Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER Applied Mathematics and Computation 161 (2005) 347–351
APPLIED MATHEMATICS AND COMPUTATION
www.elsevier.com/locate/amc

Maximum probability/entropy translating of contiguous categorical observations into frequencies ☆

M. Grendar Jr. [a,b,*], M. Grendar ✠

[a] *Institute of Mathematics and Computer Science of Mathematical Institute of Slovak Academy of Sciences (SAS) and of Matej Bel University, Severná ulica 5, 974 00 Banská Bystrica, Slovakia*
[b] *Institute of Measurement Science of SAS, Dúbravská cesta 9, 841 04 Bratislava, Slovakia*

Abstract

Maximum Probability method is used to translate possibly contiguous and overlapping categorical observations into frequencies.

Keywords: Maximum entropy method; Maximum probability method; Fuzzy categorical observations; Kullback–Leibler distance

1. Problem formulation

Consider the following (GK) problem (cf. [5]): As a result of a survey, there are n observations of an ordinal random variable which could take J different values (categories) from a set $\mathscr{X} \triangleq \{x_1, x_2, \ldots, x_J\}$. Contiguous grouping is allowed (i.e., respondents can select a range of contiguous categories). Given the observations, it is desired to calculate a measure which is based on frequency distribution (an entropy, for instance).

☆ The work was supported by the grant VEGA 1/0264/03 from the Scientific Grant Agency of the Slovak Republic. Gary King is thanked for posing the interesting problem. The thanks extend to George Judge and Viktor Witkowský.

* Corresponding author.
E-mail address: umergren@savba.sk (M. Grendar Jr.).
✠ Deceased.

0096-3003/$ - see front matter
doi:10.1016/j.amc.2003.12.024

Table 1
Feasible set of types

n_1	n_2	n_3	n_4
1	1	0	1
1	0	1	1
1	0	0	2
0	2	0	1
0	1	1	1
0	1	0	2
0	0	2	1
0	0	1	2

Clearly, if solely sharp responses were observed, the frequency-based measure would be directly computable, since the observations straightforwardly translate into frequencies. It is the presence of contiguous responses, which makes the problem interesting, since contiguous (and in this sense 'fuzzy') observations cannot be directly translated into frequencies.

There are several ways how to tackle the GK problem. One way—which we persuade here—is to recognize, that it falls into a class of under-determined inverse problems. The survey results define a feasible set of frequency vectors (called types in Information Theory, cf. [1]) from which it is necessary to pick up 'the best one', by some (preferably reasonable) selection scheme (criterion). Once the 'best' type is selected, the desired quantity (entropy of the type, for instance) can be calculated. Obviously, choice of the selection scheme needs a justification.

2. Example

Let $n = 3$, $J = 4$ and $\mathscr{X} = \{1, 2, 3, 4\}$. Let the following responses were observed: [1–3], [2–4], 4; i.e., the first responded answer was a fuzzy one: 'any number 1, 2 or 3'; the second answer laid in range 2–4 and the third one was sharp: 4.

The following are all possible sequences which conform with the observed responses: $\{1,2,4\}$, $\{1,3,4\}$, $\{1,4,4\}$, $\{2,2,4\}$, $\{2,3,4\}$, $\{2,4,4\}$, $\{3,2,4\}$, $\{3,3,4\}$, $\{3,4,4\}$ and all their permutations (since the order in which the responses were obtained is immaterial).

All types (i.e., frequency vectors) which could be based on the listed sequences are in the Table 1 (un-normalized [1]). These types define a feasible set of types $\mathscr{H}_n$ from which we would like to select one or more types by some selection scheme.

[1] Where it should cause no confusion, term type will be used for absolute vector of frequencies as well as for the proper, relative one.