning has taken place in the second group since there were only 144/62 errors per student in Test 2 and 196/80 errors per student in Test 1.)

The left-hand columns present the observed error frequencies for each item listed and for all students. The no-carry column represents the no carry-output errors for Performance Model I. The numbers in that column are added to the corresponding entries of the output column for Performance Model II. For Performance Model III all entries are added to the corresponding entries of the first column except for the items 639 + 212 and 5267 + 283; there are only 17 no carry-output errors for Model III.

Consider, for example, the data given for the problem 14 + 15 in Table 4.2. For Performance Model I: $n_1 = n_2 = 62$, $n_3 = 0$, $n_1 - t_1 = 3$ and $n_2 - t_2 = 3$; for Performance Model II and III $n_1 = 124$, $n_2 = 0$ and $n_1 - t_1 = 6$. The data given for the problem 639 + 212 in the same table are, for Performance Models I and III: $n_1 = n_2 = n_3 = 62$, $n_1 - t_1 = 3$, $n_2 - t_2 = 8$ and $n_3 - t_3 = 4$; for Performance Model II: $n_1 = 124$, $n_2 = 0$, $n_3 = 62$, $n_1 - t_1 = 11$ and $n_3 - t_3 = 4$.

The predicted values for the three models are calculated by using the maximum likelihood estimates (2.2) and Eq. (3.1) for Performance Models I and III, and Eq. (3.3) for Performance Model II. The maximum likelihood estimates and the variances of total errors due to each error type for all three models are given in Table 4.3. Note that the estimates are for the $q_i$'s; these are simply the ratio of errors per all digits of a given type. For example

$$\hat{q}_3 = (1-\hat{c}\hat{b}) = \left(1- \frac{t_3}{n_3}\right) = \frac{n_3 - t_3}{n_3}$$

the variance estimates in Table 4.3 were calculated using Eq. (3.2) for Models I and III and Eq. (3.4) for Model II.

TABLE 4.1

| | ERRORS | | | | PREDICTED | |
|---|---|---|---|---|---|---|
| item | output | no carry | carry | total | I | III |
| 17 +2 | 5 | 1 | 0 | 6 | 6.7 | 5.5 |
| 14 +15 | 1 | 4 | 0 | 5 | 6.7 | 5.5 |
| 6 +13 | 1 | 2 | 0 | 3 | 6.7 | 5.5 |
| 363 +214 | 1 | 2 | 0 | 3 | 10.1 | 8.3 |
| 416 +212 | 2 | 2 | 0 | 4 | 10.1 | 8.3 |
| 27 +4 | 3 | 0 | 4 | 7 | 10.9 | 10.3 |
| 8 +32 | 4 | 0 | 7 | 11 | 10.9 | 10.3 |
| 66 +14 | 1 | 0 | 3 | 4 | 10.9 | 10.3 |
| 639 +212 | 4 | 8 | 6 | 18 | 14.3 | 18.8 |
| 5267 +283 | 3 | 9 | 18 | 30 | 21.9 | 26.4 |
| 378 +125 | 6 | 0 | 11 | 17 | 18.5 | 17.9 |
| 557 +256 | 6 | 0 | 18 | 24 | 18.5 | 17.9 |
| 3986 +4735 | 3 | 0 | 25 | 28 | 26.1 | 25.5 |
| 7657 +1875 | 7 | 0 | 29 | 36 | 26.1 | 25.5 |
| total | 47 | 28 | 121 | 196 | 196 | 196 |

PERFORMANCE MODEL TEST 1

TABLE 4.2

| | ERRORS | | | | PREDICTED | | |
|---|---|---|---|---|---|---|---|
| item | output | no carry | carry | total | I | II | III |
| 17<br>+ 2 | 2 | 0 | 0 | 2 | 5.0 | 4.8 | 3.6 |
| 14<br>+15 | 3 | 3 | 0 | 6 | 5.0 | 4.8 | 3.6 |
| 6<br>+13 | 0 | 2 | 0 | 2 | 5.0 | 4.8 | 3.6 |
| 363<br>+214 | 0 | 3 | 0 | 3 | 8.0 | 7.1 | 5.4 |
| 416<br>+212 | 1 | 2 | 0 | 3 | 8.0 | 7.1 | 5.4 |
| 27<br>+4 | 5 | 0 | 3 | 8 | 7.6 | 7.9 | 7.4 |
| 8<br>+32 | 1 | 0 | 4 | 5 | 7.6 | 7.9 | 7.4 |
| 66<br>+14 | 0 | 0 | 4 | 4 | 7.6 | 7.9 | 7.4 |
| 639<br>+212 | 3 | 8 | 4 | 15 | 10.6 | 10.3 | 15.8 |
| 5267<br>+283 | 2 | 9 | 21 | 32 | 16.1 | 15.9 | 26.4 |
| 378<br>+125 | 4 | 0 | 8 | 12 | 13.1 | 13.5 | 12.9 |
| 557<br>+256 | 2 | 0 | 9 | 11 | 13.1 | 13.5 | 12.9 |
| 3986<br>+4735 | 2 | 0 | 21 | 23 | 18.6 | 19.1 | 18.5 |
| 7657<br>+1875 | 3 | 0 | 15 | 18 | 18.6 | 19.1 | 18.5 |
| total | 28 | 27 | 89 | 144 | 144 | 144 | 144 |

PERFORMANCE MODEL TEST 2

TABLE 4.3

| Model | output $\hat{q}_1$ | no carry $\hat{q}_2$ | carry $\hat{q}_3$ | output $\hat{V}(X_1)$ | no carry $\hat{V}(X_2)$ | carry $\hat{V}(X_3)$ | Total $\hat{V}(T/Q)$ Eq.(3.2) |
|---|---|---|---|---|---|---|---|
| Test 1 | | | | | | | |
| I | .042 | .038 | .09 | 3.21 | 3.84 | 12.17 | 19.22 |
| III | .034 | .106 | .094 | 3.99 | 7.59 | 12.17 | 23.76 |
| Test 2 | | | | | | | |
| I | .032 | .048 | .09 | 1.93 | 3.6 | 9.0 | 14.6 |
| II | .039 | -- | .09 | 3.77 | -- | 9.0 | 12.78 |
| III | .029 | .137 | .09 | 2.6 | 7.3 | 9.0 | 18.97 |

PERFORMANCE MODELS

Error estimates and total error variances due to each error type

III.4.2 The Distribution of Item Performance Rates with Homogeneous Individuals

Using the estimators $\hat{q}_i$'s, calculated in the last section, we are now able to determine the distribution of the prior performance rates. For Performance Model II this is done by rewriting (3.5) in the following manner:

$$n_1^*(n_1^*-1)V(q_1) + n_3(n_3-1)V(q_3) = V^*(T) - V(T|q_1,q_3) \quad (4.1)$$

We replace $V^*(T)$ by the observed total variance, $\mathring{V}(T)$ and $V(T|q_1,q_3)$ by its estimate (Table 4.3). In order to solve for $V(q_1)$ and $V(q_3)$ we let $\hat{E}(q_i) = \hat{q}_i$ and apply Eq. (2.14). We now have,

$$\mathring{V}(T) - \hat{V}(T|q_1q_3) =$$

$$n_1^*(n_1^*-1)\frac{\hat{E}(q_1)[1-\hat{E}(q_1)]}{\hat{A}+1} + n_3(n_3-1)\frac{\hat{E}(q_3)[1-\hat{E}(q_3)]}{\hat{A}+1} \quad (4.2)$$

The resulting equation solving for $\hat{A}$ is

$$\hat{A} = \frac{n_1^*(n_1^*-1)\hat{E}(q_1)[1-\hat{E}(q_1)]}{\mathring{V}(T) - \hat{V}(T|q_1,q_3)} + \frac{n_3(n_3-1)\hat{E}(q_3)[1-\hat{E}(q_3)]}{\mathring{V}(T) - \hat{V}(T|q_1q_3)} - 1 \quad (4.3)$$

The estimators $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are finally calculated by Eqs. (2.12) and (2.13)

$$\hat{\alpha}_1 = \hat{A}\hat{E}(q_1) = \hat{A}\hat{q}_1$$

$$\hat{\alpha}_3 = \hat{A}[1-\hat{E}(q_3)] = \hat{A}(1-\hat{q}_3)$$

$$\hat{\alpha}_2 = \hat{A}(1-\hat{E}(q_1)-1+\hat{E}(q_3)) = \hat{A}(\hat{E}(q_3)-\hat{E}(q_1)) = \hat{A}(\hat{q}_3-\hat{q}_1) \quad (4.4)$$

As an example, consider the distribution of item performance rates for Test 2 Model II. The observed variance $\mathring{V}(T) = 75.20$ and the estimate of the conditional variance $V(T|q_1,q_3) = 12.78$. Using (4.3), $\hat{A} = 21.04$, $\hat{\alpha}_1 = .820$, $\hat{\alpha}_3 = 19.14$, $\hat{\alpha}_2 = 1.07$, $\hat{V}(q_1) = .007$ and $\hat{V}(q_3) = .037$. That is, there is not too much variance in items due to the output factor, but a noticeable variance due to carry.

## III.5 DISCUSSION AND CONCLUSIONS

### III.5.1 The Conditional Models

The results of this study demonstrate that asymptotic performance data, in the context of computer-assisted instruction in elementary mathematics, can successfully be accounted for by probabilistic automaton models with few parameters.

Educationally more important is the fact that these models serve as excellent tools for determining the structural features of items. There is no doubt that being able to identify these features is a prerequisite if one is to use difficulty factors in order to develop a sound theory of instruction as well as sensible testing procedures.

The main conceptual strength of these models is their ability to provide explicit temporal analysis of the steps being taken by the student in solving a problem. The analysis which led to Performance Model III is a case in point; we were easily able to determine that a no-carry difficulty is raised only if a carry was previously encountered. The advantage provided here in identifying the latent structure of the data seems to be more impressive than the gain provided, say, from a Factor Analytic approach.

From the point of view of an analysis of variance, a second advantage of Performance models is immediate description of the models' adequacy. Let the total error statistics $T$ be a linear combination of the errors due to $n$ variables $X_1$, $X_2$, ..., $X_n$ and an error variable $\epsilon$, i.e., $T = \sum_{i=1}^{n} X_i + \epsilon$. If the $X_i$'s and $\epsilon$ are mutually independent, then

$$E(T) = \sum_{i=1}^{n} E(X_i) + E(\epsilon) \qquad (5.1)$$

and

$$V(T) = \sum_{i=1}^{n} V(X_i) + V(\epsilon) \qquad (5.2)$$

The additivity and independence assumption may now be tested by analyzing the observed discrepancy between $V(T)$ and $\sum_{i=1}^{n} V(X_i)$.

This procedure was actually used in our data, where $X_1$ was the total errors due to output, $X_2$ was the total errors due to carry-output. The magnitude of the observed $V(\epsilon)$ was less than 20 per cent of the total variance, $V(T)$.

III.5.2 The Unconditional Models

One question remains to be answered: can we improve the predictions when item differences are considered? The unconditional predictions, (3.5), depend on the expectations $E(q_1)$ and $E(q_3)$. We can always do as well as the conditional predictions by letting $\hat{E}(q_i) = \hat{q}_i$. However, as long as our estimation procedures are based on the first moment estimates we cannot improve the prediction.

The obvious question now is: can we estimate $E(q_i)$ by some other method. Toward this end, we tried to estimate $\alpha_1$, $\alpha_2$, and $\alpha_3$ by maximizing the unconditional likelihood, i.e., the integral of (2.1) with respect to the Dirichlet prior, and had exactly the same results that we arrived at by the moment estimation procedure. The reason we arrive at the same results using the two estimation methods is due to the fact that the estimators are a function of the mean total errors only.

Theoretically, item differences should have an effect on the predictive power of the models. This is so in view of Eq. (2.14) which can be written as

$$E(q_i)[1-E(q_i)] = (A+1)V(q_i).$$

In other words, unless the variance $V(q_i) = 0$, $E(q_i)$ does depend on item differences, since the $E(q_i)$'s are a function of $V(q_i)$'s.

In general, we may conclude that the properties of the models are sensitive to item differences - to the degree that such differences exist. This fact is demonstrated by noticing the weight attached to the $V(q_i)$'s in the expression for the variance of total errors, $V^*(T)$, in Eq. (3.6).

The aggregate of item differences is expressed as the sum of differences in each performance category. By observing the discrepancy between the conditional variance of total errors for each factor $V(X_i|q_i)$ and the observed variance of total errors due to each factor, we may decide the source of differences between items.

It is true that as $q_i \to 0$ this difference for factor i, e.g., carry, is small. The converse is not true however; $q_i$ may approach unity and $V(q_i)$ may still approach zero. Item differences may, therefore, be viewed as a convex function of correct and incorrect responses summed over performance factors. For either extreme of the function, all responses correct or all responses incorrect, there are no item differences for that factor. We had exactly this situation in mind when we discussed the limiting behavior of the prior distribution in Section III.2.3.

The output performance factor serves as a good example. It has a Beta prior distribution with parameters (.8,20.2). The total probability is concentrated toward $q_1 = 0$. In addition, the discrepancy between the observed output variance and the conditional output variance, $V(X_1|q_1)$ is very small. The same discrepancy between the output-carry variances was ten times as large.

Having the exact prior distribution on performance rates will enable us to derive the posterior distribution of these rates after a new presentation of items. Future extensions may include, therefore, sequential instruction strategies based on Bayesian procedures.

REFERENCES


1. Atkinson, R. C., Bower, G. H. and Crothers, E. J., An Introduction to Mathematical Learning Theory, New York: Wiley, 1965.

2. Atkinson, R. C. and Crothers, E. J., A comparison of paired associate learning models having different acquisition and retention axioms, Journal of Mathematical Psychology, 1964, 1, 285-315.

3. Birnbaum, A., Statistical theory of logistic mental test models with a prior distribution of ability, Journal of Mathematical Psychology, 1969, 6, 258-276.

4. _____ Some latent trait models and their use in inferring an examinee's ability, in F. M. Lord and N. R. Novick (Eds.), Statistical Theories of Mental Test Scores, Reading, Mass.: Addison-Wesley, 1968, Chapters 17-20.

5. Bower, G. H., Application of a model to paired-associate learning, Psychometrika, 1961, 26, 255-280.

6. Bruner, J. S., Some theorems on instruction stated with reference to Mathematics, in E. R. Hilgard (Ed.) Theories of Learning and Theories of Instruction, 63rd NSSE yearbook, Part 1, 1964, pp. 306-335.

7. Bush. R. R. and Mosteller, F., A comparison of eight models, in R. R. Bush and W. K. Estes (Ed.) Studies in Mathematical Learning Theory: Stanford University Press, 1959, pp. 293-307.

8. Erdelyi, A., et al., Higher Transcendental Function, Vol. 1, New York: McGraw Hill Book Co., 1953.

9. Galanter, E., and Bush, R. R., Some T-maze experiments, in R. R. Bush and W. K. Estes (Eds.) Studies in Mathematical Learning Theory, Stanford University Press, 1959, pp. 265-289.

10. Glaser, R., Some implications of previous work on learning and individual differences, in R. E. Gagne (Ed.), Learning and Individual Differences, Columbus, Ohio: Charles E. Merrill, 1967, pp. 1-18.

11. Gregg, L. W and Simon, H. A., Process models and stochastic theories of simple concept formation, Journal of Mathematical Psychology, 1967, 4, 246-276.

12. Groen, G. J. and Atkinson, R. C., Models of optimizing the learning process, Tech Rep. 92, Inst. for Mathematical Studies in the Soc. Sciences, Stanford University, 1965.

13. Hilgard, E. R. (Ed.) Theories of Learning and Theories of Instruction, 63rd NSSE yearbook, Part 1, 1964.

14. Holland, P W., Minimum chi-square procedures; unpublished doctoral dissertation, Stanford University, 1965.

15. Laubsch, J. H., An adaptive teaching system for optimal item allocation, Tech. Rep. 151, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1969.

16. Lindgren, B. W., Statistical Theory, New York, The Macmillan Company, 1962.

17. Matheson, J., Optimum teaching procedures derived from mathematical learning models, Report CC51, Institute in Engineering-Economic Systems, Stanford University, 1964.

18. Murdock, B. B., Jr., Short term retention of single paired-associates, Psychol. Rep., 1961, 8, 280.

19. _____ Short-term memory and paired-associate learning, J. Verb. Learn. and Verb. Behav., 1963, 2, 320-328.

20. Offir, J. D., Adaptive computer tutorial; unpublished manuscript, 1968.

21. Peterson, L. R., Saltzman, Dorothy, Hillner, K., and Land, Vera, Recency and Frequency in paired-associate learning, J. Exp. Psychology, 1962, 63, 396-402.

22. Polya, G., Patterns of Plausible Inference, (Vol. 2 of Mathematics and Plausible Reasoning) Princeton: Princeton University Press, 1954.

23. Raiffa, H. and Schlaifer, R., Applied Statistical Decision Theory, Cambridge, Massachusetts: The MIT Press, 1961.

24. Silberman, H. F., Characteristics of some recent studies of instructional methods, in E. J. Coulson (Ed.) Programmed Learning and Computer-Based Instruction, New York: Wiley, 1962, pp. 13-24.

25. Silver, E., Markovian decision process with uncertain transition probabilities or reward, Tech. Rept. 1, O. R. Center, MIT, 1963.

26. Smallwood, R. D., Quantitative methods in computer-directed teaching system, Final Report, Institute in Engineering-Economic Systems, Stanford University, 1967.

27. Sternberg, S. H., Stochastic learning theory, in R. D. Luce, R. R. Bush and E. Galanter (Eds.) Handbook of Mathematical Psychology, Vol. 2, New York: Wiley, 1963, pp. 1-120.

28. Suppes, P., Stimulus response theory of finite automata, Tech. Rep. 133, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1968.

29. Suppes, P., Hyman, L. and Jerman, M., Linear structural modules for response and latency performance on computer controlled terminals, in J. P. Hill (Ed.) Minnesota Symposia of Child Psychology, Minneapolis: University of Minnesota Press, 1967, pp. 160-200.

## APPENDIX

## THE $\chi^2$ MINIMIZATION PROGRAM

The tables included in the appendix present two or usually three of the best sets of prior parameter estimates for both the OEM* and the LM* for all eight experiments. For each model-experiment combination, 16 in all, each table of the appendix describes also the predicted frequencies of the $O_i$ events based on the first set of the "refined point" listed, which is not necessarily the best set of estimates of r , s , m and n in that table. Also listed are the prior means and variances of c and g associated with this set of estimates.

The numerical computations were written in Fortran IV. The program was adapted to be run on the PDP 10 at the Institute for Mathematical Studies in the Social Sciences, Stanford University.

The program itself consists of two subprograms. The first, named Paraest and written by Tom Wickens[†], is a routine which utilizes general hill-climbing procedures to find the minima of an arbitrary function over a multi-dimensional space. Values of the function are provided by the second subprogram, Stat, which was written by the present author. Stat calculates the $\chi^2$ values (Eq. 3.2) associated with the predicted $O_i$ values which are

[†]Department of Psychology, UCLA

calculated in turn by the equations of Tables 3.3 or 3.4 -- for given values of r, s, m and n. The author provided the range of the search space for each parameter. It was determined from a few pilot runs that the range of 0.5 to 70 was wide enough. The range of the search is tabled under the heading of Minimum and Maximum. The precision was controlled to 0.5, i.e., the worst estimate would be accurate up to 0.5.

Paraest takes over by first calling for function values at points in a rectangular grid over the relevant portion of the parameter space. From this scan a number of points are selected which give the smallest $\chi^2$ values, supplied in turn by Stat. The best points are denoted as Scan Points. A second routine, Refine, works from the previously given estimates of the minimum. Function values are called at points around the estimate, along each of the parameter axes, and from these values the gradient of the function is estimated at the original point, and a "downhill" direction found. Proceeding along the gradient a minimum is approached. The points calculated by the refine procedure are denoted as refine points.

Since Paraest calls on Stat for each new parameter value on the grid from 0.5 to 70 with increments of size 1, Stat is called about $(70)^4$ times by the Scan procedure alone. Each of the 16 $O_i$ predictions calls for a product of two Betas, i.e., 6 products and ratios of the Gamma function.

Thus, the number of subroutine calls for the Scan procedure alone is about 2.5 million. This rough calculation should serve as an indication of the amount of time that was required to run each and every experiment.

Finally we would like to make a remark associated with the results for OEM* experiment Ia. In this appendix there are two tables given for the OEM* experiment Ia. The result of 5.10 for the minimum $\chi^2$ given in the first table seems out of place with respect to the Scan value of 40.766 for almost the same parameter estimates. The same experiment was run, therefore, a few more times and the $\chi^2$ value reached usually as low as 7.15, the value listed in the second table of experiment Ia. The parameter estimated in both tables are almost the same but the value of the prior variance of c went down from .00531 to .00237 as noted in Table 3.7**.

JEM PROBABILITIES EXPERIMENT IA

PROGRAM WILL FIND A MINIMUM.

INPUT DATA:

| | | | | |
|---|---|---|---|---|
| 1.23000E+02 | 3.00000E+00 | 6.00000E+00 | 1.00000E+00 | 1.60000E+01 |
| 3.00000E+00 | 5.00000E+00 | 2.00000E+00 | 4.30000E+01 | 1.00000E+00 |
| 7.00000E+00 | 2.00000E+00 | 1.50000E+01 | 0.00000E-01 | 6.00000E+00 |
| 1.00000E+00 | | | | |

POINT NO.: 1

| DEFN | SCAN POINT | REFINE POINT | MINIMUM | MAXIMUM | PRECISION |
|---|---|---|---|---|---|
| R | 6.25000E+01 | 6.15000E+01 | 5.25E+01 | 7.00E+01 | 5.00E-01 |
| S | 5.85000E+01 | 5.75000E+01 | 5.25E+01 | 6.50E+01 | 5.00E-01 |
| M | 3.35000E+01 | 3.35000E+01 | 2.05E+01 | 3.40E+01 | 5.00E-01 |
| N | 5.95000E+01 | 6.00000E+01 | 4.05E+01 | 6.50E+01 | 5.00E-01 |
| CHI-SQUARE: | 0.71802E+01 | 0.71673E+01 | | | |

POINT NO.: 2

| DEFN | SCAN POINT | REFINE POINT | MINIMUM | MAXIMUM | PRECISION |
|---|---|---|---|---|---|
| R | 6.25000E+01 | 6.18706E+01 | 5.25E+01 | 7.00E+01 | 5.00E-01 |
| S | 5.85000E+01 | 5.78711E+01 | 5.25E+01 | 6.50E+01 | 5.00E-01 |
| M | 3.35000E+01 | 3.40000E+01 | 2.05E+01 | 3.40E+01 | 5.00E-01 |
| N | 6.15000E+01 | 6.14915E+01 | 4.05E+01 | 6.50E+01 | 5.00E-01 |
| CHI-SQUARE. | 0.71823E+01 | 0.71567E+01 | | | |

POINT NO.: 3

| DEFN | SCAN POINT | REFINE POINT | MINIMUM | MAXIMUM | PRECISION |
|---|---|---|---|---|---|
| R | 5.85000E+01 | 5.82500E+01 | 5.25E+01 | 7.00E+01 | 5.00E-01 |
| S | 5.45000E+01 | 5.47500E+01 | 5.25E+01 | 6.50E+01 | 5.00E-01 |
| M | 3.35000E+01 | 3.40000E+01 | 2.05E+01 | 3.40E+01 | 5.00E-01 |
| N | 6.15000E+01 | 6.10000E+01 | 4.05E+01 | 6.50E+01 | 5.00E-01 |
| CHI-SQUARE: | 0.71863E+01 | 0.71675E+01 | | | |

FLOATING CONSTANTS: 0.23400E+03

THE OPTIMAL EXPECTED FREQUENCIES ARE:

O( 1)= 1.26656E+02 O( 2)= 2.73512E+00 O( 3)= 5.52943E+00 O( 4)= 2.51978E+00

O( 5)= 1.39493E+01 O( 6)= 2.51978E+00 O( 7)= 5.13522E+00 O( 8)= 2.39882E+00

O( 9)= 3.96704E+01 O(10)= 2.51978E+00 O(11)= 5.13522E+00 O(12)= 2.39882E+00

O(13)= 1.31444E+01 O(14)= 2.39882E+00 O(15)= 4.92923E+00 O(16)= 2.35982E+00

THE PRIOR MEANS ARE: G=.51681E+00 C=.35829E+00

THE VARIANCES ARE: VG= .20810E-02 VC=.24330E-02

OEM EXPERIMENT Ia

program will find a minimum.

input data:

| | | | | |
|---|---|---|---|---|
| 1.23000e+02 | 3.00000e+00 | 6.00000e+00 | 1.00000e+00 | 1.60000e+01 |
| 3.00000e+00 | 5.00000e+00 | 2.00000e+00 | 4.30000e+01 | 1.00000e+00 |
| 7.00000e+00 | 2.00000e+00 | 1.50000e+01 | 0.00000e-01 | 6.00000e+00 |
| 1.00000e+00 | | | | |

point no.: 1

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 5.40000e+01 | 5.39378e+01 | 5.05e+01 | 5.50e+01 | 5.00e-01 |
| s | 5.35000e+01 | 5.35953e+01 | 4.85e+01 | 5.50e+01 | 5.00e-01 |
| m | 1.55000e+01 | 1.54437e+01 | 1.45e+01 | 2.00e+01 | 5.00e-01 |
| n | 2.75000e+01 | 2.70939e+01 | 2.65e+01 | 3.00e+01 | 5.00e-01 |
| chi-square: | 0.40759e+02 | 0.53174e+01 | | | |

point no.: 2

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 5.50000e+01 | 5.50000e+01 | 5.05e+01 | 5.50e+01 | 5.00e-01 |
| s | 5.35000e+01 | 5.36265e+01 | 4.85e+01 | 5.50e+01 | 5.00e-01 |
| m | 1.55000e+01 | 1.55960e+01 | 1.45e+01 | 2.00e+01 | 5.00e-01 |
| n | 2.75000e+01 | 2.71251e+01 | 2.65e+01 | 3.00e+01 | 5.00e-01 |
| chi-square: | 0.40766e+02 | 0.51044e+01** | | | |

floating constants: 0.23400e+03

the optimal expected frequencies are:

| | | | |
|---|---|---|---|
| o( 1)= 1.25499e+02 | o( 2)= 1.90450e+00 | o( 3)= 5.98323e+00 | o( 4)= 1.86986e+00 |
| o( 5)= 1.42501e+01 | o( 6)= 1.86986e+00 | o( 7)= 5.94860e+00 | o( 8)= 1.90449e+00 |
| o( 9)= 4.06220e+01 | o(10)= 1.86986e+00 | o(11)= 5.94860e+00 | o(12)= 1.90449e+00 |
| o(13)= 1.43716e+01 | o(14)= 1.90449e+00 | o(15)= 6.13716e+00 | o(16)= 2.01230e+00 |

the prior means are: g=.50159e+00 c=.36306e+00

the variances are: vg= .23034e-02 vc=.53114e-02

DEM EXPERIMENT Ib

```
                          program will find a minimum.

 input data:
 1.25000e+02  3.00000e+00  1.00000e+01  4.00000e+00  2.10000e+01
 0.00000e-01  6.00000e+00  3.00000e+00  5.50000e+01  5.00000e+00
 1.00000e+01  2.00000e+00  3.00000e+01  1.00000e+00  6.00000e+00
 7.00000e+00
   point no.:          1
defn            scan point   refine point   minimum   maximum   precision
  r             5.60000e+01  5.56875e+01    5.35e+01  6.50e+01  5.00e-01
  s             6.50000e+01  6.50000e+01    6.25e+01  7.00e+01  5.00e-01
  m             3.10000e+01  3.10000e+01    2.85e+01  4.00e+01  5.00e-01
  n             7.10000e+01  6.97500e+01    6.35e+01  8.00e+01  5.00e-01

 chi-square:    0.20289e+02  0.20263e+02
   point no.:          2
defn            scan point   refine point   minimum   maximum   precision
  r             5.60000e+01  5.47500e+01    5.35e+01  6.50e+01  5.00e-01
  s             6.50000e+01  6.37500e+01    6.25e+01  7.00e+01  5.00e-01
  m             3.10000e+01  2.97500e+01    2.85e+01  4.00e+01  5.00e-01
  n             6.60000e+01  6.72500e+01    6.35e+01  8.00e+01  5.00e-01

 chi-square:    0.20607e+02  0.20291e+02
   point no.:          3
defn            scan point   refine point   minimum   maximum   precision
  r             5.60000e+01  5.41279e+01    5.35e+01  6.50e+01  5.00e-01
  s             6.50000e+01  6.31230e+01    6.25e+01  7.00e+01  5.00e-01
  m             3.60000e+01  3.41238e+01    2.85e+01  4.00e+01  5.00e-01
  n             7.60000e+01  7.66253e+01    6.35e+01  8.00e+01  5.00e-01

 chi-square:    0.20651e+02  0.20210e+02*
 floating constants:  0.28800e+03
 the optimal expected frequencies are:
o( 1)=  1.31815e+02 o( 2)=  3.60557e+00 o( 3)=  6.89969e+00 o( 4)=  4.12511e+00
o( 5)=  1.72090e+01 o( 6)=  4.12511e+00 o( 7)=  7.96038e+00 o( 8)=  4.87553e+00
o( 9)=  4.99261e+01 o(10)=  4.12511e+00 o(11)=  7.96038e+00 o(12)=  4.87553e+00
o(13)=  2.01787e+01 o(14)=  4.87553e+00 o(15)=  9.48991e+00 o(16)=  5.95351e+00

  the prior means are: g=.46142e+00 c=.30769e+00

 the variances are: vg= .20422e-02 vc=.20935e-02
```

OEM EXPERIMENT II

program will find a minimum.

input data:

| | | | | |
|---|---|---|---|---|
| 3.03000e+02 | 1.40000e+01 | 1.90000e+01 | 1.20000e+01 | 5.40000e+01 |
| 1.70000e+01 | 3.20000e+01 | 1.80000e+01 | 1.25000e+02 | 1.50000e+01 |
| 2.50000e+01 | 1.70000e+01 | 6.10000e+01 | 1.90000e+01 | 3.00000e+01 |
| 1.90000e+01 | | | | |

point no.: 1

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 5.00000e+01 | 4.92939e+01 | 3.75e+01 | 5.00e+01 | 5.00e-01 |
| s | 5.50000e+01 | 5.50000e+01 | 4.25e+01 | 5.50e+01 | 5.00e-01 |
| m | 1.40000e+01 | 1.46688e+01 | 1.15e+01 | 2.00e+01 | 5.00e-01 |
| n | 4.50000e+01 | 4.43011e+01 | 3.75e+01 | 5.00e+01 | 5.00e-01 |
| chi-square: | 0.52724e+01 | 0.45188e+01 | | | |

point no.: 2

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 4.50000e+01 | 4.45442e+01 | 3.75e+01 | 5.00e+01 | 5.00e-01 |
| s | 5.00000e+01 | 5.00894e+01 | 4.25e+01 | 5.50e+01 | 5.00e-01 |
| m | 1.40000e+01 | 1.44051e+01 | 1.15e+01 | 2.00e+01 | 5.00e-01 |
| n | 4.50000e+01 | 4.36130e+01 | 3.75e+01 | 5.00e+01 | 5.00e-01 |
| chi-square: | 0.53055e+01 | 0.44951e+01 | | | |

point no.: 3

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 4.50000e+01 | 4.67508e+01 | 3.75e+01 | 5.00e+01 | 5.00e-01 |
| s | 5.50000e+01 | 5.40384e+01 | 4.25e+01 | 5.50e+01 | 5.00e-01 |
| m | 1.40000e+01 | 1.40028e+01 | 1.15e+01 | 2.00e+01 | 5.00e-01 |
| n | 4.00000e+01 | 4.19882e+01 | 3.75e+01 | 5.00e+01 | 5.00e-01 |
| chi-square: | 0.53330e+01 | 0.44541e+01 * | | | |

floating constants: 0.78000e+03

the optimal expected frequencies are:

| | | | |
|---|---|---|---|
| o( 1)= 3.08104e+02 | o( 2)= 1.43289e+01 | o( 3)= 2.36237e+01 | o( 4)= 1.56436e+01 |
| o( 5)= 4.99162e+01 | o( 6)= 1.56436e+01 | o( 7)= 2.59930e+01 | o( 8)= 1.77294e+01 |
| o( 9)= 1.25502e+02 | o(10)= 1.56436e+01 | o(11)= 2.59930e+01 | o(12)= 1.77294e+01 |
| o(13)= 5.58626e+01 | o(14)= 1.77294e+01 | o(15)= 2.96967e+01 | o(16)= 2.08606e+01 |

the prior means are: g=.47264e+00 c=.24875e+00

the variances are: vg= .23672e-02 vc=.31161e-02

OFM EXPERIMENT III

```
                              program will find a minimum.

 input data:
  1.60000e+02  1.30000e+01  1.60000e+01  1.10000e+01  2.40000e+01
  6.00000e+00  1.80000e+01  7.00000e+00  5.70000e+01  9.00000e+00
  2.70000e+01  1.40000e+01  3.30000e+01  2.50000e+01  2.40000e+01
  3.60000e+01
    point no.:          1
 defn            scan point   refine point  minimum   maximum   precision
   r             3.00000e+00  2.68750e+00   5.00e-01  6.00e+01  5.00e-01
   s             3.00000e+00  3.31250e+00   5.00e-01  6.00e+01  5.00e-01
   m             8.00000e+00  8.31250e+00   5.00e-01  6.00e+01  5.00e-01
   n             4.80000e+01  4.67500e+01   5.00e-01  6.00e+01  5.00e-01
 chi square:  0.25432e+02  0.23181e+02

    point no.:          2
 defn            scan point   refine point  minimum   maximum   precision
   r             3.00000e+00  2.68750e+00   5.00e-01  6.00e+01  5.00e-01
   s             3.00000e+00  3.00000e+00   5.00e-01  6.00e+01  5.00e-01
   m             3.00000e+00  3.31250e+00   5.00e-01  6.00e+01  5.00e-01
   n             1.80000e+01  1.67500e+01   5.00e-01  6.00e+01  5.00e-01
 chi square:  0.25699e+02  0.22963e+02

    point no.:          3
 defn            scan point   refine point  minimum   maximum   precision
   r             3.00000e+00  2.68750e+00   5.00e-01  6.00e+01  5.00e-01
   s             3.00000e+00  3.00000e+00   5.00e-01  6.00e+01  5.00e-01
   m             8.00000e+00  8.00000e+00   5.00e-01  6.00e+01  5.00e-01
   n             4.30000e+01  4.42500e+01   5.00e-01  6.00e+01  5.00e-01
 chi square:  0.25951e+02  0.23067e+02
  floating constants:  0.48000e+03
   the optimal expected frequencies are:
 o( 1)=  3.06309e-01 o( 2)=  2.54118e-02 o( 3)=  3.43733e-02 o( 4)=  2.33788e-02
 o( 5)=  5.78657e-02 o( 6)=  2.33788e-02 o( 7)=  3.38593e-02 o( 8)=  3.36813e-02
 o( 9)=  1.32075e-01 o(10)=  2.33788e-02 o(11)=  3.38593e-02 o(12)=  3.36813e-02
 o(13)=  7.15564e-02 o(14)=  3.36813e-02 o(15)=  5.43985e-02 o(16)=  7.91120e-02

   the priormeans are: g=.44792e+00 c=.15096e+00

  variances are: vg= .35327e-01 vc=.22863e-02
```

OPM PROBABILITIES EXPERIMENT IV

program will find a minimum.

input data:

| | | | | |
|---|---|---|---|---|
| 1.17000e+02 | 3.00000e+00 | 1.00000e+01 | 1.00000e+00 | 1.50000e+01 |
| 3.00000e+00 | 9.00000e+00 | 6.00000e+00 | 5.40000e+01 | 7.00000e+00 |
| 9.00000e+00 | 1.00000e+01 | 3.40000e+01 | 8.00000e+00 | 2.20000e+01 |
| 1.20000e+01 | | | | |

point no.: 1

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 1.30000e+01 | 1.26875e+01 | 1.05e+01 | 2.20e+01 | 5.00e-01 |
| s | 2.20000e+01 | 2.07500e+01 | 1.95e+01 | 3.00e+01 | 5.00e-01 |
| m | 2.40000e+01 | 2.40000e+01 | 1.65e+01 | 3.00e+01 | 5.00e-01 |
| n | 6.80000e+01 | 6.76875e+01 | 4.55e+01 | 7.00e+01 | 5.00e-01 |
| chi-square: | 0.12647e+02 | 0.12538e+02 | | | |

point no.: 2

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 1.30000e+01 | 1.33125e+01 | 1.05e+01 | 2.20e+01 | 5.00e-01 |
| s | 2.20000e+01 | 2.16875e+01 | 1.95e+01 | 3.00e+01 | 5.00e-01 |
| m | 1.90000e+01 | 1.90000e+01 | 1.65e+01 | 3.00e+01 | 5.00e-01 |
| n | 5.30000e+01 | 5.33125e+01 | 4.55e+01 | 7.00e+01 | 5.00e-01 |
| chi-square: | 0.12704e+02 | 0.12632e+02 | | | |

point no.: 3

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 1.30000e+01 | 1.33125e+01 | 1.05e+01 | 2.20e+01 | 5.00e-01 |
| s | 2.20000e+01 | 2.16875e+01 | 1.95e+01 | 3.00e+01 | 5.00e-01 |
| m | 2.40000e+01 | 2.27500e+01 | 1.65e+01 | 3.00e+01 | 5.00e-01 |
| n | 6.30000e+01 | 6.33125e+01 | 4.55e+01 | 7.00e+01 | 5.00e-01 |
| chi-square: | 0.12993e+02 | 0.12566e+02 | | | |

floating constants: 0.32000e+03

the optimal expected frequencies are:

| | | | |
|---|---|---|---|
| o( 1)= 1.18415e+02 | o( 2)= 3.46032e+00 | o( 3)= 6.37500e+00 | o( 4)= 5.12421e+00 |
| o( 5)= 1.66283e+01 | o( 6)= 5.12421e+00 | o( 7)= 9.75577e+00 | o( 8)= 8.51696e+00 |
| o( 9)= 5.45882e+01 | o(10)= 5.12421e+00 | o(11)= 9.75577e+00 | o(12)= 8.51696e+00 |
| o(13)= 2.73329e+01 | o(14)= 8.51696e+00 | o(15)= 1.68218e+01 | o(16)= 1.59431e+01 |

the prior means are: g=.37944e+00 c=.26176e+00

the variances are: vg= .68375e-02 vc=.20849e-02

OEM EXPERIMENT Va

program will find a minimum.

input data:

| | | | | |
|---|---|---|---|---|
| 8.20000e+01 | 1.10000e+01 | 1.40000e+01 | 1.30000e+01 | 2.20000e+01 |
| 2.10000e+01 | 2.00000e+01 | 3.10000e+01 | 5.80000e+01 | 1.30000e+01 |
| 3.40000e+01 | 1.80000e+01 | 3.40000e+01 | 2.10000e+01 | 2.60000e+01 |
| 6.20000e+01 | | | | |

point no.: 1

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 6.50000e+00 | 6.50000e+00 | 5.00e-01 | 5.00e+01 | 5.00e-01 |
| s | 1.05000e+01 | 1.05000e+01 | 5.00e-01 | 5.00e+01 | 5.00e-01 |
| m | 2.50000e+00 | 2.50000e+00 | 5.00e-01 | 2.00e+01 | 5.00e-01 |
| n | 2.25000e+01 | 2.15000e+01 | 5.00e-01 | 5.00e+01 | 5.00e-01 |

chi-square: 0.21570e+02 0.21360e+02

point no.: 2

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 6.50000e+00 | 6.50000e+00 | 5.00e-01 | 5.00e+01 | 5.00e-01 |
| s | 1.05000e+01 | 1.02500e+01 | 5.00e-01 | 5.00e+01 | 5.00e-01 |
| m | 2.50000e+00 | 2.25000e+00 | 5.00e-01 | 2.00e+01 | 5.00e-01 |
| n | 1.85000e+01 | 1.95000e+01 | 5.00e-01 | 5.00e+01 | 5.00e-01 |

chi-square: 0.23149e+02 0.21430e+02

point no.: 3

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 1.05000e+01 | 1.07304e+01 | 5.00e-01 | 5.00e+01 | 5.00e-01 |
| s | 1.85000e+01 | 1.76411e+01 | 5.00e-01 | 5.00e+01 | 5.00e-01 |
| m | 2.50000e+00 | 2.47448e+00 | 5.00e-01 | 2.00e+01 | 5.00e-01 |
| n | 1.85000e+01 | 1.99786e+01 | 5.00e-01 | 5.00e+01 | 5.00e-01 |

chi-square: 0.23216e+02 0.21920e+02

floating constants: 0.48000e+03

the optimal expected frequencies are:

| | | | |
|---|---|---|---|
| o( 1)= 8.54230e+01 | o( 2)= 1.18764e+01 | o( 3)= 1.47279e+01 | o( 4)= 1.60681e+01 |
| o( 5)= 2.30276e+01 | o( 6)= 1.60681e+01 | o( 7)= 2.04404e+01 | o( 8)= 2.67802e+01 |
| o( 9)= 4.95864e+01 | o(10)= 1.60681e+01 | o(11)= 2.04404e+01 | o(12)= 2.67802e+01 |
| o(13)= 3.51244e+01 | o(14)= 2.67802e+01 | o(15)= 3.51884e+01 | o(16)= 5.56204e+01 |

the prior means are: g=.38235e+00 c=.10417e+00

the variances are: vg= .13120e-01 vc=.37326e-02

```
oem experiment 7: vc

                          program will find a minimum.

input data:
 1.44000e+02  1.80000e+01  2.30000e+01  9.00000e+00  2.80000e+01
 1.40000e+01  1.20000e+01  1.30000e+01  6.20000e+01  1.40000e+01
 2.50000e+01  1.40000e+01  2.80000e+01  2.00000e+01  2.10000e+01
 3.50000e+01
   point no.:         1
defn           scan point   refine point   minimum   maximum   precision
  r            3.00000e+00  3.00000e+00    5.00e-01  4.00e+00  5.00e-01
  s            3.00000e+00  3.00000e+00    5.00e-01  4.00e+00  5.00e-01
  m            2.00000e+00  2.00000e+00    5.00e-01  5.00e+00  5.00e-01
  n            1.30000e+01  1.27500e+01    9.50e+00  1.40e+01  5.00e-01

chi-square:    0.10002e+02  0.99184e+01

   point no.:         2
defn           scan point   refine point   minimum   maximum   precision
  r            3.00000e+00  3.00000e+00    5.00e-01  4.00e+00  5.00e-01
  s            3.00000e+00  3.00000e+00    5.00e-01  4.00e+00  5.00e-01
  m            2.00000e+00  2.00000e+00    5.00e-01  5.00e+00  5.00e-01
  n            1.20000e+01  1.22500e+01    9.50e+00  1.40e+01  5.00e-01

chi-square:    0.10010e+02  0.99176e+01 x
0floating constants:  0.48000e+03
0 the optimal expected frequencies are:
o( 1)=  1.43899e+02 o( 2)=  1.68473e+01 o( 3)=  2.06982e+01 o( 4)=  1.34779e+01
o( 5)=  2.99663e+01 o( 6)=  1.34779e+01 o( 7)=  1.73287e+01 o( 8)=  1.68473e+01
o( 9)=  5.63100e+01 o(10)=  1.34779e+01 o(11)=  1.73287e+01 o(12)=  1.68473e+01
o(13)=  2.96861e+01 o(14)=  1.68473e+01 o(15)=  2.32654e+01 o(16)=  3.36947e+01

  the prior means are: g=.50000e+00 c=.13559e+00

  the variances are: vg= .35714e-01 vc=.74418e-02
```

```
OPM  EXPERIMENT Ve

                              program will find a minimum.

 input data:
 2.16000e+02  4.00000e+00  1.70000e+01  6.00000e+00  3.40000e+01
 1.60000e+01  1.20000e+01  1.20000e+01  6.60000e+01  4.00000e+00
 1.70000e+01  7.00000e+00  2.90000e+01  8.00000e+00  1.90000e+01
 1.30000e+01

   point no.:          1
 defn          scan point   refine point  minimum   maximum   precision
   r           1.05000e+01  1.05000e+01   5.00e-01  5.00e+01  5.00e-01
   s           1.05000e+01  1.05000e+01   5.00e-01  6.00e+01  5.00e-01
   m           3.00000e+00  3.00000e+00   5.00e-01  3.00e+01  5.00e-01
   n           8.00000e+00  8.00000e+00   5.00e-01  5.00e+01  5.00e-01

 chi-square:   0.17691e+02  0.17691e+02*

   point no.:          2
 defn          scan point   refine point  minimum   maximum   precision
   r           1.45000e+01  1.45000e+01   5.00e-01  5.00e+01  5.00e-01
   s           1.45000e+01  1.45000e+01   5.00e-01  6.00e+01  5.00e-01
   m           3.00000e+00  3.00000e+00   5.00e-01  3.00e+01  5.00e-01
   n           8.00000e+00  8.00000e+00   5.00e-01  5.00e+01  5.00e-01

 chi-square:   0.17737e+02  0.17737e+02

   point no.:          3
 defn          scan point   refine point  minimum   maximum   precision
   r           1.85000e+01  1.82500e+01   5.00e-01  5.00e+01  5.00e-01
   s           1.85000e+01  1.85000e+01   5.00e-01  6.00e+01  5.00e-01
   m           3.00000e+00  3.00000e+00   5.00e-01  3.00e+01  5.00e-01
   n           8.00000e+00  8.00000e+00   5.00e-01  5.00e+01  5.00e-01

 chi-square:   0.17849e+02  0.17832e+02

 floating constants:  0.48000e+03

 the optimal expected frequencies are:
o( 1)=  2.09111e+02 o( 2)=  9.83391e+00 o( 3)=  1.4983 e+01 o( 4)=  9.04720e+00
o( 5)=  2.94016e+01 o( 6)=  9.04720e+00 o( 7)=  1.41966e+01 o( 8)=  9.83391e+00
o( 9)=  7.30380e+01 o(10)=  9.04720e+00 o(11)=  1.41966e+01 o(12)=  9.83391e+00
o(13)=  2.99881e+01 o(14)=  9.83391e+00 o(15)=  1.59641e+01 o(16)=  1.26436e+01

  the prior means are: g=.50000e+00 c=.27273e+00

 the variances are: vg= .11364e-01 vc=.16529e-01
```

LINEAR MODEL EXPERIMENT IA

PROGRAM WILL FIND A MINIMUM.

INPUT DATA:

| | | | | |
|---|---|---|---|---|
| 1.23000E+02 | 3.00000E+00 | 6.00000E+00 | 1.00000E+00 | 1.60000E+01 |
| 3.00000E+00 | 5.00000E+00 | 2.00000E+00 | 4.30000E+01 | 1.00000E+00 |
| 7.00000E+00 | 2.00000E+00 | 1.50000E+01 | 0.00000E-01 | 6.00000E+00 |
| 1.00000E+00 | | | | |

POINT NO.: 1

| DEFN | SCAN POINT | REFINE POINT | MINIMUM | MAXIMUM | PRECISION |
|---|---|---|---|---|---|
| R | 6.50000E+00 | 6.25000E+00 | 5.00E-01 | 6.00E+01 | 5.00E-01 |
| S | 1.45000E+01 | 1.35000E+01 | 5.00E-01 | 6.00E+01 | 5.00E-01 |
| M | 2.50000E+00 | 2.25000E+00 | 5.00E-01 | 3.50E+01 | 5.00E-01 |
| N | 2.50000E+00 | 2.25000E+00 | 5.00E-01 | 5.00E+01 | 5.00E-01 |

CHI-SQUARE: 0.12091E+02 0.11632E+02

POINT NO.: 2

| DEFN | SCAN POINT | REFINE POINT | MINIMUM | MAXIMUM | PRECISION |
|---|---|---|---|---|---|
| R | 1.05000E+01 | 9.32240E+00 | 5.00E-01 | 6.00E+01 | 5.00E-01 |
| S | 2.25000E+01 | 2.23850E+01 | 5.00E-01 | 6.00E+01 | 5.00E-01 |
| M | 2.50000E+00 | 2.09324E+00 | 5.00E-01 | 3.50E+01 | 5.00E-01 |
| N | 2.50000E+00 | 1.96540E+00 | 5.00E-01 | 5.00E+01 | 5.00E-01 |

CHI-SQUARE: 0.12177E+02 0.10717E+02

POINT NO.: 3

| DEFN | SCAN POINT | REFINE POINT | MINIMUM | MAXIMUM | PRECISION |
|---|---|---|---|---|---|
| R | 1.45000E+01 | 1.35913E+01 | 5.00E-01 | 6.00E+01 | 5.00E-01 |
| S | 3.05000E+01 | 3.14316E+01 | 5.00E-01 | 6.00E+01 | 5.00E-01 |
| M | 2.50000E+00 | 1.93143E+00 | 5.00E-01 | 3.50E+01 | 5.00E-01 |
| N | 2.50000E+00 | 1.75687E+00 | 5.00E-01 | 5.00E+01 | 5.00E-01 |

CHI-SQUARE: 0.12250E+02 0.10317E+02

FLOATING CONSTANTS: 0.23400E+03

THE OPTIMAL EXPECTED FREQUENCIES ARE:
O( 1)= 1.12466E+02 O( 2)= 5.17847E+00 O( 3)= 9.11035E+00 O( 4)= 1.60495E+00
O( 5)= 1.31023E+01 O( 6)= 2.49325E+00 O( 7)= 3.83023E+00 O( 8)= 1.23916E+00
O( 9)= 4.43742E+01 O(10)= 4.43785E+00 O(11)= 7.14357E+00 O(12)= 1.92675E+00
O(13)= 1.26921E+01 O(14)= 2.35591E+00 O(15)= 4.15112E+00 O(16)= 1.89316E+00

THE PRIOR MEANS ARE: G=.31646E+00 C=.50000E+00

THE VARIANCES ARE: VG= .10425E-01 VC=.45455E-01

LINEAR MODEL EXPERIMENT IB

PROGRAM WILL FIND A MINIMUM.

INPUT DATA:

```
 1.25000E+02  3.00000E+00  1.00000E+01  4.00000E+00  2.10000E+01
 0.00000E-01  6.00000E+00  3.00000E+00  5.50000E+01  5.00000E+00
 1.00000E+01  2.00000E+00  3.00000E+01  1.00000E+00  6.00000E+00
 7.00000E+00
```

POINT NO.: 1

| DEFN | SCAN POINT | REFINE POINT | MINIMUM | MAXIMUM | PRECISION |
|---|---|---|---|---|---|
| R | 3.00000E+00 | 3.00000E+00 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| S | 2.30000E+01 | 2.17500E+01 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| M | 3.00000E+00 | 3.00000E+00 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| N | 3.00000E+00 | 3.00000E+00 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| CHI-SQUARE: | 0.21814E+02 | 0.21773E+02 | | | |

POINT NO.: 2

| DEFN | SCAN POINT | REFINE POINT | MINIMUM | MAXIMUM | PRECISION |
|---|---|---|---|---|---|
| R | 8.00000E+00 | 8.00000E+00 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| S | 5.80000E+01 | 5.76875E+01 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| M | 3.00000E+00 | 3.00000E+00 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| N | 3.00000E+00 | 3.00000E+00 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| CHI-SQUARE: | 0.21898E+02 | 0.21898E+02 | | | |

POINT NO.: 3

| DEFN | SCAN POINT | REFINE POINT | MINIMUM | MAXIMUM | PRECISION |
|---|---|---|---|---|---|
| R | 3.00000E+00 | 2.68750E+00 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| S | 1.30000E+01 | 1.92500E+01 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| M | 3.00000E+00 | 3.00000E+00 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| N | 3.00000E+00 | 3.00000E+00 | 5.00E-01 | 6.50E+01 | 5.00E-01 |
| CHI-SQUARE: | 0.21927E+02 | 0.21746E+02 * | | | |

FLOATING CONSTANTS: 0.28300E+03

THE OPTIMAL EXPECTED FREQUENCIES ARE:

```
O( 1)=  1.11790E+02 O( 2)=  5.30350E+00 O( 3)=  1.03136E+01 O( 4)=  1.65
982E+00
O( 5)=  2.28152E+01 O( 6)=  2.96531E+00 O( 7)=  5.03925E+00 O( 8)=  1.56
251E+00
O( 9)=  6.54637E+01 O(10)=  6.55977E+00 O(11)=  1.15377E+01 O(12)=  3.04
997E+00
O(13)=  2.27650E+01 O(14)=  5.13243E+00 O(15)=  8.14042E+00 O(16)=  3.89
154E+00
```

THE PRIOR MEANS ARE: G=.12121E+00 C=.50000E+00

THE VARIANCES ARE: VG= .41367E-02 VC=.35714E-01

```
linear model experiment     II
                             program will find a minimum.

 input data:
  3.03000e+02  1.40000e+01  1.90000e+01  1.20000e+01  5.40000e+01
  1.70000e+01  3.20000e+01  1.80000e+01  1.25000e+02  1.50000e+01
  2.50000e+01  1.70000e+01  6.10000e+01  1.90000e+01  3.00000e+01
  1.90000e+01
    point no.:          1
 defn           scan point   refine point   minimum   maximum   precision
  r             3.00000e+00  1.81744e+00    5.00e-01  6.50e+01  5.00e-01
  s             3.00000e+00  1.79041e+00    5.00e-01  6.50e+01  5.00e-01
  m             3.00000e+00  2.46735e+00    5.00e-01  6.00e+01  5.00e-01
  n             8.00000e+00  8.84412e+00    5.00e-01  6.50e+01  5.00e-01
  chi-square:    0.86053e+02  0.68262e+02 *

    point no.:          2
 defn           scan point   refine point   minimum   maximum   precision
  r             3.00000e+00  1.57223e+00    5.00e-01  6.50e+01  5.00e-01
  s             3.00000e+00  1.66394e+00    5.00e-01  6.50e+01  5.00e-01
  m             8.00000e+00  6.56989e+00    5.00e-01  6.00e+01  5.00e-01
  n             2.80000e+01  2.65719e+01    5.00e-01  6.50e+01  5.00e-01
  chi-square:    0.10224e+03  0.87666e+02

    point no.:          3
 defn           scan point   refine point   minimum   maximum   precision
  r             1.30000e+01  1.20425e+01    5.00e-01  6.50e+01  5.00e-01
  s             1.80000e+01  1.63966e+01    5.00e-01  6.50e+01  5.00e-01
  m             3.00000e+00  2.74300e+00    5.00e-01  6.00e+01  5.00e-01
  n             8.00000e+00  7.10884e+00    5.00e-01  6.50e+01  5.00e-01
  chi-square:    0.10830e+03  0.99614e+02

  floating constants:  0.78000e+03
  the optimal expected frequencies are:
o( 1)=  2.54912e+02 o( 2)=  3.68286e+01 o( 3)=  4.91345e+01 o( 4)=  1.38690e+01
o( 5)=  6.88747e+01 o( 6)=  1.84014e+01 o( 7)=  2.41616e+01 o( 8)=  1.11723e+01
o( 9)=  1.05577e+02 o(10)=  2.64311e+01 o(11)=  3.55757e+01 o(12)=  1.56793e+01
o(13)=  5.06643e+01 o(14)=  2.11726e+01 o(15)=  2.79491e+01 o(16)=  1.95963e+01

   the prior means are: g=.50375e+00 c=.21813e+00

  the variances are: vg= .54252e-01 vc=.13853e-01
```

LINEAR MODEL EXPERIMENT III

program will find a minimum.

input data:

| | | | | |
|---|---|---|---|---|
| 1.60000e+02 | 1.30000e+01 | 1.60000e+01 | 1.10000e+01 | 2.40000e+01 |
| 6.00000e+00 | 1.80000e+01 | 7.00000e+00 | 5.70000e+01 | 9.00000e+00 |
| 2.70000e+01 | 1.40000e+01 | 3.30000e+01 | 2.50000e+01 | 2.40000e+01 |
| 3.60000e+01 | | | | |

point no.: 1

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 3.50000e+00 | 1.96468e+00 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| s | 3.50000e+00 | 2.01270e+00 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| m | 3.50000e+00 | 2.06802e+00 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| n | 2.15000e+01 | 1.99581e+01 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| chi-square: | 0.10350e+03 | 0.85085e+02 | | | |

point no.: 2

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 3.50000e+00 | 1.13152e+00 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| s | 3.50000e+00 | 1.14048e+00 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| m | 3.50000e+00 | 2.06454e+00 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| n | 2.75000e+01 | 2.43052e+01 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| chi-square: | 0.11009e+03 | 0.66556e+02* | | | |

point no.: 3

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 3.50000e+00 | 9.61171e-01 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| s | 3.50000e+00 | 1.12603e+00 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| m | 9.50000e+00 | 6.95903e+00 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| n | 6.35000e+01 | 6.30905e+01 | 5.00e-01 | 6.60e+01 | 5.00e-01 |
| chi-square: | 0.11520e+03 | 0.67366e+02 | | | |

floating constants: 0.48000e+03

the optimal expected frequencies are:

| | | | |
|---|---|---|---|
| o( 1)= 9.76817e+01 | o( 2)= 2.68590e+01 | o( 3)= 3.10303e+01 | o( 4)= 1.48277e+01 |
| o( 5)= 3.65672e+01 | o( 6)= 1.73762e+01 | o( 7)= 2.02561e+01 | o( 8)= 1.53097e+01 |
| o( 9)= 4.45329e+01 | o(10)= 2.10997e+01 | o(11)= 2.48966e+01 | o(12)= 1.87470e+01 |
| o(13)= 3.02255e+01 | o(14)= 2.25753e+01 | o(15)= 2.68562e+01 | o(16)= 3.11589e+01 |

the prior means are: g=.49396e+00 c=.93890e-01

the variances are: vg= .50220e-01 vc=.36947e-02

LINEAR MODEL EXPERIMENT IV

program will find a minimum.

input data:

| | | | | |
|---|---|---|---|---|
| 1.17000e+02 | 3.00000e+00 | 1.00000e+01 | 1.00000e+00 | 1.50000e+01 |
| 3.00000e+00 | 9.00000e+00 | 6.00000e+00 | 5.40000e+01 | 7.00000e+00 |
| 9.00000e+00 | 1.00000e+01 | 3.40000e+01 | 8.00000e+00 | 2.20000e+01 |
| 1.20000e+01 | | | | |

point no.: 1

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 3.00000e+00 | 1.75000e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| s | 6.30000e+01 | 6.17500e+01 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| m | 3.00000e+00 | 3.31250e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| n | 3.00000e+00 | 4.25000e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| chi-square: | 0.52954e+02 | 0.42517e+02 | | | |

point no.: 2

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 3.00000e+00 | 1.75000e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| s | 5.80000e+01 | 5.67500e+01 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| m | 3.00000e+00 | 3.31250e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| n | 3.00000e+00 | 4.25000e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| chi-square: | 0.53720e+02 | 0.42483e+02 | | | |

point no.: 3

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 3.00000e+00 | 1.75000e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| s | 5.30000e+01 | 5.17500e+01 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| m | 3.00000e+00 | 3.31250e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| n | 3.00000e+00 | 4.25000e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| chi-square: | 0.54659e+02 | 0.42455e+02 * | | | |

floating constants: 0.32000e+03

the optimal expected frequencies are:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| o( 1)= | 8.62222e+01 | o( 2)= | 6.03138e+00 | o( 3)= | 1.16425e+01 | o( 4)= | 2.17928e+00 |
| o( 5)= | 2.55126e+01 | o( 6)= | 4.07053e+00 | o( 7)= | 7.03065e+00 | o( 8)= | 2.43101e+00 |
| o( 9)= | 7.39288e+01 | o(10)= | 9.90143e+00 | o(11)= | 1.74640e+01 | o(12)= | 5.40485e+00 |
| o(13)= | 3.47323e+01 | o(14)= | 9.51865e+00 | o(15)= | 1.53633e+01 | o(16)= | 8.56657e+00 |

the prior means are: g=.27559e-01 c=.43802e+00

the variances are: vg= .41550e-03 vc=.28748e-01

LINEAR MODEL EXPERIMENT Va

program will find a minimum.

input data:

| | | | | |
|---|---|---|---|---|
| 8.20000e+01 | 1.10000e+01 | 1.40000e+01 | 1.30000e+01 | 2.20000e+01 |
| 2.10000e+01 | 2.00000e+01 | 3.10000e+01 | 5.80000e+01 | 1.30000e+01 |
| 3.40000e+01 | 1.80000c+01 | 3.40000e+01 | 2.10000e+01 | 2.60000e+01 |
| 6.20000e+01 | | | | |

point no.: 1

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 3.00000e+00 | 2.87151e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| s | 3.00000e+00 | 2.92304e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| m | 3.00000e+00 | 1.55779e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| n | 6.30000e+01 | 6.15544e+01 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| chi-square: | 0.73757e+02 | 0.71624e+02 | | | |

point no.: 2

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 3.00000e+00 | 2.84140e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| s | 3.00000e+00 | 2.93644e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| m | 3.00000e+00 | 1.56624e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| n | 5.80000e+01 | 5.65637e+01 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| chi-square: | 0.75069e+02 | 0.70999e+02 | | | |

point no.: 3

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 3.00000e+00 | 2.22292e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| s | 3.00000e+00 | 2.20378e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| m | 3.00000e+00 | 1.28702e+00 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| n | 5.30000e+01 | 5.12866e+01 | 5.00e-01 | 6.50e+01 | 5.00e-01 |
| chi-square: | 0.77067e+02 | 0.65881e+02 | | | |

floating constants: 0.48000e+03

the optimal expected frequencies are:

| | | | |
|---|---|---|---|
| o( 1)= 6.40072e+01 | o( 2)= 2.85775e+01 | o( 3)= 2.98238e+01 | o( 4)= 2.08692e+01 |
| o( 5)= 3.12109e+01 | o( 6)= 2.19369e+01 | o( 7)= 2.30461e+01 | o( 8)= 2.43700e+01 |
| o( 9)= 3.27737e+01 | o(10)= 2.31689e+01 | o(11)= 2.43896e+01 | o(12)= 2.59701e+01 |
| o(13)= 2.57891e+01 | o(14)= 2.76302e+01 | o(15)= 2.93531e+01 | o(16)= 4.70836e+01 |

the prior means are: g=.49555e+00 c=.24683e-01

the variances are: vg= .36791e-01 vc=.37549e-03

LI NEAR MODEL EXPERIMENT Vc

program will find a minimum.

```
input data:
 1.44000e+02  1.80000e+01  2.30000e+01  9.00000e+00  2.80000e+01
 1.40000e+01  1.20000e+01  1.30000e+01  6.20000e+01  1.40000e+01
 2.50000e+01  1.40000e+01  2.80000e+01  2.00000e+01  2.10000e+01
 3.50000e+01
```

point no.: 1

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 1.00000e+00 | 1.00000e+00 | 5.00e-01 | 6.00e+00 | 5.00e-01 |
| s | 1.00000e+00 | 1.00000e+00 | 5.00e-01 | 6.00e+00 | 5.00e-01 |
| m | 1.00000e+00 | 1.00000e+00 | 5.00e-01 | 6.00e+00 | 5.00e-01 |
| n | 5.00000e+00 | 5.00000e+00 | 5.00e-01 | 1.50e+01 | 5.00e-01 |
| chi-square: | 0.85497e+01 | 0.85497e+01 | | | |

point no.: 2

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 1.00000e+00 | 1.00000e+00 | 5.00e-01 | 6.00e+00 | 5.00e-01 |
| s | 1.00000e+00 | 1.00000e+00 | 5.00e-01 | 6.00e+00 | 5.00e-01 |
| m | 1.00000e+00 | 1.00000e+00 | 5.00e-01 | 6.00e+00 | 5.00e-01 |
| n | 6.00000e+00 | 5.75000e+00 | 5.00e-01 | 1.50e+01 | 5.00e-01 |
| chi-square: | 0.10069e+02 | 0.93614e+01 | | | |

point no.: 3

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 1.00000e+00 | 1.00000e+00 | 5.00e-01 | 6.00e+00 | 5.00e-01 |
| s | 2.00000e+00 | 1.75000e+00 | 5.00e-01 | 6.00e+00 | 5.00e-01 |
| m | 1.00000e+00 | 1.00000e+00 | 5.00e-01 | 6.00e+00 | 5.00e-01 |
| n | 3.00000e+00 | 2.75000e+00 | 5.00e-01 | 1.50e+01 | 5.00e-01 |
| chi-square: | 0.11405e+02 | 0.68001e+01 ** | | | |

```
floating constants:  0.48000e+03
 the optimal expected frequencies are:
o( 1)=  1.51964e+02 o( 2)=  2.09504e+01 o( 3)=  2.60009e+01 o( 4)=  9.65567e+00
o( 5)=  3.41039e+01 o( 6)=  1.18701e+01 o( 7)=  1.45974e+01 o( 8)=  1.08572e+01
o( 9)=  4.98104e+01 o(10)=  1.58462e+01 o(11)=  2.01896e+01 o(12)=  1.41538e+01
o(13)=  2.74545e+01 o(14)=  1.80000e+01 o(15)=  2.25455e+01 o(16)=  3.20000e+01

 the prior means are: g=.50000e+00 c=.16667e+00

 the variances are: vg= .83333e-01 vc=.19841e-01
```

linear model experiment Ve
program will find a minimum.

input data:

| | | | | |
|---|---|---|---|---|
| 2.16000e+02 | 4.00000e+00 | 1.70000e+01 | 6.00000e+00 | 3.40000e+01 |
| 1.60000e+01 | 1.20000e+01 | 1.20000e+01 | 6.60000e+01 | 4.00000e+00 |
| 1.70000e+01 | 7.00000e+00 | 2.90000e+01 | 8.00000e+00 | 1.90000e+01 |
| 1.30000e+01 | | | | |

point no.: 1

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 3.00000e+00 | 1.48826e+00 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| s | 3.00000e+00 | 1.36704e+00 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| m | 3.00000e+00 | 2.42083e+00 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| n | 8.00000e+00 | 7.07466e+00 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| chi-square: | 0.72003e+02 | 0.47046e+02 * | | | |

point no.: 2

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 8.00000e+00 | 6.75000e+00 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| s | 8.00000e+00 | 6.75000e+00 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| m | 3.00000e+00 | 2.68750e+00 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| n | 8.00000e+00 | 7.68750e+00 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| chi-square: | 0.95963e+02 | 0.89024e+02 | | | |

point no.: 3

| defn | scan point | refine point | minimum | maximum | precision |
|---|---|---|---|---|---|
| r | 3.00000e+00 | 2.90184e+00 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| s | 3.00000e+00 | 2.23059e+00 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| m | 8.00000e+00 | 6.92337e+00 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| n | 2.80000e+01 | 2.80307e+01 | 5.00e-01 | 6.00e+01 | 5.00e-01 |
| chi-square: | 0.98463e+02 | 0.78800e+02 | | | |

floating constants: 0.48000e+03

the optimal expected frequencies are:

| | | | |
|---|---|---|---|
| o( 1)= 1.86020e+02 | o( 2)= 2.02038e+01 | o( 3)= 2.78737e+01 | o( 4)= 6.79735e+00 |
| o( 5)= 4.07695e+01 | o( 6)= 9.27299e+00 | o( 7)= 1.25155e+01 | o( 8)= 5.32513e+00 |
| o( 9)= 6.64048e+01 | o(10)= 1.38969e+01 | o(11)= 1.93185e+01 | o(12)= 7.75598e+00 |
| o(13)= 2.86660e+01 | o(14)= 1.07777e+01 | o(15)= 1.45899e+01 | o(16)= 9.81199e+00 |

the prior means are: g=.52123e+00 c=.25495e+00

the variances are: vg= .64729e-01 vc=.18098e-01

(Continued from inside front cover)

96 R. C. Atkinson, J. W. Breisford, and R. M. Shiffrin. Multi-process models for memory with applications to a continuous presentation task. April 13, 1966. (J. math. Psychol., 1967, 4, 277-300).
97 P. Suppes and E. Crothers. Some remarks on stimulus-response theories of language learning. June 12, 1966.
98 R. Bjork. All-or-none subprocesses in the learning of complex sequences. (J. math. Psychol., 1968, 1, 182-195).
99 E. Gammon. The statistical determination of linguistic units. July 1, 1966.
100 P. Suppes, L. Hyman, and M. Jerman. Linear structural models for response and latency performance in arithmetic. In J. P. Hill (ed.), Minnesota Symposia on Child Psychology. Minneapolis, Minn.: 1967. Pp. 160-200).
101 J. L. Young. Effects of intervals between reinforcements and test trials in paired-associate learning. August 1, 1966.
102 H. A. Wilson. An investigation of linguistic unit size in memory processes. August 3, 1966.
103 J. T. Townsend. Choice behavior in a cued-recognition task. August 8, 1966.
104 W. H. Batchelder. A mathematical analysis of multi-level verbal learning. August 9, 1966.
105 H. A. Taylor. The observing response in a cued psychophysical task. August 10, 1966.
106 R. A. Bjork. Learning and short-term retention of paired associates in relation to specific sequences of interpresentation intervals. August 11, 1966.
107 R. C. Atkinson and R. M. Shiffrin. Some Two-process models for memory. September 30, 1966.
108 P. Suppes and C. Ihrke. Accelerated program in elementary-school mathematics--the third year. January 30, 1967.
109 P. Suppes and I. Rosenthal-Hill. Concept formation by kindergarten children in a card-sorting task. February 27, 1967.
110 R. C. Atkinson and R. M. Shiffrin. Human memory: a proposed system and its control processes. March 21, 1967.
111 Theodore S. Rodgers. Linguistic considerations in the design of the Stanford computer-based curriculum in initial reading. June 1, 1967.
112 Jack M. Knutson. Spelling drills using a computer-assisted instructional system. June 30, 1967.
113 R. C. Atkinson. Instruction in initial reading under computer control: the Stanford Project. July 14, 1967.
114 J. W. Breisford, Jr. and R. C. Atkinson. Recall of paired-associates as a function of overt and covert rehearsal procedures. July 21, 1967.
115 J. H. Stelzer. Some results concerning subjective probability structures with semiorders. August 1, 1967
116 D. E. Rumelhart. The effects of interpresentation intervals on performance in a continuous paired-associate task. August 11, 1967.
117 E. J. Fishman, L. Keller, and R. E. Atkinson. Massed vs. distributed practice in computerized spelling drills. August 18, 1967.
118 G. J. Groen. An investigation of some counting algorithms for simple addition problems. August 21, 1967.
119 H. A. Wilson and R. C. Atkinson. Computer-based instruction in initial reading: a progress report on the Stanford Project. August 25, 1967.
120 F. S. Roberts and P. Suppes. Some problems in the geometry of visual perception. August 31, 1967. (Synthese, 1967, 17, 173-201)
121 D. Jamison. Bayesian decisions under total and partial ignorance. D. Jamison and J. Kozielecki. Subjective probabilities under total uncertainty. September 4, 1967.
122 R. C. Atkinson. Computerized instruction and the learning process. September 15, 1967.
123 W. K. Estes. Outline of a theory of punishment. October 1, 1967.
124 T. S. Rodgers. Measuring vocabulary difficulty: An analysis of item variables in learning Russian-English and Japanese-English vocabulary parts. December 18, 1967.
125 W. K. Estes. Reinforcement in human learning. December 20, 1967.
126 G. L. Wolford, D. L. Wessel, W. K. Estes. Further evidence concerning scanning and sampling assumptions of visual detection models. January 31, 1968.
127 R. C. Atkinson and R. M. Shiffrin. Some speculations on storage and retrieval processes in long-term memory. February 2, 1968.
128 John Holmgren. Visual detection with imperfect recognition. March 29, 1968.
129 Lucille B. Mlodnosky. The Frostig and the Bender Gestalt as predictors of reading achievement. April 12, 1968.
130 P. Suppes. Some theoretical models for mathematics learning. April 15, 1968. (Journal of Research and Development in Education, 1967, 1, 5-22)
131 G. M. Olson. Learning and retention in a continuous recognition task. May 15, 1968.
132 Ruth Norene Hartley. An investigation of list types and cues to facilitate initial reading vocabulary acquisition. May 29, 1968.
133 P. Suppes. Stimulus-response theory of finite automata. June 19, 1968.
134 N. Moler and P. Suppes. Quantifier-free axioms for constructive plane geometry. June 20, 1968. (In J. C. H. Gerretsen and F. Oort (Eds.), Compositio Mathematica. Vol. 20. Groningen, The Netherlands: Wolters-Noordhoff, 1968. Pp. 143-152.)
135 W. K. Estes and D. P. Horst. Latency as a function of number of response alternatives in paired-associate learning. July 1, 1968.
136 M. Schlag-Rey and P. Suppes. High-order dimensions in concept identification. July 2, 1968. (Psychom. Sci., 1968, 11, 141-142)
137 R. M. Shiffrin. Search and retrieval processes in long-term memory. August 15, 1968.
138 R. D. Freund, G. R. Loftus, and R.C. Atkinson. Applications of multiprocess models for memory to continuous recognition tasks. December 18, 1968.
139 R. C. Atkinson. Information delay in human learning. December 18, 1968.
140 R. C. Atkinson, J. E. Holmgren, and J. F. Juola. Processing time as influenced by the number of elements in the visual display. March 14, 1969.
141 P. Suppes, E. F. Loftus, and M. Jerman. Problem-solving on a computer-based teletype. March 25, 1969.
142 P. Suppes and Mona Morningstar. Evaluation of three computer-assisted instruction programs. May 2, 1969.
143 P. Suppes. On the problems of using mathematics in the development of the social sciences. May 12, 1969.
144 Z. Domotor. Probabilistic relational structures and their applications. May 14, 1969.
145 R. C. Atkinson and T. D. Wickens. Human memory and the concept of reinforcement. May 20, 1969.
146 R. J. Titiev. Some model-theoretic results in measurement theory. May 22, 1969.
147 P. Suppes. Measurement: Problems of theory and application. June 12, 1969.
148 P. Suppes and C. Ihrke. Accelerated program in elementary-school mathematics--the fourth year. August 7, 1969.
149 D. Rundus and R.C. Atkinson. Rehearsal in free recall: A procedure for direct observation. August 12, 1969.
150 P. Suppes and S. Feldman. Young children's comprehension of logical connectives. October 15, 1969.

( Continued on back cover )

( Continued from inside back cover )

151 Joaquim H. Laubsch. An adaptive teaching system for optimal item allocation. November 14, 1969.
152 Roberta L. Klatzky and Richard C. Atkinson. Memory scans based on alternative test stimulus representations. November 25, 1969.
153 John E. Holmgren. Response latency as an indicant of information processing in visual search tasks. March 16, 1970.
154 Patrick Suppes. Probabilistic grammars for natural languages. May 15, 1970.
155 E. Gammon. A syntactical analysis of some first-grade readers. June 22, 1970.
156 Kenneth N. Wexler. An automaton analysis of the learning of a nimiature system of Japanese. July 24, 1970.
157 R. C. Atkinson and J.A. Paulson. An approach to the psychology of instruction. August 14, 1970.
158 R.C. Atkinson, J.D. Fletcher, H.C. Chetin, and C.M. Stauffer. Instruction in initial reading under computer control: the Stanford project. August 13, 1970.
159 Dewey J. Rundus. An analysis of rehearsal processes in free recall. August 21, 1970.
160 R.L. Klatzky, J.F. Juola, and R.C. Atkinson. Test stimulus representation and experimental context effects in memory scanning.
161 William A. Rottmayer. A formal theory of perception. November 13, 1970.
162 Elizabeth Jane Fishman Loftus. An analysis of the structural variables that determine problem-solving difficulty on a computer-based teletype. December 18, 1970.
163 Joseph A. Van Campen. Towards the automatic generation of programmed foreign-language instructional materials. January 11, 1971.
164 Jamesine Friend and R.C. Atkinson. Computer-assisted instruction in programming: AID. January 25, 1971.
165 Lawrence James Hubert. A formal model for the perceptual processing of geometric configurations. February 19, 1971.
166 J. F. Juola, I.S. Fischler, C.T.Wood, and R.C. Atkinson. Recognition time for information stored in long-term memory.
167 R.L. Klatzky and R.C. Atkinson. Specialization of the cerebral hemispheres in scanning for information in short-term memory.
168 J.D. Fletcher and R.C. Atkinson. An evaluation of the Stanford CAI program in initial reading ( grades K through 3 ). March 12, 1971.
169 James F. Juola and R.C. Atkinson. Memory scanning for words versus categories.
170 Ira S. Fischler and James F. Juola. Effects of repeated tests on recognition time for information in long-term memory.
171 Patrick Suppes. Semantics of context-free fragments of natural languages. March 30, 1971.
172 Jamesine Friend. Instruct coders' manual. May 1, 1971.
173 R.C. Atkinson and R.M. Shiffrin. The control processes of short-term memory. April 19, 1971.
174 Patrick Suppes. Computer-assisted instruction at Stanford. May 19, 1971.
175 D. Jamison, J.D. Fletcher, P. Suppes and R.C.Atkinson. Cost and performance of computer-assisted instruction for compensatory education.
176 Joseph Offir. Some mathematical models of individual differences in learning and performance. June 28, 1971.