Proceedings of the 9th WSEAS International Conference on Applied Mathematics, Istanbul, Turkey, May 27-29, 2006 (pp436-440)

Estimation by MinxEnt Principle

ALADDIN SHAMILOV* , ZERRIN ASAN* and CIGDEM GIRIFTINOGLU*

*Anadolu University
Faculty of Science
Department of Statistics
26470 Eskisehir
TURKEY

Abstract - In this study, the proportions of electricity consumption of Turkey in respect of some industrial sectors are estimated by using Kullback's Minimum Cross-Entropy Principle which is a method of deriving probability distribution. This estimation problem is considered as an entropic optimization problem. The problem is transformed to a functional equation by Lagrange's multipliers method and solved by Newton's Method. A Visual Basic program is written for the whole of the solution process in order to estimate the proportions. Furthermore, these estimated proportions are compared to observed proportions and the results are interpreted.

Keywords: Kullback-Leibler Entropy Measure, Kullback's Minimum Cross-Entropy Principle, Estimation of Probability Distribution.

1 Introduction

The term entropy as a scientific concept was first used in thermodynamics (Clausius 1850). Its probabilistic interpretation in the context of statistical mechanics is attributed to Boltzmann (1877). Although it is originated in the literature on thermodynamics, its usage has penetrated almost all disciplines, mainly because of its association with the concept of information as envisaged by Claude Shannon in the mathematical theory of communication [13]. Jaynes [8] developed the maximum entropy principle (MaxEnt) which is a method of deriving a probability distribution when information about a statistical problem is given in terms of some averages, such as the mean, variance, etc. Thus, MaxEnt has been successfully invoked in problems arising out of several disciplines such as engineering and physics. Kullback [10;5] exposed minimum cross-entropy principle which rests on completely different bases although its relationship with MaxEnt. These optimization principles are widely used in many studies and disciplines like statistics, thermodynamics, statistical mechanics [1;2;3;4;7;15]. Especially Kapur and Kesavan [6;11;12] have dealt with the applications to pattern recognition, spectral analysis, queuing theory, parameter estimation, portfolio analysis, transportation and analysis of information in contingency tables etc. Whereas there are such applications of MinxEnt principle, in this study, we taken into consideration that this principle can also be used in estimation of the proportions of electricity consumption of Turkey in respect of by sectors when an estimated mean is given as a constraint.

2 Kullback's Minimum Cross-Entropy Principle

Kulback's minimum cross-entropy principle (MinxEnt) is an entropy optimization principle. This principle is also known as *the minimum directed divergence principle* or *the minimum discrimination information principle*. The information used here is in the form of prescribed values of linear moments of the probability distribution.

Let a random variate X take values $x_1, x_2, \ldots, x_n$ with the corresponding probabilities $p_1, p_2, \ldots, p_n$. The probabilities $p_1, p_2, \ldots, p_n$ may be used to find the mean, variance, and other algebraic moments of X.

The natural relation is

$$\sum_{i=1}^{n} p_i = 1, \qquad (1)$$

and the expected values of the functions $g_1(X), g_2(X), \ldots, g_m(X)$ may be found as follows:

$$\sum_{i=1}^{n} p_i g_r(x_i) = \eta_r, \quad r = 1,2,\ldots,m. \qquad (2)$$

Particularly, these functions are chosen in the form of algebraic moments and they are expressed as the following:

$$\sum_{i=1}^{n} p_i x_i^r = \eta_r, \qquad r = 1,2,...,m. \qquad (3)$$

Before demonstrating the idea and mathematical base of the mentioned principle, the following three concepts are discussed:

- Probabilistic distance or cross-entropy of P with respect to Q and Kullback-Leibler (K-L) measure,
- An a priori probability distribution,
- Minimization of the K-L measure subject to the specified linear moment constraints.

The emphasis in MinxEnt is on the concept of "cross-entropy" or "directed divergence" of a probability distribution P from another probability distribution Q. In order to describe the directed divergence the K-L measure should be used. In the statistics literature this term arises as expected value of logarithm of likelihood ratio.

Let $p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$ be two probability distributions such that $p = (p_1, p_2, \ldots, p_n)$ satisfies the given constraints and $q = (q_1, q_2, \ldots, q_n)$ is the given distribution based on experience, intuition, or theory. Then, K-L measure is defined as

$$D(p:q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \qquad (4)$$

where q_i's are called a priori probability distribution.

Compatible distributions are numerous when some mean values are given as constraints. Based on minimum cross-entropy principle, Kullback suggested that out of all probability distributions satisfying the given constraints, choose the distribution which is closest to the given a priori distribution according to K-L Measure. As it is seen easily, this is a constrained optimization problem. That is, mentioned Kullback's MinxEnt principle consists of minimizing (4) subject to the constraints (1) and (3) when a priori distribution q is given.

By taking $r = 1$, the constraint (3) is transformed to

$$\sum_{i=1}^{n} p_i x_i = \eta. \qquad (5)$$

Lagrange's method would be preferred for solving this constrained optimization problem. Thus Lagrange's auxiliary function is constructed as the following:

$$L \equiv \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} + \lambda_0 \sum_{i=1}^{n} (p_i - 1) + \lambda_1 \sum_{i=1}^{n} (x_i p_i - \eta).$$

If we differentiate the function L with respect to p_i's and equate to zero then we get

$$\frac{\partial L}{\partial p_i} = \left(\ln \frac{p_i}{q_i} + \frac{1}{p_i} p_i \right) + \lambda_0 + \lambda_1 x_i = 0,$$

$$\ln \frac{p_i}{q_i} = -\lambda_1 x_i - \lambda_0 - 1,$$

$$p_i = q_i \left(e^{-\lambda_1 x_i} e^{-\lambda_0 - 1} \right) \qquad (6)$$

where λ_0 and λ_1 may be determined by using the constraints (1) and (5). In order to remove λ_0, the following process is pursued:

In Eq. (6), let $e^{-\lambda_0 - 1} = A$ and $e^{-\lambda_1} = w$. By substituting these values into (6),

$$p_i = q_i w^{x_i} A \qquad (7)$$

is obtained. If (7) is taken into account in constraint (2), then

$$\sum_{i=1}^{n} q_i w^{x_i} A = 1,$$

and A can be expressed in terms of w.

Indeed,

$$A = \frac{1}{\sum_{i=1}^{n} q_i w^{x_i}} . \qquad (8)$$

If (8) is put into (7), then p_i's are found in dependence on only w as

$$p_i = \frac{q_i w^{x_i}}{\sum_{i=1}^{n} q_i w^{x_i}} \qquad i = 1,2,\ldots,n. \qquad (9)$$

Putting (9) into constraint (5), we have

$$\frac{\sum_{i=1}^{8} x_i q_i w^{x_i}}{\sum_{i=1}^{8} q_i w^{x_i}} = \eta \qquad (10)$$

or

$$g(w) = \sum_{i=1}^{n} q_i (x_i - \eta) w^{x_i} = 0. \qquad (11)$$

In this study, a Visual Basic program is written. The program constructs Eq.(11), calculates p_i's expressed by (9) and finds the root w by using Newton's method when values of x_i's, q_i's and η are given.

3 An Application of MinxEnt Principle on the Electricity Consumption

3.1 Importance of Estimation of Electricity Consumption Proportions

All around the world, energy is one of the requirements of the economical and social development and it should be provided reliably on time. As well, there is a close relationship between energy consumption and economical activity. In other words, national income increases as energy consumption becomes greater. This shows that energy consumption is one of the most important components of national economy. Furthermore, energy consumption per capita is also added to the criteria which show development level of the countries.

After the industrial revolution, usage of electrical appliances and machines has been widespread due to the mechanization in agriculture and industry sectors. So, the demand of electricity energy rose. It is crucial that the demand can be afforded. Not only insufficiency of electricity energy means to stay in the dark, but also production is getting fewer, unemployment increases and all of the industry sectors are paralyzed. Briefly, it means that national income decreases. For this reason, the establishments of electricity production and distribution have to go forward in direction of determined marks. If these peculiarities are taken into account, the importance of distributing electricity to the sectors also arises. These proportions give information about the trend of the country's economy.

In this study, the data sets are taken from [14] in order to estimate the energy consumption proportions. These data sets consist of quantity, sale value and proportions of electricity consumption of the first period in 2004 according to the sectors. Table 1 shows the energy consumption of administrative building, industrial establishment with auto producer, trading establishment, dwelling, irrigate, construction, street lighting and other sectors.

Table 1 Electricity consumption of the sectors in the first period of 2004.

	2004			
	I.Period (October,November.December)			
Sectors	**Quantity (Gwh)**	**Value (YTL.)**	**Unit Price**	**%**
Total	27495	2735348167	**99485**	**100**
Ind. Est. +auto producer	12732	742272918	**58300**	**46.31**
Street lighting	934	94134586	**100786**	**3.40**
Other	2073	243555399	**117489**	**7.54**
Irrigate	97	12533058	**129207**	**0.35**
Dwelling	6757	879293525	**130131**	**24.58**
Admin. building	1322	179995358	**136154**	**4.81**
Const.	298	47094299	**158035**	**1.08**
Trading est.	3282	536469024	**163458**	**11.93**

The forth column of Table 1 shows the unit prices obtained by dividing the third column to the second column. We have also proportions of electricity consumption of the sectors in the same year. We want to make use of this information for the following year and we can estimate it via MinxEnt principle. In order to apply this principle, we consider these consumption proportions as $q=(q_1,q_2,\ldots,q_8)$ a priori distribution and the unit prices per the sectors as $X=(x_1,x_2,\ldots,x_8)$ values. In addition, we take the average unit price as mean constraint for the first period in 2005. This mean constraint can be obtained by using various methods such as ARIMA which is mostly used in time series analyses. However, in this study the average unit price observed in the first period of 2005 was used. Thus, in the constraint (5), the value of η is determined as 107830 monetary unit. If the problem is reviewed again, numeric values of n, x_i, q_i and η can be written as follows:

- $n=8$,
- $x_i=($58300, 100786, 117489, 129207, 130131, 136154, 158035, 163458),
- $q_i=($ 0.4631, 0.0340, 0.0754, 0.0035, 0.2458, 0.0481, 0.0108, 0.1193),
- $\eta=107831$.

The proportions of electricity consumption of eight sectors are obtained for the first period in 2005 using the method mentioned above. Observed and estimated values are compared in Table 2.

Table 2 Comparison of observed and estimated proportions

	2005 I. Period	**2005 I.period**
Sector	**Observed**	**Estimated**
Industrial. Estab. +auto prod.	0.4215	0.3752
Street lighting	0.0322	0.0298
Other	0.0927	0.0718
Irrigate	0.0041	0.0036
Dwelling	0.2599	0.2757
Adm. building	0.0465	0.0585
Construction	0.0122	0.0142
Trading establishment	0.1309	0.1708

It is seen that observed proportions are very close to the estimated proportions. So, this is an evidence case which show that this principle obtains successful results and can be used for estimation problems.

4 Conclusion

According to the information theory, uncertainty decreases about a system as information increases. From this point of view, it is very important to make use of available information. MinxEnt principle enables us to use the available information. If we have information in the form of some mean values and a priori distribution based on experience, intuition, or theory about a system, then the usage of MinxEnt principle becomes convenient. The foundation of this principle rests on a very simple idea**:** Out of all probability distributions satisfying the given mean values, choose the distribution that is closest to the a priori distribution. The mentioned closeness in the sense of probabilistic distance is expressed by K-L measure. Thus, it is required to minimize K-L measure subject to the given constraint as mean value.

In this study, for the first period in 2004 electricity consumption proportions as a priori distribution and the unit prices concerning per using sectors as X values are considered. Besides, average unit price of the first period of 2005 is taken as mean constraint and this mean constraint is supposed to be given. The distribution compatible with the given mean value is found by means of the MinxEnt Principle. However, estimation of this value by statistical methods can be thought as further research related to this paper. Likewise, as a development of this study, proportions of electricity consumption of the sectors for the first period of 2006 which are not declared yet can be estimated.

References:

[1] Papoulis, A., *Probability, Random* variables, and stochastic processes, McGraw-Hill, Singapore, USE, 1991.

[2] Cover, T.M. and Thomas, J.A., Elements of Information Theory, A Wiley-Interscience Publication, New York, USA, 1992.

[3] Golan, A., Judge, G. and Miller, D., Maximum Entropy Econometrics: Robust Estimation with Limited Data, John Wiley & Sons, New York, USA, 1996.

[4] Karmeshu and Pal N.R., Entropy Measures, Maximum Entropy Principle and Emerging Applications, Springer, New York, USA, 2003.

[5] Kullback, S., Information Theory and Statistics, Dover Publications, New York, USA, 1968.

[6] Kapur, J.N. and Kesavan, H.K., Entropy Optimization Principles with Applications, Academic Press, New York, USA, 1992.

[7] Wu, N.,The Maximum Entropy Method, Springer, Berlin, Germany 1997.

[8] Jaynes, E.T. “Information Theory and Statistical Mechanics.” *Physical Reviews*,106 : 620-630, 1957.

[10] Kullback, S. and Leibler, R.A. “On Information and Sufficiency.” Ann. Math. Stat., 22: 79-86, 1951.

[11] Kapur, J.N. and Kesavan, H.K., “Inverse MaxEnt and MinxEnt Principles and their Applications.” In PaulF. Fougere, editor, *Maximum Entropy and Bayesian Methods*, pp. 433-450, Kluwer Academic Publishers, New York, 1990.

[12] Kesavan, H.K. and Kapur, J.N., “Maximum Entropy and Minimum Cross-Entropy Principles: Need for a Broader Perspective.” In Paul F. Fougere, editor, *Maximum Entropy and Bayesian Methods*, pp. 419-432, Kluwer Academic Publishers, New York, 1990.

[13] Shannon, C.E., “A Mathematical Theory of Communication.” *Bell System Tech. J.*, 27: 379-423, 623-659, 1948.

[14]http://www.die.gov.tr/TURKISH/SONIST/ENERJI/enerji.html

[15] Shamilov A., *Comparision of Classical Estimate Methods Based On MaximumEntropy Distribution*, Ordered statistical data: Approximation, Bounds and characterizations, Izmir, Turkey, 2005.