MaxDiff Adjusted to Items Non-Availability and Network Effects

Stan Lipovetsky and W. Michael Conklin
GfK North America, 8401 Golden Valley Rd, Minneapolis, MN 55427

Abstract
Maximum Difference (MaxDiff), also known as Best-Worst Scaling, is a discrete choice technique widely known in marketing research for finding utilities and probabilities among multiple items. It was proposed by J. Louviere and widely used for finding utilities and preference probabilities among multiple items. It can be seen as an extension of the paired comparison techniques for simultaneous presentation of several items together to each respondent. A respondent identifies the best and the worst ones and estimation of utilities is usually performed in a multinomial-logit (MNL) as a discrete choice modeling. It produces utilities and choice probabilities of the compared items. This work considers how to obtain robust probability estimation adjusted to possible absence of some items on shelves in actual purchasing. For this aim we apply Markov chain modeling in form of Chapman-Kolmogorov system of differential equations and its steady-state solution which can be reduced to eigenproblem by a stochastic matrix and solved analytically. The obtained closed-form solution suggests a robust modification of choice probabilities with accounted cases of items non-availability. Adjustment to choice probability with network effects is also considered. Numerical example by marketing research data is used and the results are discussed.

Key Words: MaxDiff, choice probability, Markov chain, Chapman-Kolmogorov equations, network effects, marketing research.

1. Introduction

Maximum Difference (MaxDiff), also known as Best-Worst scaling, is a modern marketing research approach to evaluation probability of choice among many compared items. This method has been originated by Jordan Louviere (1991, 1993), and developed in multiple works (for instance, Marley and Louviere, 2005; Bacon et al., 2007, 2008). It can be seen as extension of Thurstone scaling from pairwise to simultaneous comparison among three or more items in a balanced plan where respondents indicate the best and worst items, with following estimation of choice probabilities in MNL or analytically (Louviere et al., 2015; Marley et al., 2008, 2016; Lipovetsky and Conklin, 2014a,b, 2015; Lipovetsky et al., 2015). Estimation of utilities and choice probabilities is usually performed in multinomial-logit (MNL) approach to the discrete choice modeling (DCM).
The recent work (Blanchet et al., 2016) considers choice models based on Markov chains with states corresponding to the items and transition probability defined by the preferences among the items, and applies this approach to assortment optimization. As shown in that work, if a product/item is not available, a customer substitutes the most preferred by another item, and such sequential transitions are described by Markov chain model, where arriving probability equals the MNL choice values, and transitional probabilities are defined as the inflated probabilities are re-estimated by exclusion of each

item from their set. The aim of the current paper is to answer the question: if some items are not actually available, how can such a scenario change the results of estimation? Using transition probabilities we describe Markov chain for discrete states in the Chapman-Kolmogorov equations, and show that this system can be solved analytically and presented in the closed-form. This technique yields a robust adjustment of the choice probabilities to situation of absence of would-be-purchased items on shelves.
Another interesting possibility to improve estimating of choice probability can be seen in accounting for possible spread and impact of opinions in social networks, when consumers tend to prefer the products of higher frequency and values endorsed by their friends and acquaintances. Some approaches using endogenous networks effects are described in various works (for instance, Anderson et al., 1992; Du et al., 2016; Wang and Wang, 2016). The current paper also includes such a consideration based on simple analytical adjustment.

2. Markov chain and steady-states in eigenproblem

Choice probability among several alternatives can be described by MNL model

$$p_j = \frac{\exp(a_j x_j)}{\sum_{k=1}^{n} \exp(a_k x_k)}, \quad j = 1, 2, ..., n \tag{1}$$

where x_j are predictor variables and a_j are utility parameters defining probability p_j of each j-th choice among all n of them. One of the shares (1) is usually taken as a reference with its parameter put to zero, for instance, a_1=0. The model can be more complicated as in a DCM with many predictors in each exponent, or less complicated with binary predictors as in BWS. All the share probabilities satisfy the evident relation of their total

$$p_1 + p_2 + ... + p_n = 1\,. \tag{2}$$

According to Blanchet et al. (2016) on Markov chain modeling of the choice probabilities, each alternative can be considered as a discrete state with arriving probabilities p_j defined by MNL (1) and transition probabilities q_{ij} defined as follows:

$$q_{ij} = \frac{p_j}{1 - p_i}, \quad i, j = 1, 2, ..., n\,. \tag{3}$$

If an i-th item is not available in assortment, the other j-th shares are renormalized by the total probability 1-p_i, so conditional probabilities (3) describe a possible substitution of each i-th item by another j-th item. Each element q_{ij} measures the preference of j-th item over i-th item that corresponds to transition to the preferred state j from the state i. A matrix with intensity of transition elements q_{ij} (3) can be seen as a connected oriented graph with n nodes of states/items and two edges between each pair of nodes – the one going to state j from state i corresponds to transition intensity q_{ij}, and the other going from state j to state i corresponds to transition intensity q_{ji}.

Markov chain transitions for discrete states and continuous time can be presented in Chapman-Kolmogorov equations defining change in probability to belong to each state P_j (those denoted by capital P) as linear aggregates of these probabilities weighted by transition intensities (Bellman, 1960). For instance, change in time of the choice probability P_1 for the 1[st] product adjusted to non-availability of products can be written as follows:

$$\frac{dP_1}{dt} = -q_{12}P_1 - q_{13}P_1 ... - q_{1n}P_1 + q_{21}P_2 + q_{31}P_3 + ... + q_{n1}P_n\,, \tag{4}$$

where the negative items correspond to P_1 diminished by transitions to other states, and the positive ones corresponds to P_1 increasing due to arrival from the other states. Using transition intensities q_{ij} (3) in (4) yields:

$$\begin{aligned}\frac{dP_1}{dt} &= -\frac{p_2}{1-p_1}P_1 - \frac{p_3}{1-p_1}P_1 - ... - \frac{p_n}{1-p_1}P_1 + \frac{p_1}{1-p_2}P_2 + \frac{p_1}{1-p_3}P_3 + ... + \frac{p_1}{1-p_n}P_n \\ &= -P_1 + \frac{p_1}{1-p_2}P_2 + \frac{p_1}{1-p_3}P_3 + ... + \frac{p_1}{1-p_n}P_n\end{aligned} \tag{5}$$

where (2) is accounted. Similarly the other states can be described, and the total system of Chapman-Kolmogorov equations is as follows:

$$\begin{cases}\dfrac{dP_1}{dt} = \quad -P_1 + \dfrac{p_1}{1-p_2}P_2 + \dfrac{p_1}{1-p_3}P_3 + ... + \dfrac{p_1}{1-p_n}P_n \\ \dfrac{dP_2}{dt} = \dfrac{p_2}{1-p_1}P_1 \quad -P_2 + \dfrac{p_2}{1-p_3}P_3 + ... + \dfrac{p_2}{1-p_n}P_n \\ - \; - \; - \; - \; - \; - \; - \; - \; - \; - \; - \; - \\ \dfrac{dP_n}{dt} = \dfrac{p_n}{1-p_1}P_1 + \dfrac{p_n}{1-p_2}P_2 + \dfrac{p_n}{1-p_3}P_3 + ... - P_n\end{cases} \tag{6}$$

As it is well-known (Bellman, 1960) the solution of a homogeneous linear system of differential equations with constant coefficients can be presented as

$$P(t) = P\,diag(\exp(\lambda t))\,c\;, \tag{7}$$

where $diag(.)$ denotes a diagonal matrix, c is a vector of constants. The vector λ and matrix P consist of eigenvalues and columns of eigenvectors, respectively, obtained in solving the eigenproblem of the matrix in (6):

$$(Q-I)P = \lambda P\,, \tag{8}$$

where I denotes the identity matrix, and Q is the matrix of transition intensities (3):

$$Q = \begin{pmatrix} 0 & \dfrac{p_1}{1-p_2} & \dfrac{p_1}{1-p_3} & ... & \dfrac{p_1}{1-p_n} \\ \dfrac{p_2}{1-p_1} & 0 & \dfrac{p_2}{1-p_3} & ... & \dfrac{p_2}{1-p_n} \\ - & - & - & - & - \\ \dfrac{p_n}{1-p_1} & \dfrac{p_n}{1-p_2} & \dfrac{p_n}{1-p_3} & ... & 0 \end{pmatrix}. \tag{9}$$

The solution (7) for the starting moment t=0 reduces to $P(0) = Pc$, and solving this linear system with respect to c yields the vector of constants $c = P^{-1}P(0)$, and P^{-1} is the inverted matrix of the eigenvectors (8). The vector of initial conditions can be defined by the arrival probabilities p_j, so $P(0) = p$, and general solution (7) of differential equations (6) reduces to

$$P(t) = P\,diag(\exp(\lambda t))\,P^{-1}p\,. \tag{10}$$

The expression $P\,diag(\exp(\lambda t))\,P^{-1}$ in (10) is known as matrix exponent. Each component of the vector *P(t)* is a linear combination of the exponents $\exp(\lambda_j t)$ which behave in accordance with the specific values of λ_j obtained in the eigenproblem (8). The reciprocal eigenvalues can be interpreted as the mean time of transitions among the states, $-1/\lambda_j = \bar{t}_j$.

3. Steady-states choice probability in the analytical closed-form solution

The choice probabilities P_j can be also obtained in explicit analytical closed-form solution. For the stabilized process with derivatives in (6) equal zero, the steady-states probabilities P_j can be found by solving the corresponding system of linear homogeneous equations:

$$\begin{cases} 0\cdot P_1 + \dfrac{p_1}{1-p_2}P_2 + \dfrac{p_1}{1-p_3}P_3 + \dots + \dfrac{p_1}{1-p_n}P_n = P_1 \\ \dfrac{p_2}{1-p_1}P_1 + 0\cdot P_2 + \dfrac{p_2}{1-p_3}P_3 + \dots + \dfrac{p_2}{1-p_n}P_n = P_2 \\ - \quad - \quad - \quad - \quad - \quad - \quad - \quad - \quad - \quad - \quad - \quad - \\ \dfrac{p_n}{1-p_1}P_1 + \dfrac{p_n}{1-p_2}P_2 + \dfrac{p_n}{1-p_3}P_3 + \dots + 0\cdot P_n = P_n \end{cases} \tag{11}$$

which can be represented in matrix form as

$$QP = \lambda P, \tag{12}$$

where $\lambda = 1$, and *P* is a vector of unknown elements $P_1, P_2, \ldots, P_n$. It is easy to see that totals in each column (9) equals one, for instance, in the 1st column:

$$\frac{p_2}{1-p_1} + \dots + \frac{p_n}{1-p_1} = \frac{1-p_1}{1-p_1} = 1, \tag{13}$$

where the relation (2) is used. A so-called stochastic matrix has totals in each row equal one, so (9) is a transposed stochastic matrix.

The equation (12) is the eigenproblem for the matrix (9), with λ and *P* its eigenvalues and eigenvectors. The eigenvalues of a matrix and its transposition coincide. Due to Perron-Frobenius theory, for a stochastic non-negative matrix its main eigenvalue equals $\lambda_{\max} = 1$, the main eigenvector is unique and non-negative, and the same concerns a transposed stochastic matrix. Thus, for solving the system of homogeneous linear equations (11) we use the eigenproblem (12) and take the main eigenvector *P* as the vector of choice probabilities adjusted to possible non-availability of products.

The eigenproblem (12) for $\lambda_{\max} = 1$ reduces to the linear homogeneous equations $(Q-I)P = 0$, with *I* as identity matrix. The matrix *Q-I* (the same used in (8)) can be presented as follows:

$$Q - I = \begin{pmatrix} p_1 \\ p_2 \\ \dots \\ p_n \end{pmatrix} \cdot \begin{pmatrix} (1-p_1)^{-1} \\ (1-p_2)^{-1} \\ \dots \\ (1-p_n)^{-1} \end{pmatrix}' - diag\begin{pmatrix} (1-p_1)^{-1} \\ (1-p_2)^{-1} \\ \dots \\ (1-p_n)^{-1} \end{pmatrix} \tag{14}$$

where prime denote transposition, so there is an outer product of two vectors minus diagonal matrix to keep zeros on the diagonal of the matrix (11).

For solving a system of homogeneous linear equations we can put one of the unknown parameters in the vector P to be a constant, for instance, P_1=1. Then skipping the first equation we reduce (11) to the nonhomogeneous system of equations of the n-1 order:

$$\left[\begin{pmatrix} p_2 \\ \dots \\ p_n \end{pmatrix} \cdot \begin{pmatrix} (1-p_2)^{-1} \\ \dots \\ (1-p_n)^{-1} \end{pmatrix}' - diag\begin{pmatrix} (1-p_2)^{-1} \\ \dots \\ (1-p_n)^{-1} \end{pmatrix}\right] P_{2:n} = -(1-p_1)^{-1}\begin{pmatrix} p_2 \\ \dots \\ p_n \end{pmatrix} \tag{15}$$

where $P_{2:n}$ denotes the vector P without the 1st element. To solve this matrix equation with respect to the vector $P_{2:n}$ we need to invert the matrix of the structure (15). For this aim we apply the known Sherman-Morrison formula

$$(A+uv')^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1+u'A^{-1}v}\ , \tag{16}$$

with which the inversion of the matrix at the left-hand side (15) yields:

$$\left[-diag\begin{pmatrix} (1-p_2)^{-1} \\ \dots \\ (1-p_n)^{-1} \end{pmatrix} + \begin{pmatrix} p_2 \\ \dots \\ p_n \end{pmatrix} \cdot \begin{pmatrix} (1-p_2)^{-1} \\ \dots \\ (1-p_n)^{-1} \end{pmatrix}'\right]^{-1} = -diag\begin{pmatrix} 1-p_2 \\ \dots \\ 1-p_n \end{pmatrix} - \frac{1}{p_1} diag\begin{pmatrix} (1-p_2)p_2 \\ \dots \\ (1-p_n)p_n \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix}' . \tag{17}$$

Then solution of the system (15) is:

$$P_{2:n} = \frac{1}{(1-p_1)p_1}\begin{pmatrix} (1-p_2)p_2 \\ \dots \\ (1-p_n)p_n \end{pmatrix}. \tag{18}$$

Combining the first element P_1=1 with the vector (18) yields the total vector of adjusted choice probabilities:

$$P = \begin{pmatrix} 1 \\ P_{2:n} \end{pmatrix} = \frac{1}{(1-p_1)p_1}\begin{pmatrix} (1-p_1)p_1 \\ (1-p_2)p_2 \\ \dots \\ (1-p_n)p_n \end{pmatrix}. \tag{19}$$

Normalizing it by the total equal one we represent (19) as follows:

$$P = \begin{pmatrix} P_1 \\ P_2 \\ \dots \\ P_n \end{pmatrix} = \frac{1}{1-\sum_{k=1}^{n} p_k^2} \cdot \begin{pmatrix} p_1 - p_1^2 \\ p_2 - p_2^2 \\ \dots \\ p_n - p_n^2 \end{pmatrix}. \tag{20}$$

Thus, having an initial set of MNL probabilities p_j we use the formula (20) for estimating the values P_j adjusted for accounting a possible non-availability of any product.

In choice-based conjoint estimations with probabilities defined via MNL (1), the robust values can be obtained by (20) which yields probabilities in explicit form:

$$P_j = \frac{p_j - p_j^2}{1 - \sum_{k=1}^{n} p_k^2} = \frac{\exp(a_j x_j) \sum_{k \neq j}^{n} \exp(a_k x_k)}{\left(\sum_{k=1}^{n} \exp(a_k x_k)\right)^2 - \sum_{k=1}^{n} \exp(2a_k x_k)} . \tag{21}$$

4. Accounting for network effects

It is also possible to improve estimation of individual and average choice probabilities taking into account the impact of opinions in social network when consumers tend to choose the products of higher total preference. Adjustment to the aggregated probability can be performed in a simple analytical approach as follows. With utility parameters a_j of the MNL model (1) we can estimate change in individual choice probabilities due to networks effects as proportional to the aggregated probability of items:

$$\Delta p_{vj} = \frac{\exp(a_j x_{vj})}{\sum_{k=1}^{n} \exp(a_k x_{vk})} p_j = p_{vj} p_j , \tag{22}$$

where v denotes the individual observations (v=1, 2,…, N – total base size), and the aggregated probability for each j-th item is

$$p_j = \frac{\sum_{v=1}^{N} p_{vj}}{\sum_{k=1}^{n} \sum_{v=1}^{N} p_{vj}} . \tag{23}$$

Re-estimation of the total probabilities adjusted to network effects yields:

$$\widetilde{p}_j = \frac{p_j + \Delta p_j}{\sum_{k=1}^{n} (p_k + \Delta p_k)} = \frac{p_j (1 + p_j)}{\sum_{k=1}^{n} \left(p_k (1 + p_k)\right)} . \tag{24}$$

Then using $\widetilde{p}_j$ (24) in place of p_j in (20) produces the choice probability estimates adjusted by network effects and by possible non-availability of the products.

5. Numerical Example

For numerical illustration we use a real marketing research project for MaxDiff prioritizing seventeen items where 3,062 respondents saw four items in ten tasks for choosing the best and worst items – more detail are given in Lipovetsky and Conklin (2014a,b) where this dataset was used. Table 1 presents original MaxDiff solution and its robust adjustment P_j (20). The adjustment mostly concerns the bigger elements in the original solution and makes it smoother. The last two columns present solution (24) with networks accounted, and its robust adjustment (20) as well. As expected, the adjustment for network effects increases the bigger elements in the vectors in comparison with those in the previous columns. The results in the last column define the analytical solution accounted for the network effects and conditioned on possible items non-availability.

Table 1. MaxDiff choice probability (% in total): solutions original and adjusted to items non-availability, also with network effects and that adjusted to items non-availability.

	MaxDiff analytical p_j	Robust adjustment P_j	With networks effects p_j	Robust adjustment with networks effects P_j
1	4.79	5.15	4.50	4.94
2	3.07	3.36	2.83	3.17
3	25.65	21.55	28.89	23.64
4	6.92	7.28	6.64	7.13
5	6.61	6.98	6.31	6.81
6	1.58	1.76	1.43	1.63
7	4.75	5.11	4.46	4.90
8	2.62	2.88	2.41	2.71
9	6.10	6.47	5.80	6.29
10	4.13	4.48	3.86	4.27
11	1.09	1.22	0.99	1.13
12	6.78	7.14	6.49	6.98
13	14.26	13.82	14.61	14.36
14	2.37	2.62	2.17	2.44
15	2.85	3.13	2.63	2.94
16	2.87	3.15	2.65	2.97
17	3.57	3.89	3.31	3.69

6. Summary

The work describes choice probability modeling conditional on some items non-availability. Using transition probabilities defined from multinomial-logit modeling in a Markov chain, we show that the problem can be solved analytically and presented in the closed-form. This technique yields a robust adjustment of the probabilities obtained by MaxDiff in a scenario of absence of some products on shelves. Adjustment to choice probability with network effects is also considered. The methods are illustrated on the MaxDiff data but can be applied to other discrete choice problems in their multinomial-logit models completed by Markov chain modeling. The considered analytical solution can be useful in various marketing research problems. More detail on it is given in Lipovetsky and Conklin (2018).

References

Anderson S.P., de Palma A., Thisse J. (1992) *Discrete Choice Theory of Product Differentiation*, MIT Press, Cambridge, MA.

Bacon L., Lenk P., Seryakova K. and Veccia E. (2007) Making MaxDiff More Informative: Statistical Data Fusion by way of Latent Variable Modeling, *Sawtooth Software Conference Proceedings*, 327-344, Santa Rosa, CA.

Bacon L., Lenk P., Seryakova K., and Veccia E. (2008) Comparing Apples to Oranges, *Marketing Research*, American Marketing Association, Spring issue, 29-34.

Bellman R. (1960) *Introduction to Matrix Analysis*, McGraw-Hill, New York.

Blanchet J., Gallego G., and Goyal V. (2016) A Markov Chain Approximation to Choice Modeling, Operations Research 64, 4, 886-905.

Du C., Cooper W., and Wang Z. (2016) Optimal pricing for a multinomial logit choice model with network effects, Operations Research, 64, 2, 441–455.

Lipovetsky S. and Conklin M. (2014a) Best-Worst scaling in analytical closed-form solution, *Journal of Choice Modelling* 10, 60–68.

Lipovetsky S. and Conklin M. (2014b) Finding Items Cannibalization and Synergy by BWS Data, *Journal of Choice Modelling*, 12, 1-9.

Lipovetsky S. and Conklin M. (2015) MaxDiff Priority Estimations with and without HB-MNL, *Advances in Adaptive Data Analysis*, 7, DOI: 10.1142/S1793536915500028.

Lipovetsky S. and Conklin M. (2018) Choice Models Adjusted to Non-Available Items and Network Effects, International Journal of Business Analytics (IJBAN), 5(2), forthcoming.

Lipovetsky S., Liakhovitski D., and Conklin M. (2015) What's the Right Sample Size for My MaxDiff Study? *Sawtooth Software Conference Proceedings*, Orlando, FL,119-141.

Louviere J.J. (1991) Best-Worst Scaling: A Model for the Largest Difference Judgments, *Working Paper*, University of Alberta.

Louviere J.J. (1993) The Best-Worst or Maximum Difference Measurement Model: Applications to Behavioral Research in Marketing, *The American Marketing Association's Behavioral Research Conference*, Phoenix, Arizona.

Louviere J.J., Flynn T.N., and Marley A.A.J. (2015) *Best-worst scaling: Theory, methods and applications*. Cambridge, UK: Cambridge University Press.

Marley A.A.J. and Louviere J.J. (2005) Some Probabilistic Models of Best, Worst, and Best-Worst Choices, *J. of Mathematical Psychology*, 49, 464–480.

Marley A.A.J., Flynn T.N., and Louviere J.J. (2008) Probabilistic Models of Set-Dependent and Attribute-Level Best-Worst Choice, *J. of Mathematical Psychology*, 52, 281-296.

Marley A.A.J., Islam T., Hawkins G.E. (2016) A formal and empirical comparison of two score measures for best–worst scaling, *Journal of Choice Modelling,* http://dx.doi.org/10.1016/j.jocm.2016.03.002i

Wang R. and Wang Z. (2016) Consumer Choice Models with Endogenous Network Effects, Management Science, forthcoming, http://dx.doi.org/10.1287/mnsc.2016.2520.