Mathematical Models in Marketing
Prof. Rajeev Kohli

Error Theory for Elimination by Aspects
Rajeev Kohli, Kamel Jedidi

Summary by Yegor Tkachenko, Shitai Zhang
09/29/2017

Order independence (weakest form of the independence of irrelevant alternatives (IIA))

Ordering of the choice probabilities of two alternatives should be independent of any other alternatives in the choice set.

Violation: Alternative A is slightly more probable than alternative B.

$$p(A) = p(B) + \delta > p(B), \qquad \delta > 0$$

Alterative C, very similar to A, is introduced, takes away half of A's choices, steals nothing from alternative B.

$$p(A) = p(C) = \frac{p(B) + \delta}{2}$$

Then, for small enough $\delta > 0$,

$$p(A) = \frac{p(B)}{2} + \frac{\delta}{2} < p(B)$$

Elimination by aspects (EBA)

Alterative is a combination of aspects. C_t is a set of alternatives. A_t is a set of aspects that are present in at least one, but not all alternatives. At each step:

(i) aspect k is selected with probability $q(k, A_t) = \frac{\alpha_k}{\sum_{j \in A} \alpha_j}$ proportional to its weight α_k;
(ii) all alternatives without this aspect are eliminated, so we update $C_t \rightarrow C_{t+1}$.

The steps are repeated until a single alternative remains $|C| = 1$.

The probability of selecting a specific alternative is then the sum of probabilities of all distinct orderings of aspect elimination that result in the selected alternative being selected.

For example, for three alternatives with profiles

$$\begin{aligned} x_1 &= (1\ 0\ 1\ 0) \\ x_2 &= (1\ 0\ 0\ 1) \\ x_3 &= (0\ 1\ 0\ 0) \end{aligned}$$

alternative x_2 can be selected from initial alternatives set C by (a) eliminating on first aspect and then eliminating on fourth aspect, or (b) directly eliminating on fourth aspect (not possessed by either x_1 or x_3).

Then probability of selecting alternative x_2 from C is

$$p(x_2, C) = q(1, A_1) \cdot q(4, A_2) + q(4, A_1) = \frac{\alpha_1}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4} \cdot \frac{\alpha_4}{\alpha_3 + \alpha_4} + \frac{\alpha_4}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}$$

EBA does not assume order independence

Let us assume that weight of aspects α_3 and α_4 is insignificant compared to other aspects, and people are indifferent between x_1 and x_2.

Also assume

$$\alpha_1 = \alpha_2 + \epsilon$$

for small $\epsilon > 0$.

Then probability of selecting x_3 is

$$p(x_3, C) = \frac{\alpha_2}{\alpha_1 + \alpha_2} = \frac{\alpha_2}{2\alpha_2 + \epsilon}$$

And probability of selecting x_1 or x_2 is

$$p(x_1, C) = p(x_2, C) = \frac{1}{2}\left(1 - \frac{\alpha_2}{2\alpha_2 + \epsilon}\right) = \frac{1}{2}\left(\frac{\alpha_2 + \epsilon}{2\alpha_2 + \epsilon}\right) < p(x_3, C)$$

for small-enough $\epsilon > 0$.

If alternative x_2 is removed, then EBA probability of selecting $x_1$is

$$p(x_1, C) = \frac{\alpha_2 + \epsilon}{2\alpha_2 + \epsilon} > \frac{\alpha_2}{2\alpha_2 + \epsilon} = p(x_3, C)$$

so EBA can model order reversal.

What had been known before this paper came out

The above EBA formulation is consistent with a random utility model. However, how do we characterize this by an additive utility function over aspects, while ensuring that ordering of aspects remains fixed on a given choice occasion, but can change between choice occasions? How would errors enter this model?

Probabilistic lexicographic rule

Deterministic lexicographic rule eliminates alternatives using a sequences of attributes until only one alternative is left. Probabilistic lexicographic first uses a probabilistic mechanism to create the sequence. In particular, it assigns individuals' utility associated with aspect k,

$$u_k = v_k + \epsilon_k$$

with v_k as a deterministic component, and ϵ_k as a stochastic component that is extreme-value distributed with following density and cumulative distribution function.

$$f_x(\epsilon_k) = \exp(-\epsilon_k e^{-\epsilon_k})$$
$$F_x(\epsilon_k) = \exp\left(-e^{-\epsilon_k}\right)$$

$s = (k_1, \dots, k_n)$ denotes a sequence representing a preference ordering of the aspects for a person who prefers k_i to k_j $i \leq j$. It is formed based on aspect utilities sampled from distribution above. A lexicographic rule uses a sequence to choose an alternative from C by first eliminating alternatives that do not have aspect k_1, then eliminating those remaining alternatives that do not have aspect k_2, and so on, until only one alternative remains.

$$B_j = \left\{j \in C_{j-1} \middle| x_{ik_j} = 1\right\}$$

$$C_j = \begin{cases} C_{j-1} & if\ B_j = \emptyset\ or\ B_j = C_{j-1} \\ B_j & otherwise \end{cases}$$
$$t_j = \begin{cases} 0 & if\ B_j = \emptyset\ or\ B_j = C_{j-1} \\ 1 & otherwise \end{cases}$$

Stop if $\left|C_j\right| = 1$ and set $t_{j+1} = t_{j+2} = \cdots = t_n = 0$. t_j is an indicator for aspect of whether some but not all alternatives possess it.

Remarks: While attribute utilities are independent distributed, they are also perfectly correlated across elimination stages. Values of the attribute utilities are fixed across elimination stages on a particular choice occasion while it can be varied across different choice occasions.

This probabilistic model addresses uncertainty in individual attributes instead of uncertainty in total utility of an alternative, which correspond to consumer decision process according to Tversky, Sattah (1979).

Equivalence between EBD and lexicographic rule

Table 1. Choices associated with lexicographic sequences.

k_1	k_2	k_3	k_4	Choice
1	2	3	4	
1	3	2	4	x_1
1	3	4	2	
1	2	4	3	
1	4	2	3	x_2
1	4	3	2	
2	1	3	4	
2	1	4	3	
2	3	1	4	x_3
2	3	4	1	
2	4	1	3	
2	4	3	1	
3	1	2	4	
3	1	4	2	
3	2	1	4	x_1
3	2	4	1	
3	4	1	2	
3	4	2	1	
4	1	2	3	
4	1	3	2	
4	2	1	3	x_2
4	2	3	1	
4	3	1	2	
4	3	2	1	

Notes. x_1 and x_2 denote two different recordings of a Beethoven symphony; x_3 denotes a recording of a Debussy suite. All aspect orderings in a partition correspond to a set of sequences that satisfy the same precedence conditions resulting in the choice of the same alternative.

The key trick here is to notice that elimination of alternatives through a randomly generated sequence of aspects in lexicographic rule corresponds to an elimination of alternatives in a particular realization of choices under EBD. Moreover, all sequences that lead to the same choice under lexicographic rule enumerate all possible realizations under EBD that lead to the same choice. Thus, the probabilities of these events should be equal. Note that probability of a subset of sequences above giving precedence to a particular aspect is denoted as K_j, and its probability is $p(K_j)$, which is equal to a sum of rank-ordered logit probabilities of each sequence in the subset. Probability of each individual sequence is

$$p(s) = \prod_{j=1}^{n-1} \frac{\exp\left(v_{k_j}\right)}{\exp\left(v_{k_j}\right) + \cdots + \exp\left(v_{k_n}\right)}$$

See Example 4.

Insight: Thus, EBA is in fact a multinomial logit model (rank-ordered logit specifically). How does it avoid IIA? By modeling probabilities over aspects rather than alternatives (see p. 520).

Probabilistic utility function

Finally, we want to

1. Specify a function that assigns a utility to each alternatives in $\mathbb{C}$.
2. Specify a probability with which the alternatives in $\mathbb{C}$ simultaneously obtain these utility values.

Utility function of each alternative is defined by aspects:

$$u(x_i) = \frac{x_{ik_1}}{2^1} + \frac{x_{ik_2}}{2^2} + \cdots + \frac{x_{ik_n}}{2^n}$$

Link between the utilities of alternatives and the EBA choice probabilities:

$$p(x_i, C) = \Pr\left[u(x_i) = \max_{l \in C} u(x_l)\right]$$

Set of sequences where i is ranked first:

$$K(i) = \{s \in S \mid i = arg \max_{l \in C} u(x_l)\}$$

Then

$$p(x_i, C) = \sum_t p(K_t(i))$$

Where $p(K_t(i))$ is as define as sum of probabilities of member sequences. See Example 6.