Program Abstracts/Algorithms

A multinomial maximum likelihood program (MUNOML)

RENWICK E. CURRY

Man-Vehicle Laboratory

*Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139*

In our research on modeling sensory and decision phenomena, we were soon confronted with the task of evaluating both old and new models using both old and new data. Rather than design an ad-hoc estimation program for each new model, as is typically done, we developed an "executive" program that provides a general method for estimating parameters and simultaneously provides flexibility for accommodating new models with a minimum amount of programming. Our experience with canned computer programs has been equivocal, so we decided to provide only the general framework and let the user accomplish the objective of estimating parameters for his particular model by writing a new subroutine within the constraints of the executive program.

The most common class of distributions for which parameters must be extracted are multinomial distributions resulting from a stimulus-response classification, e.g., binary responses (YES-NO or two alternative forced choice methods), the method of successive categories (rating scales), or transition probabilities in a Markov chain. Although a number of methods exist for estimating such parameters (Restle, 1971) we have chosen the maximum likelihood method and have implemented the scoring of Rao to adjust the parameters from one iteration to the next. We have chosen the maximum likelihood (ML) method because (1) it is a member of the class of consistent asymptotically normal estimators (CAN) under suitable conditions (Rao, 1973); (2) it will easily handle situations in which all the responses fall into one category; (3) there are many situations in which the maximum likelihood estimator can be shown to yield unique estimates for parameters (other estimation techniques may have this property as well); and (4) it is the only one exhibiting first order efficiency (Rao, 1973).

Other authors have successfully used the maximum likelihood method with Rao's scoring method in signal detection type models (e.g., Dorfman and Alf, 1968; Dorfman and Alf, 1969; Dorfman, Beavers, & Saslow, 1973; Grey & Morgan, 1972; Ogilvie & Creelman, 1968). This program differs primarily in its ability to estimate parameters for a wide variety of models.

A BRIEF REVIEW OF MAXIMUM LIKELIHOOD ESTIMATION

Necessary Conditions for a Maximum

In this section, we summarize the pertinent aspects of the maximum likelihood (ML) estimators pertaining to the practical considerations of carrying out the estimation process. [See the excellent descriptions by Rao (1973) and Kendall & Stuart (1967) for a more detailed discussion of ML estimators.] We will formulate the problem as it is usually encountered in behavioral research, that is, in the form of a multinomial distribution containing parameters to be estimated. We specify by P_{ij} the theoretical conditional probability of observed response

Sponsored by NASA Grant NGR 22-009-733, Man-Machine Integration Branch, NASA-Ames Research Center.

$j(j = 1, \dots, n)$ given stimulus $i(i = 1, \dots, m)$. Thus, we have a series of multinomial distributions (multimultinomial), and it is assumed that the responses in each of these distributions are independent. Under these conditions, the likelihood function for the multinomial distribution is proportional to $L(x)$ (Rao, 1973).

$$L(x) = \prod_{i=1}^m \prod_{j=1}^n P_{ij}^{r_{ij}}(x), \quad (1)$$

where x is the ℓ -dimensional parameter vector to be estimated and r_{ij} is the number of "j" responses observed when the stimulus is specified at condition "i."

As is usual, we will consider the natural logarithm of the likelihood function, since it is easier to deal with numerically and analytically, and its maximum occurs at the same point as that of the likelihood function.

$$\ell_n L(x) = \ell(x) = \sum_{i=1}^m \sum_{j=1}^n r_{ij} \ell_n P_{ij}(x). \quad (2)$$

The first order necessary conditions for an extremum point of the likelihood function are given by the (components of the) vector equation:

$$\frac{\partial \ell}{\partial x} = \sum_{i=1}^m \sum_{j=1}^n r_{ij} \frac{1}{P_{ij}(x)} \frac{\partial P_{ij}}{\partial x} = 0. \quad (3)$$

Whether any extremum point satisfying (3) is a local maximum, minimum, or saddle point is found by examining the eigenvalues of the matrix of second partial derivatives or the Hessian matrix

$$\begin{aligned} \frac{\partial^2 \ell}{\partial x^2} = & - \sum_{i=1}^m \sum_{j=1}^n \frac{r_{ij}}{P_{ij}^2(x)} \left[\frac{\partial P_{ij}}{\partial x} \right] \left[\frac{\partial P_{ij}}{\partial x} \right]^T \\ & + \sum_{i=1}^m \sum_{j=1}^n \frac{r_{ij}}{P_{ij}(x)} \frac{\partial^2 P_{ij}}{\partial x^2}. \end{aligned} \quad (4)$$

Finding the Maximum

Since the necessary conditions for an extremum (3) are rarely, if ever, analytically tractable, an iterative approach is called for. The general form of the parameter updating technique usually takes the form

$$x_{i+1} = x_i + A_i^{-1} \frac{\partial \ell}{\partial x}, \quad (5)$$

where x_i is the parameter estimate at the i th iteration, $\partial \ell / \partial x$ is the gradient evaluated at the point x_i , and the matrix A_i determines the algorithm by which one seeks the maximum of the likelihood function. One approach, Rao's scoring method, is to construct the matrix A from an approximation to the information matrix

$$A = \sum_{i=1}^m \sum_{j=1}^n \left[\frac{r_{ij}}{P_{ij}^2(x)} \right] \left[\frac{\partial P_{ij}}{\partial x} \right] \left[\frac{\partial P_{ij}}{\partial x} \right]^T. \quad (6)$$

Note that the matrix A_i is the negative of the Hessian matrix (4) with the last set of terms omitted. An advantage of the method of scoring is that the matrix A_i can be calculated from the gradient itself, and the second partial derivatives of P_{ij} are not needed. When inverted, as it must be for the iterations, the inverse becomes the Cramer-Rao lower bound covariance matrix of estimation errors in the parameter x . See Curry (Note 1) for a more detailed discussion of these points.

MUNOML: The Executive Search Program

As can be seen from Equations (3), (5), and (6), the only quantities needed to carry out the search function are the $m \times n$ quantities P_{ij} , and the $m \times n \times \ell$ quantities $\partial P_{ij} / \partial x_k$. We have written MUNOML to carry out the search operation using only the numerical values of the probabilities and the partial derivatives. These numerical values are provided by a user supplied subroutine which is called once in each iteration cycle. This provides an extremely flexible tool for estimating parameters of grouped data, i.e., for multinomial distributions. Thus, it will accommodate any behavioral model in which the stimuli and responses are drawn from discrete element sets as long as $\partial P_{ij} / \partial x_k$ is continuous. The theoretical form of the underlying distributions within the model are contained in the subroutine. The range of distributions and models is almost limitless: One may treat continuous sensory models (signal detection theory); Thurstonian scaling; discrete sensory models (rectilinear ROC curves); predictions of judgments and choice by the method of successive categories (more Thurstonian scaling). The distributions may belong to the same location/scale family (Gaussian, logistic, arc sine, etc.) or they may involve imbedded parameters (e.g., the beta distribution).

In addition to the maximization of the likelihood function with respect to the parameters, we have included a χ^2 goodness-of-fit test for the final parameter estimates, given by

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{(r_{ij} - r_i P_{ij})^2}{r_i P_{ij}}, \text{ where } r_i = \sum_{j=1}^n r_{ij}; \quad (7)$$

$$df = m(n-1) - \ell$$

The degrees of freedom are obtained by finding the number of independent measurements and subtracting the number of parameters being estimated. We also have included another χ^2 computation in which all cells for which the expected number of responses is less than four are pooled with adjacent cells and the degrees of freedom are reduced accordingly.

MODELS IN SIGNAL DETECTION AND RECOGNITION**Psychological Continuum (Thurstonian) Models**

Rating scales—The method of successive categories: In the method of successive categories (Bock & Jones, 1968), or the rating method in psychophysics (Green & Swets, 1966), the observed measures are P_{ij} , the probability of eliciting response R_j , given stimulus S_i . These sensory-continuum models assume that each presentation of a stimulus maps to a point on an internal continuum, and these same probabilities, as predicted by the models, take the form

$$P_{ij} = P(R_j/S_i) = P(c_j \leq u < c_{j+1}/S_i)$$

$$= P\left(\frac{c_j - \mu_i}{\sigma_i} \leq \frac{u - \mu_i}{\sigma_i} < \frac{c_{j+1} - \mu_i}{\sigma_i}\right)$$

$$= F(Z_{i,j+1}) - F(Z_{i,j})$$

$$i = 1, 2, \dots, m \quad j = 1, 2, \dots, n. \quad (8)$$

$F(\cdot)$ is the theoretical distribution function; and the unit deviates are

$$Z_{i,j} = \frac{c_j - \mu_i}{\sigma_i} = b_i c_j - a_i. \quad (9)$$

Here c_j is the j th criterion level; σ_i and μ_i are the mean and standard deviation of the discriminial dispersion of the i th

stimulus; and $b_i = 1/\sigma_i$ and $a_i = \mu_i/\sigma_i$ are an alternative set of parameters introduced for convenience.

To remove the ambiguity of the origin and scale factor of the distributions, we assign

$$b_1 \equiv 1, \text{ which fixes scale}$$

$$a_1 \equiv 0, \text{ which fixes origin.} \quad (10)$$

Equations (10) are particularly useful when using rating scale methods in psychophysics, since one can arbitrarily assign the noise process to stimulus S_1 , and all distributions are scaled with noise standard deviation so that b_i becomes the ratio of noise standard deviation to signal standard deviation. In the equal-variance case, a_i is the normalized separation between "noise" and "signal" distributions, often called d'_i .

Multiple SNR pairs with equal variance: One of the new models that we have investigated is for decision behavior at multiple signal-to-noise ratios (SNR) in which the experimental conditions were such that an equal-variance Gaussian model is apropos. Rather than work with criterion cut-off points, c_j it is more meaningful to deal with the log likelihood ratio which, for equal-variance Gaussian distributions, can be shown to be

$$\lambda(u) = d'_i [b_i u - \frac{1}{2}(a_i + a_{i+1})] \quad i = 1, 3, \dots, m-1. \quad (11)$$

In the above expression, we have used the convention that the stimuli are ordered by equal SNR pairs, i.e., S_1 and S_2 have the same SNR, S_3 and S_4 are at another SNR, etc. When this expression for the log likelihood ratio is substituted into Equation (9) for the deviates with $u = c_{ij}$, we obtain

$$Z_{i+1,j} = \lambda_{i,j}/d'_i + d'_i/2 \quad i = 1, 3, \dots, m-1$$

$$Z_{i,j} = \lambda_{i,j}/d'_i - d'_i/2 \quad j = 2, 3, \dots, n. \quad (12)$$

If one assumes a constant likelihood decision rule, then $\lambda_{i,j} = \lambda_j$ for all i .

Discrete Sensory Models

The discrete sensory models lead to the prediction of rectilinear ROC curves (e.g., see Krantz, 1969 & Luce, 1963). These may be expressed by

$$P_H = p_0 + \sum_{i=1}^N \Delta p_i S(c - i) \quad P_{FA} = q_0 + \sum_{i=1}^N \Delta q_i S(c - i)$$

$$(13)$$

with the constraint

$$p_0 + \sum_{i=1}^N \Delta p_i = q_0 + \sum_{i=1}^N \Delta q_i = 1,$$

where N is the number of limbs in the ROC; P_H and P_{FA} are the theoretical values for the probability of a hit and false alarm; p_0 , Δp_i , q_0 , Δq_i describe the form of the ROC curve; c is the criterion level; and $S(\cdot)$ is a saturation function.

$$S(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 \leq x \leq 1. \\ 1 & 1 \leq x \end{cases} \quad (14)$$

Equations (13) and (14) are suitable for use with the method of scoring as an iterative technique since the gradient is reasonably

well behaved, although only piecewise continuous with respect to c .

SUMMARY

We have found MUNOML to be extremely useful in day-to-day operation since it provides the maximum amount of flexibility with a minimum amount of duplicated effort. This is especially true when minor changes to a model's structure are made because there will be a great deal of unchanged code within the user supplied subroutine.

MUNOML has excellent convergence qualities and rarely goes beyond 10 iterations; when it does, it is usually because of programming errors in the calculation of the probabilities or their derivatives.

We are collecting a library of subroutines for use with MUNOML. There are currently subroutines for calculating parameters for a successive categories model and for models based on the theory of signal detectability: unequal variance ROC parameters from binary responses; equal-variance signal-to-noise ratio pairs; constant likelihood ratio decision rules; and Neyman-Pearson decision rules (using objective and subjective probabilities). Complete listings and/or card decks (about 500 cards in all) are available on request. Operational details are described in Curry (Note 1).

REFERENCE NOTE

1. Curry, R. E. MUNOML: A multinomial maximum likelihood program for behavioral research. Man-Vehicle

Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1974.

REFERENCES

- Bock, R. D., & Jones, L. V., *The measurement and prediction of judgement and choice*. San Francisco: Holden-Day, 1968.
- Dorfman, D. D., & Alf, E., Jr. Maximum-likelihood estimation of parameters of signal-detection theory—a direct solution. *Psychometrika*, 33, 117-124, 1968.
- Dorfman, D. D., & Alf, E., Jr. Maximum-likelihood estimation of parameters of signal detection theory and determination of confidence intervals—rating method data. *Journal of Mathematical Psychology*, 6, 487-496, 1969.
- Dorfman, D. D., Beavers, L. L., & Saslow, C. Estimation of signal detection theory parameters from rating-method data: A comparison of the method of scoring and direct search. *Bulletin of the Psychonomic Society*, 1, 207-208, 1973.
- Green, D., & Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- Grey, D. R., & Morgan, B. J. T. Some aspects of ROC curve fitting: Normal and logistic models. *Journal of Mathematical Psychology*, 9, 128-139, 1972.
- Kendall, M. G., & Stuart, A. *The advanced theory of statistics* (4th ed., Vol. II). London: Griffin, 1967.
- Krantz, D. H. Threshold theory of signal detection. *Psychology Review*, 76, 308-324, 1969.
- Luce, R. D. A threshold theory for simple detection experiments. *Psychology Review*, 70, 61-79, 1963.
- Ogilvie, J. C., & Creelman, C. D. Maximum likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology*, 5, 377-391, 1968.
- Rao, C. R. *Linear statistical inference and its applications* (2nd ed.). New York: Wiley, 1973.
- Restle, F. *Mathematical psychology an introduction*. London: Penguin, 1971.

Behavior Research Methods & Instrumentation
1975, Vol. 7 (3), 307

Computer-generated computation exercises

PAUL G. SALMON and ROBERT J. TRACY
DePaul University, Chicago, Illinois 60614

Four programs were designed to aid instructors in developing students' statistical skills by generating exercises on (1) z scores, (2) Pearson r correlation with regression or matched groups t test, (3) independent groups t test, and (4) chi square. The programs generate exercises by drawing random samples from populations, a procedure which produces a unique exercise for each student and thus discourages plagiarism. The exercises generated consist of data values plus instructions and are given to students, who do the required calculations either with paper and pencil or electronic calculators and submit them for grading.

For each student exercise that is printed, a corresponding solution is produced for the instructor's use in checking and grading. The solutions contain not only the final answer to problems, but also intermediate summary statistics so that students' computation errors may be located and partial credit given. The printed output of the programs is designed so that the solutions can be conveniently separated from the exercises and the exercises easily separated from one another for distributing to students.

Designed for the convenience of the instructor, the programs require no preparation of raw data. Instead, the instructor enters such information as the number of exercises to be produced, the difficulty level of the exercises, and the parameters (e.g., μ , σ) of the populations from which the data values are randomly generated. The parameter values can be chosen to produce marginal statistical significance, causing some of the generated exercises to yield statistically significant results and others to be nonsignificant.

In addition to producing computation exercises, the present programs are suitable for sampling experiments based on random samples drawn from normal populations with specified parameters. It is possible to systematically vary sample sizes and different parametric values to study their effects on a number of theoretical propositions. Such experiments might be used to help students understand theoretical concepts.

Computer and language. The programs are designed for time-sharing use. They are written in BASIC and were tested on Hewlett-Packard computers (Models 2100A, 2000E, and 2000F), but designed to be compatible with a variety of computers. Because the time-sharing mode minimizes turn-around time, the programs may be used in a classroom or statistics lab setting, perhaps for running sampling experiments.

Availability. Program listings and a manual giving detailed use information and sample output may be obtained at no charge from Paul Salmon, Department of Psychology, 2219 North Kenmore Avenue, Chicago, Illinois 60614.