ISSN(E):2522-2260
ISSN(P):2522-2252

Indexing/Abstracting

Journal of Quantitative Methods

The Journey from Entropy to Generalized Maximum Entropy

Author(s)
Amjad D. Al-Nasser[1]

Affiliations
[1]Professor, Department of Statistics, Science Faculty, Yarmouk University, Irbid 21163, Jordan.

Manuscript Information

Submission Date: December 24, 2018
Acceptance Date: February 27, 2019

Citation in APA Style

Al-Nasser, A., D. (2019). The journey from entropy to generalized maximum entropy, *Journal of Quantitative Methods*, *3*(1), 1-7.

This manuscript contains references to 23 other manuscripts.

The online version of this manuscript can be found at
https://journals.umt.edu.pk/sbe/jqm/volume3issue1.aspx#

DOI: https://doi.org/10.29145/2019/jqm/030101

Published by

Department of Quantitative Methods
School of Business and Economics
University of Management and Technology, Lahore, Pakistan

Additional Information

Subscriptions and email alerts: editorasst.jqm@umt.edu.pk
For further information, please visit
http://journals.umt.edu.pk/sbe/jqm/Home.aspx

Editorial

The Journey from Entropy to Generalized Maximum Entropy

Amjad D. Al-Nasser[1]

https://doi.org/10.29145/2019/jqm/030101

1. Introduction

Currently we are witnessing the revaluation of huge data recourses that should be analyzed carefully to draw the right decisions about the world problems. Such big data are statistically risky since we know that the data are combination of (useful) signals and (useless) noise, which considered as unorganized facts that need to be filtered and processed. Using the signals only and discarding the noise means that the data restructured and reorganized to be useful and it is called information. So for any data set, we need only the information. In context of information theory, the entropy is used as a statistical measure to quantify the maximum amount of information in a random event. Therefore, entropy is a very important measure to study the lack of knowledge and quantify the randomness in a data. Thermodynamically, entropy measures the amount of energy lost when doing useful work. Moreover, in our real life it can be used to measure the degree of organized lives, such that high entropy value means highly unorganized person.

The entropy measure passed through three stages; started in 1870 when Boltzmann proposed the **entropy** as a measure of information to define the thermodynamic state of a physical system. In 1948, the mathematical theory of the entropy measure is provided by Shannon. He used the entropy measure concept to measure the uncertainty of a message that contains a noise and then to measure the amount of information. In 1957, Jaynes proposed the **maximum entropy (ME)** principle as a new estimation tool in statistical inference. In 1968, Jaynes used the ME principle as a tool in probability theory for solving ill-posed problems by finding the optimal probability distribution of a random variable subject to data constraints.

[1]Department of Statistics, Science Faculty, Yarmouk University, Jordan.
Email: amjadyu@yahoo.com

Over the years, the use of the maximum entropy formalism accelerated faster, and entered many branches of sciences (Mead & Papanicolaou, 1984; Zellner & Highfield, 1988; Csiszar, 1991; Press, 1996). Since the 1990's many efforts have been made to integrate the ME method. In 1996, Golan, Judge Miller proposed the idea of the **generalized maximum entropy (GME)** as a new estimation method for fitting the general linear models. Unlike the ME, the GME approach idea is to solve any mathematical system even if it is not in probability form (Golan and Ullah, 2017; Al-Nasser, 2005; Ciavolino, Carpita &Al-Nasser, 2015). In the general linear models, this can be done by reparametrizing the unknown model coefficients as well as the additive error terms in expected value of a discrete random variable form, then solving a mathematical programming problem using the joint entropies as an objective function subject to the data model (Golan, Judge & Perloff, 1997; Golan & Gzyl, 2012; Ciavolino & Al-Nasser, 2009; Ciavolino & Al-Nasser, 2010; Al-Nasser, 2011; Al-Nasser, 2012; Al-Rawwash & Al-Nasser, 2013 and Al-Nasser, 2014).

2. Entropy Stage

The mathematical entropy measure formula as defined by Shannon (1948) can be defined as a negative average of the logarithm of a probability density (mass) function which is called the amount of self-information for a given event. For a discrete event; say $y = \{y_1, y_2, \dots, y_k\}$ with probabilities P = $\{p_1, p_2, \dots, p_k\}$, then the entropy measure is

$$H(P) = -\sum_{i=1}^{k} p_i \ln(p_i)$$

where 0ln(0) = 0. In terms of statistical experiments, in 1962; E. T Jayens credits Graham Wallis with the experiment (Jaynes, 2003) derived a measure of entropy using multinomial experiment. He defined the probability of success as $p_i = \frac{n_i}{N}, i = 1,2,\dots,K$, where K is the number of all possible outcomes within each trail, n_i is number of times that the i[th] event occurs among the N trials, and $\sum_i n_i = N$ is the total number of all trails. Accordingly, the multinomial coefficient of this experiment can be defined as

$$W = \frac{N!}{(N\,p_1)!(N\,p_2)!....(N\,p_K)!} = \frac{N!}{\prod_{i=1}^{K} n_i!}$$

When N becomes large, one can use Striling's approximation to simplify this function, and show that

$$\frac{\ln(W)}{N} = -\sum_i p_i \ln(p_i) = H(p)$$

which gives the same Shannon's entropy measure which concerns in quantifying the amount of uncertainty using consistent measure.

3. Maximum Entropy Stage

Jaynes (1968) proposed the ***Maximum Entropy (ME)*** principle as a mathematical programming problem in which the objective function is the entropy measure and the constraints are derived from the data in terms of moments of a probability density (mass) function. The main idea behind the ME is to find the optimal probability distribution containing the largest amount of uncertainty subject to its constraints. ME optimisation problem is given in figure 1, which can be considered as a non-linear programming system that used to determine the optimal probability distribution.

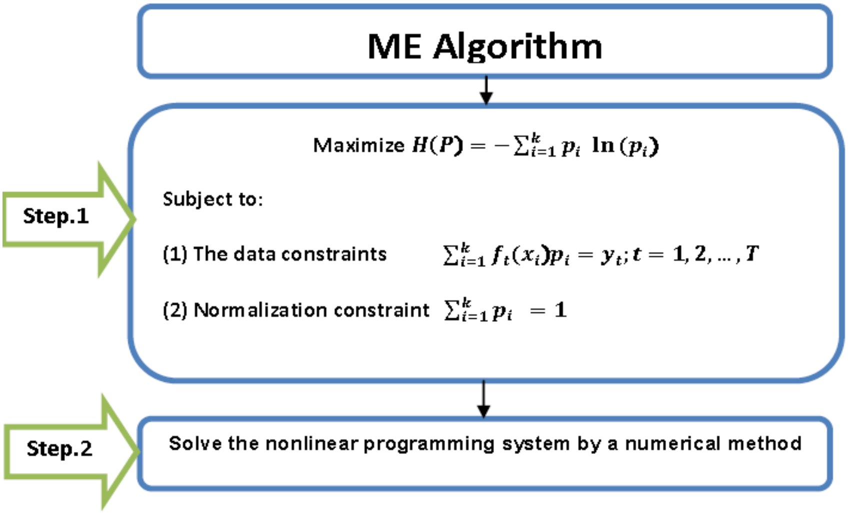

Figure 1: Maximum Entropy Optimization Problem

where $\sum_{i=1}^{k} f_t(x_i)\, p_i = y_t$ is the t$^{\text{th}}$ moment constraints.

The entropy measure for a continuous random variable is defined with exchanging the summation symbol by the integral symbol as

$$H = -\int f(x)\ln(f(x))\partial x$$

The solution for the maximum entropy system will be the same as in the discrete case. Many researchers focused their work in the continuous case to obtain the ME distribution. Zellner and Highfield (1988) derived the ME distribution with respect to high order moments constraints. Press (1996) presented a summary of some ME distributions.

3. Generalized Maximum Entropy Stage

Golan, Judge and Miller (1996) proposed the GME estimation procedure as an extension of the ME. The main idea behind the GME estimation procedure is to solve the underlying problem using the data information even if these information are limited, partial or incomplete known. Therefore, the optimization problem steps that are given in figure 1 should be extended to recover the unknown parameters when they are not in probability terms. This can be done by reparametrizing the unknown parameters and noises in terms of probabilities and then rewrite the model by using the new formulation. After adding these two steps, the ME algorithm can be applied. For example, consider the general linear model

$$y = X\beta + e$$

where **y** is response variable of size N, β is unknown parameters, **X** is a predictor variables of size K*N and **e** is unobservable random errors with finite location and scale parameters. Given information about the predictors (X) and response (**y**) variables, the objective of GME procedure is to recover the unknown model coefficients (parameters) β and the disturbance term **e**. This can be done in four steps as shown in figure 2.

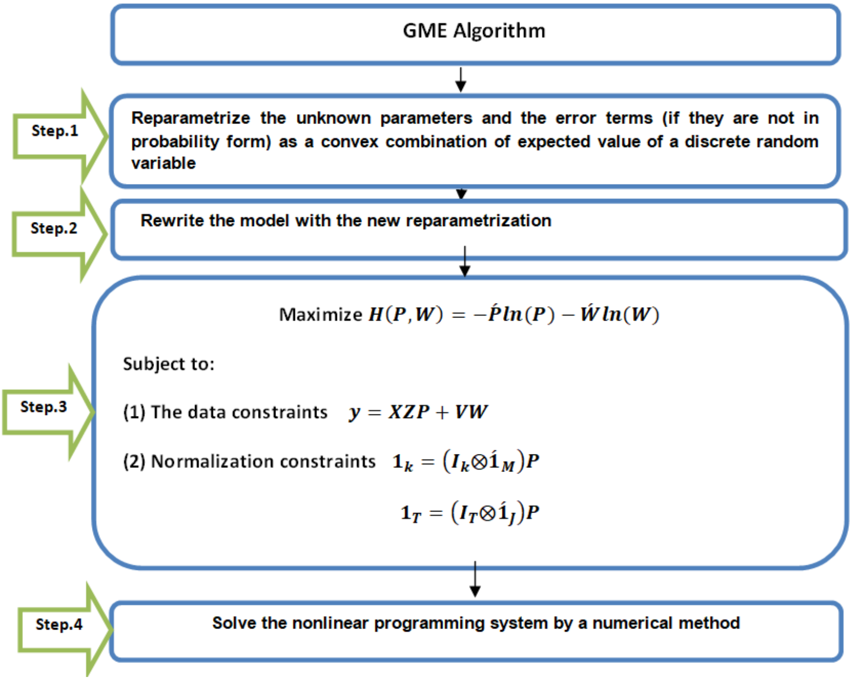

Figure 2: Generalized Maximum Entropy Estimation Procedure

where $\beta = ZP$; $e = VW$, $\otimes$ is the Kronecker product, 1_T is a T-dimensional vector. Then the optimal solution of this system will be of the following form:

$$\hat{\beta} = Z\hat{P} \text{ and } \hat{e} = V\hat{W}$$

More details can be obtained from Golan, Judge and Miller(1996), Golan (2014) and Al-Nasser (2010).

References

Al-Nasser, A. D. (2010). *Measuring customer satisfaction: An information-theoretic approach*. Germany: LAP Lambert Academic Publishing AG & Co. KG.

Al-Nasser, A. D. (2005). Entropy type estimator to simple linear measurement error models. *Austrian Journal of Statistics*, *34*(3), 283-294. https://doi.org/10.17713/ajs.v34i3.418.

Al-Nasser, A. D. (2011). An information-theoretic alternative to maximum likelihood estimation method in ultrastructural measurement error model. *Hacettepe Journal of Mathematics and Statistics*, *40*(3), 469-481.

Al-Nasser, A. D. (2012). Human Development Index in the Arab States. *Journal of Applied Sciences*, *12*(4), 326-335. https://doi.org/10.3923/jas.2012.326.335.

Al-Nasser, A. D. (2014). Two steps generalized maximum entropy estimation procedure for fitting linear regression when both covariates are subject to error. *Journal of Applied Statistics, 41*(8), 1708-1720. https://doi.org/10.1080/02664763.2014.888544.

Al-Rawwash, M., & Al-Nasser, A. D. (2013). Repeated measures and longitudinal data analysis using higher-order entropies. *Statistica Neerlandica*, *67*(1), 100-111. https://doi.org/10.1111/j.1467-9574.2012.00534.x.

Boltzmann L. (1970) Weitere studien uber das Warmegleichgewicht unter Gasmolekülen. In: Kinetische Theorie II. *WTB wissenschaftliche taschenbucher*, (pp. 115-225). Wiesbaden, Germany: Vieweg Teubner Verlag. https://doi.org/10.1007/978-3-322-84986-1_3.

Ciavolino, E. & Al-Nasser A. D. (2010). Information theoretic estimation improvement to the nonlinear Gompertz's model based on ranked set sampling. *Journal of Applied Quantitative Methods*. *5*(2), 317-330.

Ciavolino, E., & Al-Nasser, A. D. (2009). Comparing generalized maximum entropy and partial least squares methods for structural equation models. *Journal of nonparametric statistics*, *21*(8), 1017-1036. https://doi.org/10.1080/10485250903009037.

Ciavolino, E., Carpita, M., & Al-Nasser, A. (2015). Modelling the quality of work in the Italian social co-operatives combining NPCA-RSM and SEM-GME approaches. *Journal of Applied Statistics*, *42*(1), 161-179. https://doi.org/10.1080/02664763.2014.938226.

Csiszar, I. (1991). Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, *19*(4), 2032-2066. https://doi.org/10.1214/aos/1176348385.

Golan, A. (2014). Information dynamics. *Minds and Machines*, *24*(1), 19-36. https://doi.org/10.1007/s11023-013-9326-2.

Golan, A., & Gzyl, H. (2012). An entropic estimator for linear inverse problems. *Entropy*, *14*(5), 892-923. https://doi.org/10.3390/e14050892.

Golan, A., & Ullah, A. (2017). Interval estimation: An information theoretic approach. *Econometric Reviews*, *36*(6-9), 781-795. https://doi.org/10.1080/07474938.2017.1307573.

Golan, A., Judge, G. G., & Miller, D. (1996). *Maximum entropy econometrics: Robust estimation with limited data*. New York: Wiley.

Golan, A., Judge, G., & Perloff, J. (1997). Estimation and inference with censored and ordered multinomial response data. *Journal of Econometrics*, *79*(1), 23-51. https://doi.org/10.1016/S0304-4076(97)00006-7.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*(4), 620. https://doi.org/10.1103/PhysRev.106.620.

Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, *4*(3), 227-241. https://doi.org/10.1109/TSSC.1968.300117.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, England.

Mead, L. R., & Papanicolaou, N. (1984). Maximum entropy in the problem of moments. *Journal of Mathematical Physics*, *25*(8), 2404-2417. https://doi.org/10.1063/1.526446.

Press, S. J. (1996). The de Finetti transform. In Hanson, k. M. & Silver, N. R. (Eds), *Maximum Entropy and Bayesian Methods* (pp. 101-108). Netherlands: Kluwer Academic https://doi.org/10.1007/978-94-011-5430-7_12.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Zellner, A., & Highfield, R. A. (1988). Calculation of maximum entropy distributions and approximation of marginal posterior distributions. *Journal of Econometrics*, *37*(2), 195-209. https://doi.org/10.1016/0304-4076(88)90002-4.

Citation: Al-Nasser, A., D. (2019). The journey from entropy to generalized maximum entropy, *Journal of Quantitative Methods*, *3*(1), 1-7.

Submission Date: December 24, 2018
Last Revised: February 08, 2019
Acceptance Date: February 27, 2019