Choice Modeling: Marketing Engineering Technical Note [1]

Table of Contents

Introduction

Firms today have access to increasing amounts of market response data at the level of individual customers, including data from scanner panels, direct marketing efforts, online retailing, loyalty programs, and the like. These data include both the marketing effort directed at a customer (e.g., price discount, or specific email sent to that customer) and the associated specific behaviors (e.g., purchase, customer support) of that customer. Consequently, there is also increasing interest among marketers in developing and using response models specified at the individual level. Analyses of individual-level data are useful for firms even for making decisions about aggregate marketing actions, such as TV advertising. After all, markets are composed of individuals, and acknowledging

and incorporating customer heterogeneity can be beneficial in a wide variety of marketing decision contexts.

One of the most widely used approaches for modeling individual customer behavior is the multinomial logit model (MNL), which can be used to explain and predict the choices that customers make (e.g., choosing brands, responding to an email, upgrading product). Other methods in Marketing for modeling behavior include Regression Analysis, Neural Networks, and Discriminant Analysis. In this note we only describe the MNL model, and describe how it can be used for customer targeting and for customer segmentation.

Description of the Multinomial Logit (MNL) Model

The theory of rational choice underlies much of modern Economics. According to this theory, individuals have well-ordered preferences for any set of choice alternatives (e.g., products, brands, candidates in an election), and they choose that alternative that maximizes their preferences. The MNL model offers a way to operationalize the theory of rational choice within a probabilistic framework. The objective of the MNL model in Marketing is to predict the probabilities that a customer would choose each of several alternatives which are available on a particular purchase/choice occasion. The MNL model is based on several core concepts: (1) The customer has an unobservable (at least to the modeler) preference or utility for each of the choice alternatives, (2) the utility of each choice alternative is composed of two additive terms, namely, a deterministic component (the intrinsic value or attractiveness of the choice alternative), and a random component that varies randomly across choice alternatives, customers, and purchase occasions, (3) the distribution of the random component can be specified, and (4) on each choice occasion, the customer chooses the alternative that provides him or her the highest utility. Below, we elaborate on these core concepts:

On each choice occasion, the (unobserved) utility that customer i gets from choice alternative k is given by:

$$U_k^i = A_k^i + \varepsilon_k^i \tag{1}$$

<u>where</u> ε_k^i is the random component of the customer's utility. We assume that

ε_k^i's are distributed independent Gumbel (i.e., type 1 extreme value) distribution.

The independence assumption implies that knowledge of the value of the random component for any customer, choice alternative, or purchase occasion does not provide any information about the value of the random component for another customer, choice alternative, or purchase occasion. Notice that utility, (U_k^i), is the sum of an observable component (A_k^i) and an unobservable component (ε_k^i), making it unobservable, or latent.

A_k^i is the overall "attractiveness" (view it as inferred preference or utility value) of alternative k to customer i

$$A_k^i = \sum_j \beta_j X_{ijk}.$$

(2)

X_{ijk} is the value (observed or measured) of a contextual variable j (e.g., Color of product; price of product, whether product j was on a special promotion on that purchase occasion) for product alternative k on a given purchase occasion.
β_j is the importance weight associated with variable j (estimated by the model – this is similar to regression coefficients).
We assume that customer i chooses the product which offers him or her the highest utility. Then, the probability that the customer i will choose alternative k is given by:

$$P_{ik} = P\{U_k^i \geq U_m^i; \text{for all m in the choice set}\} \qquad (3)$$

That is P_{ik} is the probability that the utility of product k will be at least as high as the utilities of any other product on that purchase occasion. Then, it can be shown that individual i's probability of choosing product 1 or choice alternative 1(P_{ik}) is given by:

$$P_{ik} = \frac{e^{A_1^i}}{\sum_k e^{A_k^i}} \qquad \text{for k =. 1, 2, ...K} \qquad (4)$$

Thus the logit model is a sequence of K equations (where K is the number of alternatives). When applied to a typical "brand choice" problem, the model components have the following interpretations:

3

X_{ijk} = customer i's evaluation of brand j on product attribute k (brand quality, for example), where the summation is over all brands that individual i is considering purchasing;

β_j = "revealed importance weight" showing the degree to which attribute j influences brand preferences (applies to all brands). These parameter estimates are revealed by an analysis of the past behavior (e.g., choice) of customers rather than by directly asking consumers. They can be broadly interpreted in much the same way as regression coefficients;

$\sum_{j} \beta_j X_{ijk}$ = Overall attractiveness (utility) of brand k for customer i

In the "aggregate Logit model," given in Eq. (4), β_j is the same for all individuals in a target market.

Properties of the MNL Model

What is the value of MNL models in Marketing? The answer, briefly, is that the structure of logit mirrors the differential sensitivities we expect in actual choice behavior. To see how this works, consider Eq. (4). The exponentiation in Eq. (4) ensures that the probabilities are always positive, since the exponentiation of any real number is always positive. Exponentiation also ensures that the probabilities do not change if all the measures of attractiveness are increased by a constant. Thus the measures of attractiveness need only form interval scales, something quite useful since most customer-based measures only achieve interval-scale quality.

S-shaped response function: An important characteristic of logit is that it produces an S-shaped curve, tracking the expected relationship between attractiveness and choice. Graphing Eq. (4) as a function of A_k^i produces an S-shaped curve that asymptotes to zero (no chance of being chosen) for very unattractive brands and to one for very attractive ones (almost certain to be chosen). In most applications of the logit model, the attractiveness of a brand (or, more generally, a choice alternative) is assumed to be a function of its characteristics. This attractiveness function is typically linear as in Eq. (3).

Inverted "U" Marginal response: The marginal impact of a change in an

attribute of an alternative X_{iji} takes a particularly simple form. For example,

considering Product 1, the derivative of P_{i1} as a function of X_{ij1} is

$$\frac{dP_{i1}}{dX_{ij1}} = w_k P_{i1}^*(1 - P_{i1}^*)$$ (5)

where P_{i1}^* is the predicted probability (as predicted by the model) that consumer i

will choose product 1 from the current choice set (Analogous expressions apply for other products in the choice set). Thus the marginal change in the probability that consumer i will choose product 1, for a unit change in variable k, turns out to

be a function of the predicted probability of choosing product 1 (P_{i1}^*). A graph of

Eq. (5) is given in Exhibit 1. The marginal impact of a given marketing effort is maximized when the probability of choosing the product is equal to .5, but the marginal impact approaches zero when the probability of choosing that product is near zero or close to one. Thus the logit model has a nice behavioral property: it ensures that the incremental impact of marketing effort directed at a product is at its peak when the consumer is "on the fence" about choosing it.
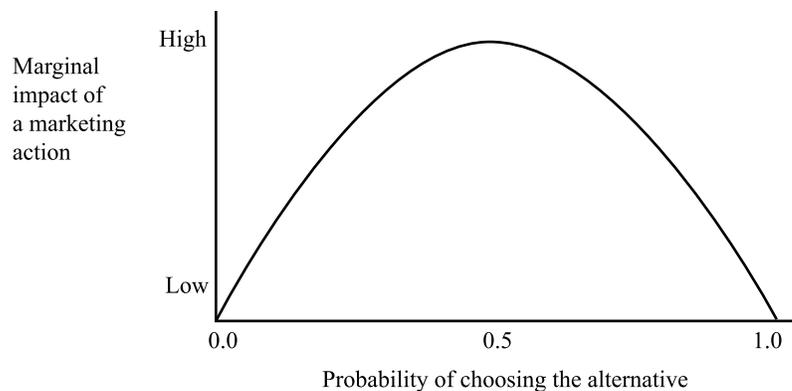
EXHIBIT 1

The marginal impact of marketing effort depends on the probability of choice.

Elasticity of response: Likewise, we can compute the elasticity of choice

probability, namely, the percentage change in the probability of choice for a 1% change in independent variable k, which is given by:

$$\frac{dP_{i1}}{dX_{ij1}}\frac{X_{i1k}}{P_{i1}} = w_k(1-P_{i1}^*)X_{ij1} \tag{6}$$

Other things equal, the response is more elastic when P_{i1}^* is lower, i.e., when product 1 has a lower probability of being chosen. In other words, low-share choice brands can gain proportionately more for their marketing efforts, as compared to high-share choice brands.

The above properties of the logit model are more credible than the properties of a linear probability model, which simply predicts P_{il} as a function of a linear combination of the X_{ilk}'s. The linear probability model assumes a constant probabilistic impact of any change in the X_{ilk}'s. That is counter to our ideas of what the impact of marketing and contextual factors on choice ought to be and can result in predicted probabilities that are less than zero or greater than one!

Proportional draw: This is another property of the logit model, which we illustrate with an example:

EXAMPLE

Suppose that someone performed a survey of shoppers in an area to understand their shopping habits and to determine the share of shoppers that a new store might attract. The respondents rated three existing stores and one proposed store (described by a written concept statement) on a number of dimensions: (1) variety, (2) quality, (3) parking, and (4) value for the money (Exhibit 2.). By fitting shoppers' choices of existing stores to their ratings through the logit model, we can estimate the coefficients [b_k]:

$$A_{ik} = b_1 X_{i1k} + b_2 X_{i2k} + ... + b_J X_{iJk} \tag{7}$$

where

    A_{ik} = attractiveness of store k (for customer i);

    X_{ijk} = customer i's rating or evaluation of store k on dimension j, $j = 1, ...$,
        J; and

b_j = importance weight for dimension j.

The data in Exhibit 2 come from a group of similar customers. Exhibit 3 gives the share of the old stores with and without the new store, the potential share of the new store, and the draw estimated from this group.

Attribute ratings by store

Store	Variety	Quality	Parking	Value for Money
1	0.7	0.5	0.7	0.7
2	0.3	0.4	0.2	0.8
3	0.6	0.8	0.7	0.4
4 (new)	0.6	0.4	0.8	0.5
Importance weight	2.0	1.7	1.3	2.2

EXHIBIT 2

Ratings and importance data for the store-selection example.

Store	(a) $A_i = w_k b_{jk}$	(b) e^{A_i}	(c) Share estimate without new store	(d) Share estimate with new store	(e) Draw [(c) – (d)]
1	4.70	109.9	0.512	0.407	0.105
2	3.30	27.1	0.126	0.100	0.026
3	4.35	77.5	0.362	0.287	0.075
4	4.02	55.7		0.206	

EXHIBIT 3

Logit model analysis of new store share example.

In column e of Exhibit 3, the draw is proportional to market share (column c). In other words, this model assumes that all individuals consider all brands in their choice process, that they do not go through any prescreening or eliminate some brands. (This prescreening is often referred to as a consideration process.) The proportional draw property implies, for example, that if a new light beer is introduced into the market, it will draw share from every product in the market (including regular beers), in proportion to the current market shares of the existing products. However, it is likely that the light beer will draw a disproportionate share from other light beers, rather than from regular beers. To

minimize the effects of such discrepancies, it is important that in applications of the logit model, we carefully specify the actual set of choices available to customers, or customer segments, based on market realities.

Researchers have also developed several ways to deal with the proportional draw problem. One way is a priori segmentation; the researcher segments the market into groups that *do* consider (different) sets of brands differently. Another alternative is to group products (rather than customers) into groups that more directly compete with one another. If we view the choice process as a hierarchy, we can then assume that consumers select among branches of a tree at each level of the hierarchy (Exhibit 4). The consumer might first choose the form of the deodorant and then, conditional on that choice, choose the brand. The form of the logit model that applies here is called the nested logit, and it incorporates an equation like Eq. (4) for the selection of product form (the upper level of the hierarchy) and a separate logit model for brand (conditional on the selection of form) at the lowest level of the hierarchy.
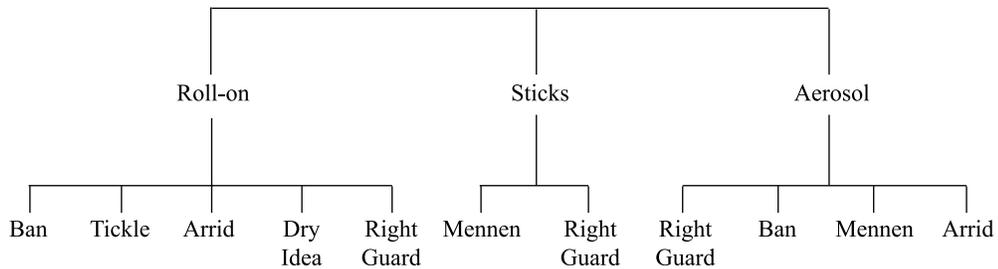
EXHIBIT 4

Consumer decision hierarchy for deodorant purchase. Source: Urban and Hauser 1980, p. 92.

The nested logit model can be represented as

$$P_{jk}^i = P_{j|k}^i P_j^i \tag{8}$$

where

P_{jk}^i = probability that customer i chooses brand k and product form j

P_j^i = probability that customer i first chooses product form j

$P_{k|j}^{i}$ = probability that customer i chooses brand k *given* he or she has

chosen product form j

(We drop the superscript i in the discussion below for simplicity.)

If we assume attractiveness is separable, we get

$$A_{jk} = A_j + A_{k|j} \tag{9}$$

where

A_{jk} = attractiveness of product form j and brand k

A_j = attractiveness of product form j

$A_{k|j}$ = attractiveness of brand k (when in product form j)

The brand choice (bottom level of the hierarchy in Exhibit 4) can be represented as a multinomial logit model as before:

$$P_{k|j} = \frac{e^{A_{k|j}}}{\sum_{\ell} e^{A_{\ell|j}}} \tag{10}$$

Under suitable assumptions, the product form probability has a similar structure:

$$P_j = \frac{e^{\mu[A_j + \ell n(\sum \exp A_{\ell|j})]}}{\sum_{\ell'} e^{\mu[A_\ell + \ell n(\sum \exp A_{\ell|\ell'})]}} \tag{11}$$

where is a normalizing constant to ensure that the sum of all choice probabilities add to 1.

Substituting Eq. (11) and Eq. (10) in Eq. (8) gives the full equation for the nested logit model. (See Roberts and Lilien, 1993, for a more complete discussion).

Logit Model Estimation via Maximum Likelihood

Individual choice models of all sorts are difficult to estimate. We outline here the general approach to estimating the simple MNL model (Eq. 4).

Let,

$$Y_k^i = \begin{cases} 1 & \text{if customer i chooses alternative } k \\ 0 & \text{if customer i does not choose alternative k} \end{cases} \qquad (12)$$

Then $P(Y_k^i = 1)$ is the probability that $U_k^i \geq U_m^i$ for all m $\neq$ k. Now consider the

likelihood that $P(Y_k^i = 1)$ for a random sample of N customers whose choices we

have observed. This sample likelihood is the product of the likelihoods that each

individual in the sample chose the alternative that they actually did, which can be

represented as:

$$L(\beta_1, \beta_2, ..., \beta_J) = \prod_{i=1}^{N} \prod_{k \in C} P(Y_k^i = 1)^{Y_k^i}, \qquad (13)$$

where C is the set of alternatives (the choice set) and βs are the unknown

parameters of the individuals' utility function to be estimated. Substitute for

$P(Y_k^i = 1)$ from Eq. (4) to get:

$$L(.) = \prod_{i=1}^{N} \prod_{k \in C} \left(\frac{e^{\sum_j \beta_j X_{jk}^i}}{\sum_k e^{\sum_j \beta_j X_{jk}^i}} \right)^{Y_k^i} \qquad (14)$$

To simplify estimation, we typically consider the logarithm of L, namely, Ln(L):

$$Ln(L) = \sum_{i=1}^{N} \sum_{k \in C} Y_k^i (\sum_j \beta_j X_{jk}^i - Ln \sum_{k \in C} e^{\sum_j \beta_j X_{jk}^i}) \qquad (15)$$

The estimates for βs can then be obtained by maximizing the Likelihood (L), or

equivalently, by maximizing *Ln(L),* by setting the partial derivatives to 0:

$$\frac{\partial Ln(L)}{\partial \beta_j} = \sum_{i=1}^{N} \sum_{k \in C} (Y_k^i - P_k^i) X_{jk}^i = 0 \quad for\, j = 1,\, 2,\, ...J \qquad (16)$$

This gives a set of J equations in J unknowns, which can be solved using

numerical methods. It can be shown that if a solution exists for this set of

equations, that solution (i.e., the maximum likelihood estimates for βs) is

unique. Further, the maximum likelihood estimates obtained this way have

many desirable statistical properties -- the estimates are consistent,

asymptotically Normal, and asymptotically efficient. The estimated βs can be

interpreted pretty much like regression coefficients.

EXAMPLE

Consider a situation where are four choice alternatives available to customers, and we also know the prices of the four alternatives. We can think of the logit model for this application as generating four equations in four unknowns: one parameter to represent the effect of product prices (β_1) and three alternative-specific constants (α_i's) to represent the intrinsic value of the four alternatives (e.g., brand image). (One of the alternative-specific constants, for example, α_1, is set to 0 to ensure that the model can be estimated).

Using Logit Models for Customer Targeting

Peppers and Rogers (1993) describe how a firm's best customers outspend its average customers by a factor of 16:1 in the retail industry, 12:1 for airlines, and 5:1 in hotels. Thus, it pays marketers to target their marketing efforts at customers who have the highest probabilities of purchase (or, more generally, the highest probability of a favorable response). An increasingly common approach to developing such targeting programs, especially in direct marketing (also called database marketing), is to use develop choice models to identify the most important factors driving customer choices. Typically, the choice model enables the firm to compute an individual's likelihood of purchase, or some other behavioral response, based on variables that the firm has in its database, such as geodemographics, past purchase behavior for similar products, attitudes or psychographics.

A firm can use the probability of choice/purchase estimated from an MNL model to calculate expected customer profitability under a particular action it takes. For example, a direct marketing firm can direct its marketing campaign to those customers (or customer segments) whose expected profitability exceeds the cost of reaching them:

Expected (gross) customer profitability = Probability of purchase

x Likely purchase volume if a purchase is made

(17)

x Profit margin (for this customer).

EXAMPLE

Exhibit 5 shows part of a direct marketing database after the firm has

completed the choice modeling step just discussed. Choice modeling provided the data in column *A*—purchase probability. The question, then, is which customers should the firm target?

Suppose that the total cost of reaching one of these customers is $3.50. What should the firm do? Firms commonly use several approaches to answer this question. First, if the firm looks at the average expected profit, it may decide to target all 10 groups and make a small profit (103($3.72−$3.50) = $2.20).

Or it may target customers 1, 3, 5, and 6 and make

$6.51+$3.62+$6.96+$6.20-(4×$3.50)=$9.29.

Notice that by using choice-based modeling the firm can target customers to improve profitability by over 400 percent.

Finally, using a more traditional segmentation by average purchase volume, the firm would target, say 30 percent, or the three largest customers in this case—2, 4, and 9—and lose $5.02!

Firms using this approach typically compute the expected customer profitability at the individual level and then sort the customer database in decreasing order of expected profitability (Exhibit 5, column *D*). They then target customers who exceed some threshold (a profitability measure) or fall into the most profitable percentage of the database. In the example below, we describe how this works.

Customer	A Purchase Probability	B Average Purchase Volume	C Margin	D Customer Profitability = A x B x C
1	30%	$ 31.00	0.70	$6.51
2	2%	$143.00	0.60	$1.72
3	10%	$ 54.00	0.67	$3.62
4	5%	$ 88.00	0.62	$2.73
5	60%	$ 20.00	0.58	$6.96
6	22%	$ 60.00	0.47	$6.20
7	11%	$ 77.00	0.38	$3.22
8	13%	$ 39.00	0.66	$3.35
9	1%	$184.00	0.56	$1.03
10	4%	$ 72.00	0.65	$1.87

Average Expected Profit = $3.72

EXHIBIT 5

Choice-based segmentation example for database marketing: target those customers whose (expected) profitability exceeds the cost of reaching them by comparing column *D* with the cost to reach that customer.

Using Logit Models for Customer Segmentation

Choice models can also be used to segment customers on the basis of the variables that most influence choices in each identified segment. In what follows, we outline the methods used for latent class choice segmentation. This approach enables marketers to understand the unobserved (latent) choice processes driving different segments of customers to behave differently in making choices. Such understanding can then be used to target different groups of customers with the appropriate marketing programs. For example, customers who are more price sensitive (as evidenced by their previous choices) can then be identified and offered special promotions not available to the less price sensitive customers, who can be offered products with enhanced features or services.

In the "aggregate logit model" summarized in Eq. (4), every customer has an identical choice process (i.e., utility function or purchase probability rule) although each customer makes different choices because of differences in the deterministic or random components in their "common" utility function. However, customers not only differ along observed characteristics (e.g., sex, race) but also with respect to the unobserved, but systematic, rules that they use for making judgments about choice alternatives. While we rarely have sufficient data about each individual to build separate individual utility functions, we may still

want to segment customers according to their latent choice rules to account for the heterogeneity that exists in the population.

Customer heterogeneity can be classified into two categories: (1) observed heterogeneity (e.g., customers differ on observable characteristics such as gender), and (2) unobserved heterogeneity (e.g., customers differ in terms of their price sensitivities). Observed heterogeneity can be modeled directly by including associated independent variables (e.g., gender) in the choice model. However, the same idea does not work for modeling unobserved heterogeneity (e.g., we cannot construct a variable for price sensitivity because we do not observe it). A common approach for accommodating unobserved heterogeneity is to use finite mixture modeling, in which each segment is assumed to follow its own choice rule. In the framework of logit models, unconditional purchase probability is then assumed to be a mixture of several conditional purchase probabilities, where each conditional probability corresponds to a segment. Then, given the actual choices people make, we can infer the most likely values of these segment-level parameters (e.g., price sensitivities for different segments) from the data, i.e., we simultaneously form segments as well as estimate the unknown choice process within each segment through the maximum likelihood estimation method. Operationally, this means that the weights (β_j's) in the logit model differ across segments, but the segments are unknown (latent) and have to be inferred from the data. To accommodate this possibility, we can specify Eq. (4) as follows:

$$(P_k^i \mid i \text{ belongs to segment } s) = \frac{e^{\sum_j \beta_{js} X_{ijk}}}{\sum_k e^{\sum_j \beta_{js} X_{ijk}}} \qquad (18)$$

There are several methods available for estimating the parameters in Eq. (7). These methods allow the estimation of: (a) the number of segments that best fit the data, (b) the parameters (β_{js}) of the utility function of each segment, (c) the proportions of the population that belong to each segment, and (d) the segment to which a particular customer is most likely belong to, whether or not that customer's purchase behavior was used to estimate the parameters of the model. A popular method for estimating the parameters via latent class analysis is the EM (Expectation Maximization) algorithm (Wedel and Kamakura, 2000). A special concern in estimating latent class models (as compared to the aggregate MNL model) is that there may issues of identifiability (i.e., insufficient number of distinct data patterns for estimating all the model parameters), a problem that is likely to exacerbated if the predictor variables are all nominal and there are not

many of them, and/or there are not sufficient number of choice alternatives for the number of parameters to be estimated.

Determining the number of latent segments in MNL models: There are several indices to assess the goodness of fit of the estimates of the MNL model that function similarly to the R^2 index associated with regression models: (1) Hit ratio – the proportion of out-of-sample observations correctly classified by the estimated model; the higher this ratio, the higher the predictive validity of the model, with a maximum possible value of 1; and (2) AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and CAIC (Consistent AIC), all of which indicate superior model performance the closer they are to 0. These indices enable the modeler to determine the number of segments in the data, i.e., to choose the model for which the number of segments results in an index value closest to 0.

The AIC criterion enables the analyst to trade off model fit against model complexity. Model fit can be improved by adding more variables, which however may increase complexity, or overweight unimportant aspects that are disproportionately present in the sample as compared to their presence in the population. In addition to accounting for the number of variables in the model, the BIC criterion accounts for sample size. We recommend the BIC criterion, unless the modeler has knowledge about the pros and cons of each index in a specific application. For further details on these indices as well as about the EM algorithm, see Jagpal (1999) and Wedel and Kamakura (2000).

EXAMPLE

Assume we have a two-brand market, with brands A and B, whose major difference is in price and that each customer i's "attractiveness" for these can be assessed as

$$\text{Attractiveness of brand A for customer i is} \quad \left[\frac{P_B}{P_A} \right]^{k_i} , \qquad (19)$$

where P_A and P_B are the prices of the brands and k_i is the price sensitivity parameter for customer *i*, where the higher the value of k, the more price sensitive the customer *i*.

15

Now, according to this model, the probability that customer i buys brand A can be assessed as

$$PROB_i(A \mid k_i) = \frac{(P_A/P_B)^{k_i}}{(P_A/P_B)^{k_i} + (P_B/P_A)^{k_i}} \tag{20}$$

and $PROB_i(B \mid k_i) = 1 - PROB_i(A \mid k_i)$

Assume that customers are of one of two types, low price sensitivity (k_l) or high price sensitivity (k_h), where we know neither the level of price sensitivity (k) nor the proportion of the population with that level of sensitivity (b_l , b_h), i.e., the mixing distribution. What can we say about the (unconditional) probability of a customer buying brand A?

Using the formula for total probability we get

$$PROB_i(A) = PROB_i(A \mid k_i)b_l + PROB_i(A \mid k_h)b_h \tag{21}$$

where $PROB_i(A/k_l)$ and $PROB_i(A/k_h)$ are determined from Eq. (20).

The challenge here is to estimate the four parameters in Eq. (21): k_l, k_h (the levels of price sensitivity) and b_l, b_h (the proportions of the population with those levels of price sensitivity—the weights in the mixing distribution), given observed choices that customers make in different price situations.

In the example above, we assumed two segments (high and low price sensitivity) and that individual purchase probabilities varied only by price sensitivity. In general, response models will have a number of parameters (like price sensitivity here) and the number of segments will not be known in advance.

Summary

In this technical note, we provided a broad overview of the MNL model focusing on its structure, properties, estimation, and uses. We also described how choice models can be used in customer targeting and segmentation: (1) Using choice probabilities estimated from an aggregate choice model for purposes of selecting target customers, and (2) Segmenting customers on the basis of their unobserved choice processes via a latent class choice model.

References

Peppers, Don, and Rogers, Martha, 1993, *The One to One Future, Building Relationships One Customer at a Time*, Currency Doubleday, New York.

Roberts, John H., and Lilien, Gary L., 1993, "Explanatory and predictive models of consumer behavior," in *Handbooks in Operations Research and Management Science, Vol. 5, Marketing*, eds. Jehoshua Eliashberg and Gary L. Lilien, Elsevier Science Publishers B.V., North Holland, New York, pp. 27–82.

Urban, Glen L., and Hauser, John R., 1980, *Design and Marketing of New Products*, Prentice Hall, Englewood Cliffs, New Jersey

Wedel, Michel, and Kamakura, Wagner A., 2000, *Market Segmentation: Conceptual and Methodological Foundations*, second edition, Kluwer Academic Press, Boston, Massachusetts