Available online at www.alphanumericjournal.com

alphanumeric journal

The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

Volume 3, Issue 1, 2015

2015.03.01.STAT.01

PERFORMANCE OF SHANNON'S MAXIMUM ENTROPY DISTRIBUTION UNDER SOME RESTRICTIONS: AN APPLICATION ON TURKEY'S ANNUAL TEMPERATURES*

Hatice ÇİÇEK[1]

Sinan SARAÇLI[2•]

[1] *Anadolu University Faculty of Sciences, Department of Statistics, Eskişehir-Turkey*
[2] *Afyon Kocatepe University Faculty of Sciences, Department of Statistics, Eskişehir-Turkey*
Received: 10 June 2014
Accepted: 07 January 2015

Abstract

Entropy has a very important role in Statistics. In recent studies it can be seen that entropy started to take place nearly in every brunch of science. In information theory, entropy is a measure of the uncertainty in a random variable. While there are different kinds of methods in entropy, the most common maximum entropy (MaxEnt) method maximizes the Shannon's entropy according to the restrictions which are obtained from the random variables. MaxEnt distribution is the distribution which is obtained by this method. The purpose of this study is to calculate the MaxEnt distribution of Turkey's Annual temperatures for last 43 years under combinations of the restrictions 1, x, x^2, lnx, $(\ln x)^2$, $\ln(1+x^2)$ and to compare this distribution with the real probability distribution by the help of Kolmogorov-Smirnov goodness of fit test. According to the results, goodness of fit statistics accept the null hypothesis that all the entropy distributions fit with the probability distribution. The results are given in related tables and figures.

Keywords: Shannon's Maximum Entropy Distribution, Lagrange Multipliers, Discrete Distributions.

Jel Code : C02, C46, C63.

---

* This study is a part of Hatice Çiçek's MS thesis supervised by Sinan Saraçlı at Afyon Kocatepe University Institute of Science, 2013 and a part of this study is presented at ENAR's Spring Meeting, March 16 – 19, 2014, Baltimore, Maryland, USA.

•Corresponding author: ssaracli@aku.edu.tr

SHANNON'UN MAKSİMUM ENTROPİ DAĞILIMININ BAZI KISITLAR ALTINDAKİ PERFORMANSI: TÜRKİYE'NİN YILLIK HAVA SICAKLIKLARI ÜZERİNE BİR UYGULAMA

Özet

İstatistik biliminde entropi oldukça önemli bir yere sahiptir. Son yıllardaki çalışmalarda entropinin neredeyse bilimin her dalında yer aldığı görülebilir. İnformasyon teorisinde, Entropi, rassal bir değişkenin belirsizliğinin bir ölçüsüdür. Entropi içerisinde farklı birçok metot olmasına rağmen, en yaygın olan Maximum Entropy (MaxEnt) metodu, rassal değişkenlerden elde edilen kısıtlara bağlı olarak Shannon'un entropisini maksimize eder. MaxEnt dağılımı ise bu metot aracılığı ile elde edilen dağılımdır.

Bu çalışmanın amacı, Türkiye'nin son 43 yıllık sıcaklık değerleri için 1, x, x^2, lnx, $(\ln x)^2$, $\ln(1+x^2)$ kısıtlarının kombinasyonları ile MaxEnt dağılımını hesaplamak ve bu dağılımı gerçek olasılık dağılımı ile Kolmogorov-Smirnov uyum iyiliği testi yardımı ile karşılaştırmaktır. Elde edilen sonuçlara göre tüm entropi dağılımlarının gerçek olasılık dağılımı ile uyum gösterdiği şeklindeki sıfır hipotezi kabul edilmektedir. Elde edilen sonuçlar ilgili tablo ve grafiklerde verilmektedir.

Anahtar Kelimeler: Shannon'un Maksimum Entropi Dağılımı, Lagrange Çarpanları, Kesikli Dağılımlar.

Jel Kodu: C02, C46, C63.

1. Introduction

Historically, many notations of entropy have been proposed. The etymology of the word entropy dates back to Clausius (Clausius 1865), in 1865, who dubbed this term from the greek tropos, meaning transformation, and a prefix en- to recall the indissociable (in his work) relation to the notion of energy (Jaynes 1980). A statistical concept of entropy was introduced by Shannon in the theory of communication and transmission of information (Lesne, 2011).

A Maximum Entropy (MaxEnt) density can be obtained by maximizing Shannon's information entropy measure subject to known moment constraints. According to Jaynes (1957), the maximum entropy distribution is "uniquely determined as the one which is maximally noncommittal with regard to missing information, and that it agrees with what is known, but expresses maximum uncertainty with respect to all other matters." The MaxEnt approach is a flexible and powerful tool for density approximation, which nests a whole family of generalized exponential distributions, including the exponential, Pareto, normal, lognormal, gamma, beta distribution as special cases (Wu, 2003).

There are many subjects in statistics, examined via Maximum entropy or minimum cross entropy application (MinxEnt) (Kullback, 1959, Kapur and Kesevan 1992; Shamilov ve Kantar Mert 2005, Usta, 2006).

There are potentially more appropriate measures of information than the variance, however, such as that developed by Shannon (1948), Shannon and Weaver (1949), Renyi (1961) and Khinchine (1957). This information theoretic approach was rigorously related to the general body of statistics by Kullback and Leibler (1951) and Kullback (1959). These authors and other current analysts such as Parzen (1990a, b) and

Brockett (1992) have continued to conduct research to show how the information theoretic approach can lead to a view of statistics which both unifies and extends the various parts of the body of statistical methods and theories (Brocket et al 1995).

2. Material and Method

As Losee (1990) mentioned; the amount of self-information that is contained in or associated with a message being transmitted, when the probability of its transmission is p, the logarithm of the inverse of the probability is as in [1].

$$h = \log\frac{1}{p} \quad or \quad h = -\log p \tag{1}$$

For a random variable X with values in a finite set R, Shannon's entropy H(x) can be defined as in [2].

$$\mathrm{H}(\mathrm{x}) = -\sum_{\mathrm{x}\in\mathrm{R}} \mathrm{p}(\mathrm{x})\log\mathrm{p}(\mathrm{x}) \tag{2}$$

The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly bits, a word suggested by J. W. Tukey (Shannon and Weaver, 1949).

Recent studies show that, when deciding the restrictions, Entropy distributions of the characterizing moments and some combinations of these moments of a known statistical distribution gives better results to model the data set. For example Wu and Stengos (2005) used x , x^2 , ln(1+ x^2) and sin x functions as the restrictions, Wu and Perloff (2007) used x , x^2 , ln(1+ x^2) and arctan x and Shamilov et al (2008) used x , x^2 , x^3 , ln x , $(\ln x)^2$ and ln(1+ x^2) as the restrictions for the entropy distribution (Usta, 2009).

In our study like these recent studies we used 1, x, x^2, lnx, $(\ln x)^2$, and $\ln(1+x^2)$ as the restrictions to calculate the entropy distributions.

When there are more than one restriction, we need to use Lagrange multipliers to solve the restricted equations at the same time. If we consider an entropy distribution with three restrictions, to find the MaxEnt distribution of a random x variable, with probabilities $p_1, p_2, \ldots, p_n$ the *H(x)* must be solved under the restrictions given below.

$$\sum_{i=1}^{n} p_i = 1 \tag{3}$$

$$\sum_{i=1}^{n} p_i g_{1i} = \mu_1 \tag{4}$$

$$\sum_{i=1}^{n} p_i g_{2i} = \mu_2 \tag{5}$$

For three restrictions like this, the Lagrange function can be obtained as in [6]. Here μ_i are the i^{th} moments of the related data.

$$L \equiv -\sum_{i=1}^{n} p_i ln p_i - \lambda_0\left(\sum_{i=1}^{n} p_i - 1\right) - \lambda_1\left(\sum_{i=1}^{n} p_i g_i - \mu_1\right) - \lambda_2\left(\sum_{i=1}^{n} p_i g_i - \mu_2\right) \tag{6}$$

If we set equation [6] to zero after derivation according to p_is, then

$$ln p_i = -1 - \lambda_0 - \lambda_1 g_{1i} - \lambda_2 g_{2i} \quad i = 1, 2, \ldots, \tag{7}$$

$$p_i = \exp\left(-1 - \lambda_0 - \lambda_1 g_{1i} - \lambda_2 g_{2i}\right) \quad i = 1, 2, \ldots, \mathrm{n} \tag{8}$$

$$\sum_{i=1}^{n} \exp\left(-1 - \lambda_0 - \lambda_1 g_{1i} - \lambda_2 g_{2i}\right) = 1 \quad i = 1, 2, \ldots, n \tag{9}$$

$$\exp\left(-1 - \lambda_0\right) = \frac{1}{\sum_{i=1}^{n} \exp\left(-\lambda_1 g_{1i} - \lambda_2 g_{2i}\right)} \quad i = 1, 2, \ldots, n \tag{10}$$

As a result we can obtain the MaxEnt

probabilities as in [11] (Değirmenci, 2011).

$$p_i = \frac{\exp(-\lambda_1 g_{1i} - \lambda_2 g_{2i})}{\sum_{i=1}^{n} \exp(-\lambda_1 g_{1i} - \lambda_2 g_{2i})} \quad (11)$$

As an illustrative example lets think that we have observations as 3, 7, 10 and 12. and lets take the restrictions as (1, x and x^2) now we may write the equations like in [13] (Çiçek, 2013).

$$p_1 + p_2 + p_3 + p_4 = 1$$

$$3p_1 + 7p_2 + 10p_3 + 12p_4 = 8$$

$$9p_1 + 49p_2 + 100p_3 + 144p_4 = 75.5 \quad [13]$$

When we adapt the given equations we can obtain the equations given [14].

$$\left.\begin{aligned} p_1 &= e^{-\lambda_0-\lambda_1 x} = e^{-\lambda_0-3\lambda_1-9\lambda_2} \\ p_2 &= e^{-\lambda_0-\lambda_1 x} = e^{-\lambda_0-7\lambda_1-49\lambda_2} \\ p_3 &= e^{-\lambda_0-\lambda_1 x} = e^{-\lambda_0-10\lambda_1-100\lambda_2} \\ p_4 &= e^{-\lambda_0-\lambda_1 x} = e^{-\lambda_0-12\lambda_1-144\lambda_2} \end{aligned}\right\} \quad [14]$$

When we solve these equations we can obtain the Lagrange multipliers as;

$\lambda_0 = 0.5618$, λ_1 =-7,80E-18 and $\lambda_2 = 0.0141$

As a result, by the help of these multipliers we may obtain the MaxEnt distribution as in Table 1.

Table 1. MaxEnt distribution of the sample for three restrictions.

p_1	0.5020
p_2	0.2851
p_3	0.1386
p_4	0.0744

3. Application

In this section of the study, MaxEnt distributions for temperature values in Turkey during the last 43 years are calculated. The data set is obtained from Turkish State Meteorological Service. To calculate MaxEnt distribution of the related data set under restrictions with the help of the Lagrange multipliers we used MATLAB software and developed a program to calculate any discrete data set under some restrictions. The frequency distribution for this data set can be seen in Table 3 and its histogram is given in Figure 1.

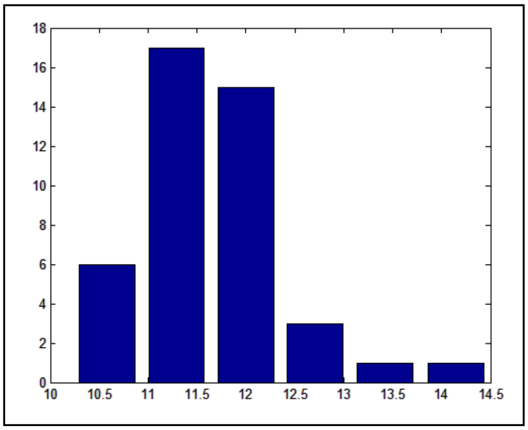

Figure 1. Histogram for the annual temperature values (in Celsius) of Turkey for last 43 years.

Figure 1 shows that the average annual temperature of Turkey in last 43 years is about 11-12 C.

Entropy values are calculated under two, three, four, five and six restrictions for this data set. The best entropy values (Minimum uncertainty amount) for the related restrictions are shown in bold and given in Table 2.

Table 2. Entropy values of the temperature distribution under given restrictions.

Restrictions	***Entropy H(x)***
(1, x)	1.77536033981
(1, x^2)	**1.72687192330**
(1, lnx)	1.78918471283
(1, $(\ln x)^2$)	1.78131812647
(1, $\ln(1+x^2)$)	1.79175946922
(1, x, x2)	1.77371720718
(1, x, lnx)	1.77554313034
(1, x, $(\ln x)^2$)	1.77549169046
(1, x, $\ln(1+x^2)$)	1.77554248388
(1, x^2, lnx)	1.72736338152
(1, x^2, $\ln(x)^2$)	1.72771821411
(1, x^2, $\ln(1+x^2)$)	**1.72735921453**
(1, lnx, $(\ln x)^2$)	1.78906743971
(1, lnx, $\ln(1+x^2)$)	1.78918513332
(1, $(\ln x)^2$, $\ln(1+x^2)$)	1.77923497943
(1, x, x^2, lnx)	1.68678640317
(1, x, x^2, $(\ln x)^2$)	1.71520751538
(1,x, x^2, $\ln(1+x^2)$)	**1.68653487964**
(1, x, lnx, $\ln(x)^2$)	1.77554313034
(1, x, lnx, $\ln(1+x^2)$)	1.77543835428
(1, x, $(\ln x)^2$, $\ln(1+x^2)$)	1.74012605933
(1, x^2, lnx, $(\ln x)^2$)	1.74973313274
(1, x^2, lnx, $\ln(1+x^2)$)	1.72716039931
(1, x^2, $(\ln x)^2$, $\ln(1+x^2)$)	1.70088754359
(1, lnx, $(\ln x)^2$, $\ln(1+x^2)$)	1.78133520000
(1, x, x^2, lnx, $(\ln x)^2$)	1.73922899405
(1, x, x^2, lnx,$\ln(1+x^2)$)	1.68628798649
(1, x, x^2, $(\ln x)^2$, $\ln(1+x^2)$)	**1.65478629163**
(1, x, lnx, $\ln(x)^2$, $\ln(1+x^2)$)	1.75994429256
(1, x^2,lnx, $(\ln x)^2$, $\ln(1+x^2)$)	1.70065491805
(1, x, x^2, lnx, $(\ln x)^2$, $\ln(1+x^2)$)	***1.65073612464***

Table 2 shows that the minimum Entropy (Maximum information) is obtained as 1.6507 under six restrictions. As a summary of the table; the minimum entropy value under restrictions (1, x^2) is 1.7268, restrictions (1, x^2, $\ln(1+x^2)$) is 1.7273, restrictions (1,x, x^2, $\ln(1+x^2)$) is 1.6865, restrictions (1, x, x^2, $(\ln x)^2$, $\ln(1+x^2)$) is 1.6547, and restrictions (1, x, x^2, lnx, $(\ln x)^2$, $\ln(1+x^2)$) is 1.6507.

It can also be seen that by increasing the number of restrictions, the entropy values decrease.

At the next step of the analysis, Kolmogorov-Smirnov goodness of fit test is applied to test whether or not each of the entropy distributions under these restrictions fit to the real probability distribution.

The two-sample Kolmogorov-Smirnov (K-S) goodness of fit test is one of the most useful and general nonparametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples.

If Fo(x) is the population cumulative distribution, and $S_N(x)$ the observed cumulative step-function of a sample (i.e., $S_N(x) = k/N$, where k is the number of observations less than or equal to x), then the sampling distribution of D= maximum $|F_o(x) - S_N(X)|$ is known, and is independent of $F_o(x)$ if $F_o(x)$ is continuous (Frank and Massey, 1951).

Null and the alternative hypothesis for K-S test can be written as:

$H_0 : F(x) = F_0(x)$ (The data follow a specified distribution)

$H_1 : F(x) \neq F_0(x)$ (The data do not follow the specified distribution)

Table 3. Temperatures, frequencies, probabilities and entropy distributions

Temperature	f_o	p_i	f_2	f_3	f_4	f_5	f_6
10.23-10.94	6	0.1395	0.2629	0.2625	0.2941	0.3164	0.3192
10.94-11.65	17	0.3953	0.2169	0.2168	0.2288	0.2355	0.2362
11.65-12.36	15	0.3488	0.1767	0.1768	0.1752	0.1726	0.1722
12.36-13.07	3	0.0697	0.1422	0.1424	0.1322	0.1248	0.1238
13.07-13.78	1	0.0232	0.1130	0.1132	0.0983	0.0889	0.0878
13.78-14.53	1	0.0232	0.0880	0.0883	0.0714	0.0618	0.0608

f_o: Observed frequencies
p_i: Probability distribution
f_{2-6}: Entropy distributions under given restrictions

Maximum Differences between the probability distribution and entropy distributions (D) according to Cumulative Density Functions (CDF) and probabilities for these differences are given in Table 4 and the graph for Cumulative Density Function for all entropy distributions is given in Figure 2.

Table 4. Goodness of fit statistics for entropy distributions and data set.

pi-Hi(X)	***D***	***p(D)***
$H_2(X)$	0,1784	0,3180
$H_3(X)$	0,1785	0,3180
$H_4(X)$	0,1736	0,3180
$H_5(X)$	0,1762	0,8096
$H_6(X)$	0,1766	0,8096

Table 4 shows that according to the probabilities (*p(D)*) of K-S test we accept the null hypothesis and we can say that the maximum entropy distributions under all restrictions statistically fit to the related data set 95% confidently.

While we obtain the maximum information from the entropy distribution under six restrictions, according to Figure 2. and the D values given in Table 4, the maximum difference is between the probability distribution (the red line in Figure 2) and the entropy distribution under three restrictions (the green line in Figure 2) according to Cumulative Density Function.

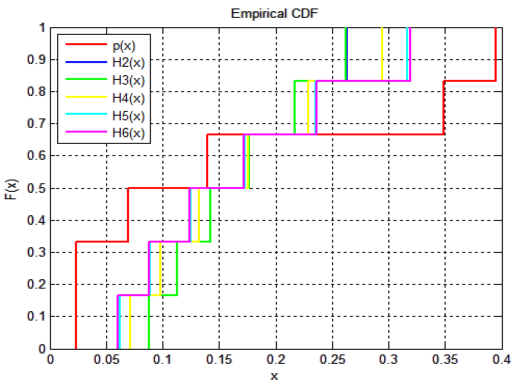

Figure 2. CDF Graph of KS test

ISSN 2148-2225
http://www.alphanumericjournal.com/

4. Results and Discussion

In this study, the performances of Shannon's maximum entropy distributions are examined under two, three, four, five and six restrictions for discrete variables and comparisons of restricted entropy distributions are concluded according to their entropy values which obtained minimum for the related restriction.

One of the importance of this study can be defined as; if any data set doesn't fit to a known statistical distribution, it can be explained via a entropy distribution.

Results show that by an increasing number of restrictions, MaxEnt distribution explains the related data set much better.

To explain the Turkey's annual temperature values for the last 43 years, the best MaxEnt distribution has the restrictions set of (1, x, x^2, lnx, $(\ln x)^2$, $\ln(1+x^2)$) with an entropy value of 1.6507.

References

Brockett P.L. (1992). Information theoretic approach to actuarial science: A unification and extension of relevant theory and applications, Society of Actuaries Transactions XLIII.

Brockett P.L., Charnes A., Cooper W.W., Learner D., Phillips F.Y., (1995). Information theory as a unifying statistical approach for use in marketing research, European Journal of Operational Research (84) 310-329

Clausius R., (1865). The mechanical theory of heat-with its applications to the steam engine and to physical properties of bodies, John van Voorst, London.

Çiçek H., (2013). Maksimum Entropi Yöntemi ile Türkiye'deki Coğrafi Bölgelerin Yıllık Hava Sıcaklık Değerlerinin İncelenmesi, Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.

Değirmenci İ., (2011). Entropi Ölçümleri ve Maksimum Entropi İlkesi, Yüksek Lisans Tezi, Hacettepe Üniversitesi.

Frank J., Massey, J.R., (1951). The Kolmogorov-Smirnov Test for Goodness of Fit, Journal of the American Statistical Association, Vol. 46, No. 253 pp. 68-78.

Jaynes E.T., (1957). Information theory and statistical mechanics. Physics Review 106, 620–630.

Jaynes E. T. (1980). The minimum entropy production principle. Ann. Rev. Phys. Chem. 31, 579-601.

Kapur J.N., Kesevan, H.K., (1992). Entropy Optimization Principle with Applications, Academic Press.

Khinchine A.I., (1957). Mathematical Foundations of Information Theory, Dover Publ., New York. (New translation of Khinchine's papers "The entropy concept in probability theory" and "On the fundamental theorems of information theory" originally published in Russian in Uspekhi Matematicheskikh VII (3) (1953) and XI (1) (1956), respectively).

Kullback S., (1959). Information Theory and Statistics, Wiley, New York.

Kullback S., Leibler R.A., (1951). On information and sufficiency, Annals of Mathematical Statistics 22, 79-86.

Lesne A., (2011). Shannon entropy: a rigorous mathematical notion at the crossroads between probability, information theory, dynamical systems and statistical physics. Under consideration for publication in Math. Struct. in Comp. Science (Source: http://preprints.ihes.fr/2011/M/M-11-04.pdf)

Losee R.M., (1990). The Science of Information: Measurement and Applications. Academic Press, San Diego.

Parzen E., (1990a). Goodness of fit tests and entropy, Department of Statistics, Texas A & M University, Tech. Report No:103.

Parzen E., (1990b). Unification of statistical methods for continuous and discrete data, Department of Statistics, Texas A & M University, Tech. Report No. 105.

Renyi A., (1961). On measures of entropy and information, Proc. 4th Berkeley Symp. Math. Statist. Probability, 1960, University of California Press, Berkeley, CA, Vol. 1, 547-561.

Shannon C.E., (1948). A mathematical theory of communication, Bell System Technical Journal 27, 379-423, 623-656.

Shannon C.E., Weaver W., (1949). The Mathematical Theory of Communication, University of Illinois Press, Urbana, Ill.

Shamilov A., and Kantar Mert Y. (2005), "On a distribution minimizing maximum entropy", Ordered Statistical Data: Approximations, Bounds and Characterizations, Izmir University of Economics.

Shamilov A., Kantar Mert, Y., Usta, I., (2008). Use of MinMaxEnt distributions defined on basis of MaxEnt method in wind power study, Energy Conversion & Management, 49(4), 660-677.

Usta İ., (2006). MaxEnt ve MinxEnt Optimizasyon Prensiplerine Bağlı Nümerik İncelemeler ve İstatistiksel Uygulamalar, Yüksek Lisans Tezi, Anadolu Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı.

Usta, İ., (2009). Moment Kısıtlarına Dayalı Genelleştirilmiş Entropi Yöntemleri, Doktora Tezi, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı.

Wu X., (2003). Calculation of maximum entropy densities with application to income distribution, Journal of Econometrics (115) 347 – 354.

Wu X., Stengos T., (2005). Partially Adaptive Estimation via Maximum Entropy Densities, Econometrics Journal, 8(3), 352-366.

Wu X., Perloff, J.M., (2007). GMM estimation of a maximum entropy distribution with interval data, Journal of Econometrics, 138(2), 532-546.