Determining Probability Distribution by Minimum Cross Entropy Method

ALADDIN SHAMILOV*
ÇİGDEM GRIFTINOGLU*
ILHAN USTA*
YELIZ MERT KANTAR*
*Anadolu University
Science and Art Faculty-Department of Statistics
16470-Eskisehir
TURKEY

Abstract: This paper presents an overview of minimum cross entropy method, which has been used extensively in many areas such as signal/image processing, pattern recognition and statistical inference. In application, we are mainly concerned with the problem of determining probability distributions of the number of divorces by age group measured in 2001 by using minimum cross entropy principle. Minimum cross entropy distributions are found for both husband and wife individuals.

Key words:- Kullback-Leibler measure, MinxEnt principle, Model-selection criteria.

1 Introduction

Minimum cross-entropy (MinxEnt) method has been applied to a wide range of problems in image processing, pattern recognition, economics, signal process and spectral analysis [6-9]. When the prior distribution is uniform, MinxEnt is equivalent to MaxEnt method. There have been a lot of studies about the application of these methods in literature. [9]. In this study, we use MinxEnt method to find probability distribution of the number of divorces.

This study consists of the following section. Section 2 introduces MinxEnt method. In same section, in order to show performance of MinxEnt, model selection criteria are given. In section 3, MinxEnt distribution of the number of divorces by age group is found. In section 4, the main results obtained from the study are summarized.

2 Minimum cross-entropy (MinxEnt) Method

Kullback (1959) presented the principle of minimum cross entropy which states that given a prior distribution *f*(*x*), choose that distribution from a set of candidate distributions which minimizes the cross-entropy and satisfies specified moment constraints

Given the moment constraints only, it is proved that the cross-entropy minimization is a uniquely correct method of probabilistic inference that satisfies all the consistency axioms.

In MixEnt principle, the object is to find a probability distribution $f(x)$ which is satisfying the given moment constraints, such that K-L measure to an a priori probability distribution $q(x)$ is a minimum.

$$D(f(x);q(x)) = \int_a^b f(x)\ln\frac{f(x)}{q(x)}dx, \qquad (1)$$

subject to constraint

$$\int_a^b g_j(x)f(x)dx = \mu_j\,, \qquad (2)$$

where $g_j(x)$ are moment vector functions, μ_j are moment vector-values.

Auxiliary functional corresponding to problem (1)-(2) is

$$\int_a^b \underbrace{\left(f(x)\ln\frac{f(x)}{q(x)}dx - \lambda_0 f(x) - \sum_{j=1}^{m}\lambda_j g_j(x) f(x)\right)}_{F} dx = \mu \qquad (3)$$

According to following Euler-Lagrange equation

$$\frac{\partial F}{\partial f(x)} - \frac{d}{dx}\left(\frac{\partial F}{\partial f'(x)}\right) = 0 \qquad (4)$$

the $f(x)$ which minimize (3).

$$\ln\frac{f(x)}{q(x)} + 1 - \lambda_0 - \sum_{j=1}^{m}\lambda_j g_j(x) = 0, \qquad (5)$$

Use of the Euler-Lagrange equation of the calculus of variations gives:

$$f(x) = q(x)\exp(\lambda_0 + \sum_{j=1}^{m}\lambda_j g_j(x)). \qquad (6)$$

(6) is called as MinxEnt distribution in [1]. In here, $\lambda_0, \lambda_1, ..., \lambda_k$ the Lagrange multiplier can be found Newton method.

3.1 Suitability Judgment Criteria

In order to show performance of MinxEnt probability distributions to model the number of divergence, we use various statistical criteria. Root mean square error (*RMSE*), Correlation coefficient (R^2) are used in statistically evaluating the performance of MinxEnt distributions.

The formula of mentioned suitability judgment criteria are

$$RMSE = \left(\frac{\sum_{i=1}^{N}(y_i - x_i)^2}{N} \right)^{\frac{1}{2}}, \qquad (7)$$

$$R^2 = \left(1 - \frac{\sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(y_i - z_i)^2} \right), \qquad (8)$$

where y_i is the *i*th probability of actual data, x_i is the *i*th predicted probability, *N* is number of all observed data, *n* is the number of parameters or the number of constrains.

The best distribution function can be determined according to the lowest values *RMSE*, the highest values R^2.

3. MinxEnt distribution of the number of divorces by age group

In this section, the distribution of the number of divorces by age group measured in 2001 is obtained by MinxEnt principle. In calculating the MinxEnt distribution for 2001, the prior distribution is taken as the distributions of 2000.

In this method, the determination of prior distribution is important part of information needed in evaluation distribution. In such a case, our best choice is to select the prior probability distribution among the distributions satisfy prior information. So, we select the prior distribution as the distributions of 2000.

In the Table 1, Table 2, f_{obs} is probability density function of observed data, f_{MxE} is MinxEnt distribution

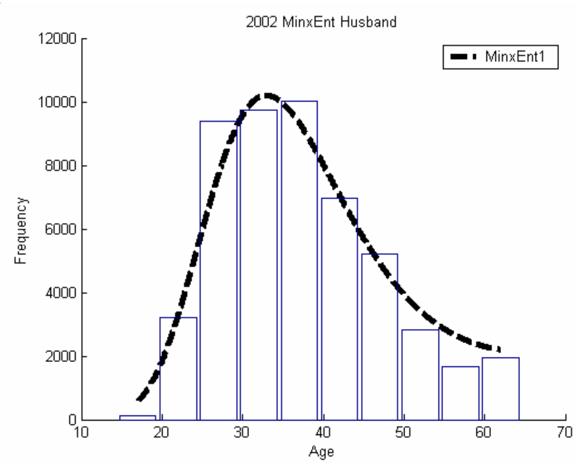
Table 1 MinxEnt probability distribution and observed data

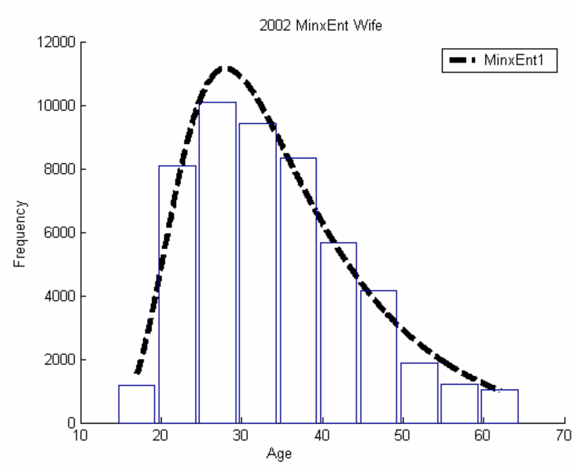
Husband individuals				
Class interval	Observed data (2001)	Observed data (2002)	f_{obs}	f_{MxE}
15-19	244	139	0.0027	0.0041
20-24	3622	3198	0.0626	0.0648
25-29	9666	9403	0.1840	0.1806
30-34	1050	9724	0.1903	0.2034
35-39	9107	10001	0.1957	0.1820
40-44	6555	6973	0.1365	0.1346
45-49	4455	5212	0.1020	0.0937
50-54	2646	2811	0.0550	0.0569
55-59	1426	1680	0.0329	0.0313
60+	2181	1955	0.0383	0.0487
RMSE				0.0075
R^2				0.98793

Normally, a value higher than 70% of R^2 is acceptable [8]. The parameters for the statistical analysis: the R^2, RMSE and Chi-square error are given in Table1 and Table 2 for all data. It can be seen that the R^2 values are bigger than 0.98 for all data if MinxEnt density functions subject to one moment constraints (f_{MxE}) are used. Also from these tables, f_{MxE} shows better fitting in terms of RMSE, R^2. The other words, Table1-2 and Figure 1-2 show the excellent agreement of the MinxEnt densities with data. For this reason the MinxEnt distributions are much convenient

Table 2 MinxEnt probability distribution and observed data

Wife individuals				
Class interval	Observed data (2001)	Observed data (2002)	f_{obs}	f_{MxE}
15-19	1620	1175	0.0027	0.0041
20-24	8645	8091	0.0626	0.0648
25-29	10751	10098	0.1840	0.1806
30-34	9500	9437	0.1903	0.2034
35-39	7440	8329	0.1957	0.1820
40-44	5172	5661	0.1365	0.1346
45-49	3395	4168	0.1020	0.0937
50-54	1918	1885	0.0550	0.0569
55-59	854	1203	0.0329	0.0313
60+	1107	1049	0.0383	0.0487
RMSE				0.0075
R^2				0.99378





4 Conclusions

This paper presents an overview of minimum cross entropy method. The power of Minimum cross entropy method are illustrated in determining probability distributions of the number of divorces by age group

References:

[1] Jaynes, E. T. Papers on Probability, Statistics, and Statistical Physics. Netherlands: D. Reidel, Dordrecht, 1983.

[2] S. Kullback and R. A. Leibler, On Information and Sufficiency, Ann. Math. Stat., Vol.22, pp. 79-86, 1961

[3] Kullback, S. *Information Theory and Statistics*. Wiley, 1959.

[4] Kapur, J.N. and Kesavan, H. K. (1992), *Entropy Optimization Principles with Applications*. London: Academic Press, Inc.

[5] Wu, X. Calculation of maximum entropy densities with application to income distribution, *Journal of Econometrics*, 2003, 115, 347-354.

[6] Li M., Li X. Investigation of wind characteristics and assessment of wind energy potential for Waterloo region, *Energy conversion management*, 46, 2005, 3014-3033

[7] Srikanth M, Kesavan H, Peter H. Probability density function estimation using the minmax measure. IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews, 2000, Vol.30, No.1, pp.77-83

[8] Celik A.N. A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey, *Renewable energy*, 29, 2003, 593-604.

[9]Pandey M.D. (2001), "Minimum cross-entropy method for extreme value estimation using peaks-over-threshold data," *Structural Safety* , 23 (4**)** , 345-363.

[10] Liou, C.-Y. A separable cross-entropy approach to power spectral estimation Musicus, B.R.;Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on Volume 38, Issue 1, Jan. 1990 Page(s):105 - 113

[11] Zaifei Liu Doucet, A., Singh S.S. The cross-entropy method for blind multiuser detection Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on27 June-2 July 2004 Page(s):510

[12] Jun-Yi Xu; Jian Yang; Ying-Ning Peng; Chao Wang; Using cross-entropy for polarimetric SAR

image classification, Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International Volume 3, 24-28 June 2002 Page(s):1917 - 1919 vol.3

[13] S. Kullback, Information Theory and Statistics, Wiley, New York, 1959, p.37

J.E. and Johnson R.W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross entropy. IEEE Trans. on Information Theory, 26(1), 26-37.