

Multinomial Modeling

Multinomial modeling is a formal approach to measuring cognitive processes, such as the capacity to store and retrieve items in memory, or to make inferences and logical deductions, or to discriminate and categorize similar stimuli. While such processes are not directly observable, theoretically they can be assumed to interact in certain ways to determine observable behaviors. The goal of multinomial modeling is to identify which underlying factors are important in a cognitive task, explain how those processes combine to create observable behavior, and then use experimental data to estimate the relative contributions of the different cognitive factors. In this way, multinomial models can be used as tools to measure unobservable cognitive processes.

This entry is organized as follows. First, the type of data used in multinomial modeling is described, and how these data can be used to develop models of this type. A detailed example of multinomial modeling is provided as an illustration. Next, a number of common aspects of multinomial modeling are discussed, including validation testing, models for complex data structures, and the use of multinomial models to test hypotheses about cognitive processing. Finally, current work and future directions for multinomial modeling are outlined.

Data Structure and Model Development

Multinomial models are developed for categorical data, where each participant's response falls into one and only one of a finite set of observable data categories. These data usually come from a cognitive experiment, where each participant in an experimental group produces a categorical response to each of a series of items; for

example, pictures are ‘recognized’ or ‘not recognized’ or letter strings are judged to be ‘words’ or ‘non-words’. Most data sets for multinomial modeling involve more than two response categories, and in addition there may be more than one type of item, each with its own system of response categories. For example, in a source-monitoring experiment, participants study a list of items from two sources, Source A or Source B (e.g., presented in a male vs. female voice, or presented visually vs. auditorily). Later, participants are given a recognition memory test consisting of three types of items, namely the two types of old list items and new distracter items, and they must classify each tested item as Source A, Source B, or New. The resulting multinomial data structure consists of three category systems, each with three response categories. If the responses in different category systems are independent and category counts within a system follow a multinomial distribution, the probability of the data structure is given by the product of three multinomial distributions, one for each category system.

To express this more formally, assume that the data structure for an experimental task consists of J categories and N experimental response observations, where n_j observations fall into category C_j , $j=1,2, \dots, J$. Then if the observations are independent and identically distributed with probability p_j of falling into category C_j , the category count vector, $\mathbf{D} = (n_1, n_2, \dots, n_J)$, follows the multinomial distribution given by

$$\Pr[\mathbf{D}|\mathbf{p} = (p_1, \dots, p_J)] = N! \prod_{j=1}^J \frac{p_j^{n_j}}{n_j!},$$

where the category probabilities are non-negative and sum to one.

The key to creating a multinomial model is to take a multinomial data structure such as the one above and express the category probabilities in terms of underlying,

cognitively interpretable parameters. One needs to specify a cognitive processing architecture along with formal computational rules that can generate the count data in terms of the parameters. Once the model is constructed and data are collected, standard tools in statistical inference can analyze the data and evaluate the adequacy of the fit of the model to the data. In addition one can estimate the values of the cognitive parameters that are likely to have created the data. In this way unobservable cognitive processes can be measured indirectly with the use of the model.

Multinomial models of various types have been used in cognitive psychology since the 1960s; however, in the 1980s and 1990s a particular approach called multinomial processing tree (MPT) modeling was developed at a general level by William Batchelder, David Riefer, and Xiangen Hu. The central characteristic of MPT models is that they have a particular type of cognitive architecture represented as a rooted tree structure. Such a structure assumes that cognitive processes follow one another, and subsequent processes are conditionally dependent on the success or failure of earlier processes. For example, if a model has parameters for item attention, item storage, and item retrieval, then successful storage depends on successful attention. In turn, successful retrieval depends on successful storage. If any of these processes fail, then responses may be governed by guessing biases corresponding to various states of incomplete information. Each series of processing possibilities leads to different observable responses, and there are usually many of these processing patterns, each represented by the “branches” of the tree architecture.

An Example: Batchelder and Riefer's Pair-Clustering Model

One early example of an MPT model is the pair-clustering model developed by Batchelder and Riefer. Their model was designed to separately measure storage capacity from retrieval capacity in human memory. The data for the model involve a specially designed free-recall task where a participant studies a list of words one at a time, and then at a later time memory is tested by having the participant recall as many of the studied words as they can in any order. The list consists of pairs of exemplars from several categories such as vehicles (taxi, car) or flowers (rose, daisy). Recall of each category pair is scored into one of four categories: (1) both words recalled successively, C_1 , (2) both words recalled but not successively, C_2 , (3) exactly one word recalled, C_3 , and (4) neither word recalled, C_4 . The model postulates three parameters each designed to measure a different cognitive process: (i) a storage parameter c representing the probability that a pair of words is 'clustered' and stored in memory during study, (ii) a retrieval parameter r representing the probability that a word pair is recalled given that it is stored as a cluster, and (iii) a parameter u for the probability that a word in a pair that was not clustered is recalled as a singleton. Because the parameters refer to probabilities of successful cognitive processes, their values must be between zero and one.

The connection of the parameters to the category probabilities is based on a combination of psychological considerations and reasonable approximations. In particular it is assumed that both members of a category pair are recalled successively if and only if the words in the pair are clustered and the cluster is retrieved (joint probability cr). Also, if a cluster was stored but not retrieved, then neither word is recalled. In contrast, with probability $(1-c)$ the words in a pair are not clustered, and in this case each

word in the pair is or is not recalled individually with probability u , subject to the condition that if both non-clustered words are recalled they are not recalled successively. These assumptions can be displayed in the processing tree found in Figure 1.

Figure 1. The Pair-clustering Model

To briefly turn to the mathematical details, it is easy to use this tree to express the category probabilities, $p_j = \Pr(C_j)$ for $j=1,2,3,4$, in terms of the parameters. The result expresses each category probability as a sum of the probabilities of the branches that lead to that category as follows: $p_1 = cr$, $p_2 = (1-c)u^2$, $p_3 = (1-c)u(1-u) + (1-c)(1-u)u$, and $p_4 = c(1-r) + (1-c)(1-u)^2$. It is a matter of simple algebra to show that if we collect data $\mathbf{D} = (n_1, n_2, n_3, n_4)$ and estimate the category probabilities by relative frequencies, $P_j = n_j/N$, then we can solve the four model equations for parameters estimates (denoted by ‘*’) yielding $u^* = 2P_2 / (2P_2 + P_3)$, $c^* = 1 - P_2 / (u^*)^2$, and $r^* = P_1 / c^*$. In order for these equations to yield estimates of the parameters in the interval (0,1) it is necessary that $(P_3)^2 < 4P_2(1 - P_2 - P_3)$.

Common Aspects of Multinomial Modeling

The example of the pair-clustering model illustrates the basic properties of multinomial modeling, which are the tree architecture and the computational rules that tie the cognitive processing parameters to the categorical data. However, the example does not illustrate three aspects typical of most applications of MPT models. First, in the example there were three parameters representing cognitive processes and only three degrees of freedom in the data structure (since the four category probabilities are required

to sum to one). In cases where there are more degrees of freedom in the categorical data than parameters, the system of equations expressing category probabilities in terms of parameters is over determined and standard techniques in mathematical statistics are used to estimate the parameters. Second, the pair-clustering model involves just one system of categories, but many MPT models are developed for several category systems, each of which is associated with its own processing tree. For example, MPT models for the source monitoring experiment discussed earlier specify three processing trees, one for each item type (Source A, Source B, or New).

Finally, unlike the example of the pair-clustering model, most applications of MPT models involve two or more experimental groups of participants, where the same model with possibly different parameter values is assumed to govern each group's category count data. In this case MPT models are used to conduct hypothesis tests in an effort to discover which cognitive processes account for differences between the groups. This approach contrasts with the usual approach in experimental psychology for analyzing data from multiple experimental groups which is to apply standard statistical tools like analysis of variance or linear regression. While these tools are well developed to detect group differences and associate them to experimental manipulations, they do not allow one to pinpoint the cognitive bases for the differences.

As can be seen, MPT models are simple statistical models that are easy to develop and analyze. But before an MPT model can be used as a measurement tool it must be validated. A validated model is one where the parameters can be shown to represent the cognitive processes they stand for. Establishing validation involves conducting simple cognitive studies where experimental manipulations are designed to affect some

parameters and not others. These experiments attempt to dissociate the parameters by showing that they can be independently manipulated in ways that are consistent with established psychological theory. For example, for the pair-clustering model providing retrieval cues during recall should increase the value of the retrieval parameter r but not the value of the cluster storage parameter c . Other manipulations such as increased study time should affect the storage parameter but probably not the retrieval parameter.

Current and Future Directions

Since the 1990s, MPT modeling has become an increasingly popular approach to cognitive modeling, and its use has been facilitated by several software packages that can perform parameter estimation and hypotheses testing. To date there have been over a hundred examples of the application of MPT modeling. Most of these applications have been in the standard cognitive areas of memory, reasoning, and perception; however, clinical, social, and developmental psychology are also areas where MPT modeling is active. There are also a number of ongoing projects that explore the statistical properties of these models. For example, there has been recent work creating hierarchical MPT models to handle variation in parameter values due to individual differences in the participants, as well as latent class MPT models that can be used to model subgroups of participants with different cognitive abilities.

William H. Batchelder

Xiangen Hu

David M. Riefer

See also: Information Processing, Memory, Signal Detection Theory

Batchelder, W. H., & Riefer, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical & Statistical Psychology*, *39*, 129-149.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548-564.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*, 57-86.

Batchelder, W.H., & Riefer, D.M. (2007). Using multinomial processing tree models to measure cognitive deficits in clinical populations. In R. Neufeld (Ed.). *Advances in Clinical Cognitive Science* (pp. 19-50). Washington, D.C.: American Psychological Association Books.

Erdfelder, E., et al. (2009). Multinomial processing tree models: A review of the literature. *Journal of Psychology*, *217*, 108-124.

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*, 21-47.

Hu, X. & Phillips, G. A. (1999). GPT.EXE: A powerful tool for the visualization and analysis of general processing tree models. *Behavior Research Methods, Instruments, & Computers*, *31*, 220-234.

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*, *71*, 7-31.

Riefer, D.M., & Batchelder, W.H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, 95, 318-339.