Maximum Probability/Entropy translating of contiguous categorical observations into frequencies

M. Grendar, Jr. and M. Grendar

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 & Institute of Measurement Science of SAS, Dúbravská cesta 9, 841 04 Bratislava, Slovakia.
umergren@savba.sk

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

1 PROBLEM FORMULATION

Consider the following (GK) problem (cf. [6]): 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 $\mathcal{X} \triangleq \{x_1, x_2, \ldots, x_J\}$. Continguous grouping is allowed (i.e. respondents can select a range of continguous categories). Given the observations, it is desired to calculate a measure which is based on frequency distribution (an entropy, for instance).

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 (types) 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 $\mathcal{X} = \{1, 2, 3, 4\}$. 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 (unnormalized[1]).

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

These types define a feasible set of types $\mathcal{H}_n$ 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.

3 MAXIMUM PROBABILITY SELECTION RULE

Maximum Probability (MaxProb, cf. [3], [4]) selection rule prescribes to choose from a feasible set of types Π_n just the type(s) $\hat{\nu}_n$ which can be drawn from a (prior) generator $\mathbf{q}$ with the highest probability, i.e. $\hat{\nu}_n \triangleq \arg\max_{\nu_n \in \Pi_n} n! \prod_{i=1}^{J} \frac{{q_i}^{n_i}}{n_i!}$, where $\nu_n \triangleq [n_1, n_2, \ldots, n_J]/n$ and n_i is number of occurrences of i-th category in the random sample. The 'prior' generator is either selected on the ground of previous observations or is set to be uniform in the case of 'ignorance'.

3.1 Example (cont'd)

Let the generator be uniform one, $\mathbf{u}$ (1/4 each category). Probability that $\mathbf{u}$ will generate type [0 2 0 1] is 0.0469. Any of other four permutations of the type which are in the feasible set $\mathcal{H}_n$ can be of course generated with the same probability. Type [1 1 0 1] is generated with a higher probability 0.0938. With the same probability also types [1 0 1 1], [0 1 1 1] are generated. Hence, the MaxProb selects three types: [1 1 0 1]/3, [1 0 1 1]/3 and [0 1 1 1]/3.

3.2 Fuzzy categorical observations and inequalities on frequencies

It is worth observing that the fuzzy categorical observations can be equivalently expressed as inequalities on frequencies of the J categories; and vice versa: inequalities on frequencies can be equivalently stated as fuzzy categorical observations. This permits to restate GK problem as Boltzmann-Jaynes inverse problem (cf. [4]) with the feasible set defined by inequality constraints on frequencies.

Example (cont'd) For instance, the observations $[1-3]$, $[2-4]$ and 4 imply the following lower and upper bounds on frequencies of the categories: $\nu_1 \in [0, 1/3]$, $\nu_2 \in [0, 2/3]$, $\nu_3 \in [0, 2/3]$ and $\nu_4 \in [1/3, 2/3]$[2]. On the other hand, if the observations were given to us in the form of the upper and lower bounds on the frequencies, they could be uniquely transformed into categorical observations: $[1-3]$, $[2-4]$, 4.

[2] Alternatively, the bounds could be restated and reinterpreted as observed 'fuzzy' frequencies: $\nu_1 = [0-1]/3$, $\nu_2 = [0-2]/3$, $\nu_3 = [0-2]/3$ and $\nu_4 = [1-2]/3$.

3.3 What happens when n gets large?

Imagine that there were $n = 30$ observations, such that responses $[1-3]$, $[2-4]$ and 4 were observed 10 times each. Then there will be 1331 types which conform with the available observations. Among them, any of the following three types is generated by $\mathbf{u}$ with the highest probability (0.0035): [7 7 6 10]/30, [7 6 7 10]/30, [6 7 7 10]/30. Stated in the form of frequencies: [0.2333 0.2333 0.2000 0.3333], and the other two.

If there were $n = 60$ observations, such that the responses $[1-3]$, $[2-4]$ and 4 were observed 20 times each, then in the feasible set of types the most probable types (in the sense of coming from uniform generator) are the following three: [14 13 13 20]/60, [13 14 13 20]/60 and [13 13 14 20]/60. Stated as frequency vectors: [0.2333 0.2167 0.2167 0.3333] and the other two.

There is a visible tendency in the MaxProb types to equalize the first three frequencies, and hence to collapse into a single type, as n grows.

TABLE 2. Convergence of MaxProb types to I-projection.

n	$\hat{n}_1/n$	$\hat{n}_2/n$	$\hat{n}_3/n$	$\hat{n}_4/n$
3	0.3333	0.3333	0	0.3333
3	0.3333	0	0.3333	0.3333
3	0	0.3333	0.3333	0.3333
30	0.2333	0.2333	0.2000	0.3333
30	0.2333	0.2000	0.2333	0.3333
30	0.2000	0.2333	0.2333	0.3333
60	0.2333	0.2167	0.2167	0.3333
60	0.2167	0.2333	0.2167	0.3333
60	0.2167	0.2167	0.2333	0.3333
$\hat{\mathbf{p}}$	0.2222	0.2222	0.2222	0.3333

3.4 Maximum Probability Theorem

Indeed, Maximum Probability Theorem (MPT) states (see [3], [4]) that MaxProb type(s) converges to I-projection(s) of $\mathbf{q}$ on Π[3].

As it was noted above, the categorical observations can be translated into frequencies. They define feasible set of pmf's: $\mathcal{H} \triangleq \{\mathbf{p} : 0 \leqq p_1 \leqq 1/3, 0 \leqq p_2 \leqq 2/3, 0 \leqq p_3 \leqq 2/3, 1/3 \leqq p_4 \leqq 2/3\}$. I-projection of $\mathbf{u}$ on $\mathcal{H}$ is unique: $\hat{\mathbf{p}} = [0.2222\ 0.2222\ 0.2222\ 0.3333]$.

[3] Recall that I-projection $\hat{\mathbf{p}}$ of a probability mass function (pmf) $\mathbf{q}$ on a feasible set of pmf's Π is defined as $\hat{\mathbf{p}} \triangleq \arg\inf_{\mathbf{p}\in\Pi} \sum_{\mathcal{X}} p_i \log \frac{p_i}{q_i}$. Thus, I-projection is such a pmf in Π which has highest value of relative entropy $H(\mathbf{p}, \mathbf{q}) \triangleq -\sum_{\mathcal{X}} p_i \log \frac{p_i}{q_i}$.

Table 2 illustrates the convergence of MaxProb types to the I-projection.

Note, that the triplicity of MaxProb types is specific to samples of size $n = k \cdot 3$, $k = 1, 2, \ldots$. For instance, if samples of size of integer multiples of 10 were considered, then there would be unique MaxProb type for fixed n, — and the MaxProb type would be of course converging to the I-projection.

4 SUMMARY AND DISCUSSION

GK problem falls into category of under-determined inverse problems. At the same time it is an inherently probabilistic problem. Maximum Probability (MaxProb) method was used here to 'regularize' the problem; or in other words to select a frequency vector (type) from the feasible set of types $\mathcal{H}_n$, which is defined by contiguous categorical observations. It was noted, that the contiguous 'fuzzy' categorical observations can be equivalently expressed as lower and upper bounds on frequencies. Thus, the bounds on frequencies are alternative way of defining the feasible set of types. From the feasible set, MaxProb type(s) was/were selected. If n, the number of observations gets large, the task of MaxProb-type(s) selection becomes burdensome. Maximum Probability Theorem (MPT) comes in rescue, since it states that for n sufficiently large, MaxProb-type(s) becomes just the I-projection(s) on the corresponding feasible set of pmf's $\mathcal{H}$.

Naturally, one might ask why just MaxProb, and not some other way of 'regularization'. Answer to this question is provided by Maximum Probability Theorem and Conditioned Weak Law of Large Numbers (see [7], [2], [1])[4]. The first one states that MaxProb type(s) converges to I-projection(s)[5]. The second one claims that the I-projection is asymptotically conditionally only-possible probability distribution. So, any other selection scheme is from this point of view doomed to lead asymptotically into selecting a distribution which $\mathbf{q}$ generates with zero conditional probability.

REFERENCES

1. Cover, T. and Thomas, J., *Elements of Information Theory*, Wiley, 1991.
2. Csiszár, I., *Ann. Prob.*, **12**, No. 3, 768-793, (1984).
3. Grendár, M., Jr. and Grendár, M., "What is the question that MaxEnt answers? A probabilistic interpretation," CP 568, A. Mohammad-Djafari (ed.), 83-94, AIP, Melville (NY), 2001.
4. Grendár M., Jr., and Grendár, M., "Maximum Entropy method with non-linear moment constraints: challenges," Technical Report of IMS SAS, Jul 2003. Presented at MaxEnt 23. Available at http://uk.arxiv.org/abs/physics/0308006.
5. Grendár M., Jr., and Grendár, M., "Maximum Probability and Maximum Entropy method: Bayesian interpretation," Technical Report of IMS SAS, Jul 2003. Presneted at MaxEnt 23. Available at http://uk.arxiv.org/abs/physics/0308005.
6. King, G., Aug 2003.
7. Vasicek, O. A., *Ann. Prob.*, **8**, 142-147, (1980).

[4] In fact, a simpler and more direct argument can be used to answer 'Why MaxProb?' question. It will be presented soon.
[5] Note that MPT can be interpreted in a bayesian manner (cf. [5]).

ACKNOWLEDGEMENTS

The work was supported by the grant VEGA 1/7295/20 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.

Aug-Sep 2003

This is a Working Paper of IM&CS, Banská Bystrica, Sep 2003.