Determination the Different Categories of Buyers Based on the Jaynes' Information Principle

A. Maron[1], M. Maron[2]

Abstract:

Purpose: The article aims to reduce the volume of statistical data, necessary for determination the buyer's structure. The correct clustering of clients is important for successful activity for both commercial and non-profit organizations. This issue is devoted to a large number of studies. Their main mathematical apparatus is statistical methods. Input data are results of buyer polls. Polls are labor-consuming and quite often annoying buyers. The problem of determination of structure (various categories) of buyers by the mathematical methods demanding a small amount of these polls is relevant.

Design/Methodology/Approach: The approach offered in this report based on the Jaynes' information principle (principle of maximum entropy). Jaynes idea is as follows. Let us consider a system in which the conditions cannot be calculated or measured by an experiment. However, each state of the system has a certain measured implication, the average value of which is known (or can be defined), and the average result of these implications is known from the statistical data. Then the most objective are probabilities of states maximizing Shannon's entropy under restrictions imposed by information about average implications of states.

Findings: In this work the task of determination of percentage of buyers for computer shop by the average check is set and solved provided that average checks for each concrete category of buyers are known. Input data for calculation are their average checks. Determination of these values requires much less statistical data, than to directly determine relative number of buyers of various categories.

Practical Implications: The results are of particular interest to marketing experts.

Originality/Value: The article deals with practical situation when initially there are only three different groups of customers. For this case, the problem of maximizing entropy under given constraints reduced to the problem of finding a solution to a system of three equations, of which only one is nonlinear. This is a completely new result.

Keywords: Statistical Decision Theory, Principle of Maximum Entropy, Marketing.

JEL code: C44, M31.

Paper type: Research article.

[1]*Associate Prof., PhD., National Research University Higher School of Economics, Faculty of Business and Management, Moscow, Russia, email:* amaron@hse.ru
[2]*Lecturer, PhD., National Research University Higher School of Economics, Faculty of Business and Management, Moscow, Russia,* mmaron@hse.ru

1. Introduction and Literature Review

The correct clustering of clients is important for successful activity for both commercial, and non-profit organizations. This issue is devoted to a large number of studies. The general principles of clusterization are discussed in several works extensively (Punj and Stewart 1983; Ketchen and Shook, 1996; Everitt *et al.,* 2001; Dolnicar, 2003; Sheenu and Sakshi, 2014). As can be seen from the above list, this problem remains relevant, despite the results obtained. For trade, an important index is the average cost of all goods bought by one client at one visit of shop – an average check. The average check is depending on to what category the client it belongs. To determine the dependence of the average check on category of buyers also devoted the significant amount of works (Punj and Stewart, 1983; Ketchen and Shook, 1996; Everitt *et al.* 2001; Chib *et al.,* 2002.; Dolnicar, 2003; Chen *et al.,* 2004; Smith, 2004).

Interest in a problem of a clustering of clients and, in particular, buyers only increases in the light of new opportunities for big data processing (Loraine and Ashok, 2011; Raorane *et al.,* 2012; Herman *et al.,* 2013; Gupta and Mamtora, 2014; Kwak *et al.,* 2014; Duvvuri and Russell, 2015). Statistical methods are the main mathematical apparatus in the works stated above. Input data are results of buyer polls (Russell and Petersen, 2000). Polls are labor-consuming and quite often annoying buyers (Miller, 2014). The problem of determination of structure (various categories) of buyers by the mathematical methods demanding a small amount of these polls is relevant. Here the significant effect can be gained with the Jaynes' information principle. Initially, Jaynes E.T. offered the principle for the solution of the most important problems of statistical thermodynamics by a definition of the most probable power state of a system (Jaynes,1957). Furthermore Janes (2003) and other scientists have expanded the field of its application (Cheeseman *et al.,* 2000; Banavar *et al.,* 2010; Maron, 2011; Baggenstoss, 2017; Holmes, 2018).

Jaynes idea is as follows. Let us consider a system in which the conditions cannot be calculated or measured by an experiment. However, each state of the system has a certain measured implication, the average value of which is known (or can be defined), and the average result of these implications is known from the statistical data. Then the most objective are probabilities of states maximizing Shannon's entropy under restrictions imposed by information about average implications of states.

In this work the task of determination of percentage of buyers for computer shop by the average check is set and solved provided that average checks for each concrete category of buyers are known.

2. Substantial Problem Definition and the Algorithm of the Decision Rule

The shop is located in a student's campus of the large university. The shop sells computers, accessories to them and also office supplies of low, average and high price categories. The main buyers are students. All students are divided into three groups:

01.00.00 Natural Sciences; 02.00.00 Engineering and Technology; 05.00.00 Social Sciences. The number of the students studying in various fields is known. It is established that consumer preferences of students of various groups significantly differ. The management of shop wants to know the share of students of each direction in total number of buyers. It is necessary both for formation of the product range, and for the exact organization of the advertizing company (Corstjens and Lal, 2000; Moon and Russell, 2008; Briesch *et al.,* 2009; Miller, 2014). At the same time, for the definition of required shares it is necessary to conduct survey of a large number of respondents, since we are talking essentially about determining probabilities. It is possible to reduce costs of polls having executed the following:

➢ To determine average checks by each group. These polls won't demand a large number of respondents as averages were defined. Let's remind that Average check is measured as the total number of sales divided by the number of buyers.
➢ To define an average check in a month.
➢ To calculate required shares of buyers of various categories based on the Jaynes' information principle.

3. Mathematical Statement and Solution of a Task

There are three categories of buyers: K_1, K_2, K_3. Their average (mean) checks in dollars are: c_1; c_2; c_3. Average check on shop in dollars – C. It is required to define relative shares of buyers: k_1, k_2, k_3.

The average check is defined by expression (1):

$$C = k_1 \cdot c_1 + k_2 \cdot c_2 + k_3 \cdot c_3 \qquad (1)$$

At the same time

$$k_1 + k_2 + k_3 = 1 \qquad (2)$$

The share of k_1 of buyers of category K_1 is equivalent, at a big flow of buyers, the probability that accidentally chosen buyer belongs to this category. The similar statement is fair for k_2 and k_3. According to Jaynes' principle, the most objective are those probabilities for which, taking into account (1) and (2), Shannon's entropy reaches its maximum:

$$H(k_1, k_2, k_3) = -k_1 \cdot \ln k_1 - k_2 \cdot \ln k_2 - k_3 \cdot \ln k_3 \rightarrow \max \qquad (3)$$

We will apply the method of Lagrange multipliers to find the maximum of the nonlinear function (3) under restrictions (1) and (2). After appropriate transformations, allowing exclusion of uncertain Lagrange multipliers, we discover

that determining the probabilities is reduced to the solution of the system that consists of (1), (2) and the following equation:

$$k_1^{c_2-c_3} \cdot k_2^{c_3-c_1} \cdot k_3^{c_1-c_2} = 1 \tag{4}$$

The system of the equations (1), (2) and (4) can be solved by methods included in the Mathcad Prime package.

4. Example

The results of the study were applied in practice. Let us give an example of calculations with real numbers, but without disclosing the names of the organizations. Average values of checks for the groups of buyers are: $c_1 = 80$ \$; $c_2 = 150$ \$; $c_3 = 58$ \$. Average check on shop 90 \$. Calculation of buyers groups relative share are shown below:

$$c_1 := 80 \qquad c_2 := 150 \qquad c_3 := 58 \qquad C := 90$$
$$k_1 := 0.5 \qquad k_2 := 0.5 \qquad k_3 := 0.5$$

$$k_1^{(c_2-c_3)} \cdot k_2^{(c_3-c_1)} \cdot k_3^{(c_1-c_2)} = 1$$

$$k_1 + k_2 + k_3 = 1$$

$$k_1 \cdot c_1 + k_2 \cdot c_2 + k_3 \cdot c_3 = C$$

$$\text{find}\,(k_1, k_2, k_3) = \begin{bmatrix} 0.352 \\ 0.264 \\ 0.385 \end{bmatrix}$$

It was known that the relative of the number of students in the campus are as follows: Natural Sciences -20%; Engineering and Technology – 50%; Social Sciences – 30%. After calculation it became apparent that the second group is captured insufficiently. Additional resources were allocated for engaging of buyers from this group.

5. Possible Extension of the Results

It is possible to extend the result to the case when the number of options (n) is more than three. This will significantly expand the scope. For example, conduct a reasonable assessment of the quality of work of the headquarters of candidates for elected posts. For them, the task of mobilizing the electorate is urgent. Important factors for clustering the electorate are age, gender, and education. When clustering conducted, the polls conducted to know the preferences of the voters of each group.

Based on this, mobilization of the electorate is organized. After the election, the percentages of the votes cast for the candidates known. Having these data, it is possible for each group of voters to determine the percentage of voters that take part in election.

The Janes' principle will be a tool to solve this problem. This will make it possible assess the quality of work of candidate headquarters.

Another example does not apply to post-factum assessment, but to planning. In the allocation of funds allocated to reduce the consequences of accidents in railways transport, the decision maker evaluates the severity of the consequences. For each specific situation the multiplication of probability by the price of losses is taken as such an estimate. As a rule, there are no statistics to determine the probabilities. The best option is to calculate them based on Janes 'principle. Then we get the following calculation scheme. Set a valid average loss value. It is equal to the sum of the multiplication of the probabilities of situations at the corresponding prices. Based on Janes 'principle calculate the least biased probabilities. Determine the minimum costs at which the average value of losses is acceptable.

The application of the Jaynes' information principle to project analysis tasks can be extremely valuable. By definition, each project is a unique sequence of work. At the same time, probability theory is oriented towards mass homogeneous random phenomena. Strictly speaking, this makes the methods of probability theory indispensable for the analysis of projects.

However, there are typical projects. For example, the construction of model houses. Of course, the scope of work when building a house on a mountain is different from the scope of work when building a house on a plain. But, many works are common. For typical projects, methods of the theory of probability are applicable, with careful use. However, collecting statistics is always difficult. So, to solve the problems of project diagnostics, it is necessary to know the probability distribution of errors in the work. It is known that a statistical calculation of probabilities requires a lot of observations. Even for typical projects, collecting this amount of data is almost impossible. However, for calculation this probabilities based on Janes 'information principle we need only data about mean values of consequences of errors.

6. Conclusion

In this article the new effective method is offered for the determination of different categories of buyers based on the Jaynes' information principle. Having carried out necessary calculations, it is possible to start formation of the necessary range and scheduling the advertising campaign. The result is of special interest to marketing specialists, and owners of different shops as the offered method is simple in realization.

References:

Baggenstoss, P.M. 2017. Uniform Manifold Sampling (UMS) Sampling the Maximum
        Entropy PDF. IEEE, Trans. Signal Process. 65, 2455-2470.
Banavar, J.R., Maritan, A., Volkov, I. 2010. Applications of the principle of maximum

entropy: From physics to ecology. J. Phys. Condens. Matter, 22p.

Briesch, R.A., Chintagunta, P.K., Fox, E.J. 2009. How does assortment affect grocery store choice? Journal of Marketing Research, 46, 176-189.

Cheeseman, P., Loredo, T., Skilling, J., Stutz, J., Smelyanskiy, V. 2000. On Bayesian inference and Maximum Entropy NASA. Ames Research Center, working paper.

Chen, Y.L., Tang, K., Shen, R.J., Hu, Y.H. 2004. Market basket analysis in a multiple store environment. Decision Support Systems, 40(2), 339-335.

Chib, S., Seetharaman P.B., Strijnev A. 2002. Analysis of multicategory purchase incidence decisions using IRI market basket data. Advances in Econometrics 16, 55-90.

Corstjens, M., Lal, R. 2000. Building store loyalty through store brands. Journal of Marketing Research 37, 281-291.

Dolnicar, S. 2003. Using Cluster Analysis for Market Segmentation: Typical Misconceptions, Established Methodological Weaknesses and Some Recommendations for Improvement. Journal of Marketing Research, 11(2), 5-12.

Everitt, B.S., Landau, S., Leese, M. 2001. Cluster Analysis, London Arnold.

Gupta S., Mamtora, R. 2014. A Survey on Association Rule Mining in Market Basket Analysis. International Journal of Information and Computationa Technology, 4(4), 409-414.

Herman, A, Forcum, L.E, Harry, J. 2013. Using Market Basket Analysis in Management Research. Journal of Management 39(7), 1799-1824.

Holmes, D.E. 2018. Entropy, Special Issue on Maximum Entropy and Its Application. MDPI Basel, Switzerland.

Jaynes, E.T. 2003. Probability theory. The logic of science. Cambridge University Press. 724p.

Jaynes, E.T. 1957. Information Theory and Statistical Mechanics. I. Phys. Rev. 106, 620-630.

Ketchen, D.J., Shook, C.L. 1996. The Application of Cluster Analysis in Strategic Management Research an Analysis and Critique. Strategic Management Journal, 17(6), 441-458.

Kwak, K., Duvvuri S.D., Russell, G.J. 2015. An analysis of assortment choice in grocery retailing. Journal of Retailing, 91, 19-33.

Loraine C., Ashok, K. 2011. Frequent Item set mining for Market Basket Data using K-Apriori algorithm. International Journal of Computational Intelligence and Informatics, Vol. 1, No. 1, 14-18.

Maron, V.I. 2011. Statisticheskie modeli na osnove informatsionnogo podhoda Dzheynsa (Statistical models based on Jaynes' information principle). Moscow. Max Press (in Russian).

Miller, T.W. 2014. Modeling Techniques in Predictive Analytics Business Problems and Solutions. Revised and Expanded Edition. PH Professional Business.

Moon, S., Russell, G. 2008. Predicting product purchase from inferred customer similarity an autologistic model approach. Management Science, 54, 71-82.

Punj, G., Stewart, D.W. 1983. Cluster Analysis in Marketing Research Review and Suggestions for Application. Journal of Marketing Research, 20(2), 134-148.

Raorane, A.A., Kulkarni, R.V., Jitkar, B.D. 2012. Association Rule – Extracting Knowledge Using Market Basket Analysis. Research Journal of Recent Sciences, 1(2), 19-27.

Russell, G.J., Petersen, A. 2000. Analysis of cross category dependence in market basket selection. Journal of Retailing 76, 367-392.

Smith, H. 2004. Supermarket choice and supermarket competition in market equilibrium. The Review of Economic Studies, 71, 235-263.

Sheenu V., Sakshi, B. 2014. An Effective Dynamic Unsupervised Clustering Algorithmic Approach for Market Basket Analysis. International Journal of Enterprise Computing and Business Systems, 4(2), 16-32.