OPEN ACCESS

1st International Electronic Conference on Entropy and Its Applications
3 - 21 November 2014

www.sciforum.net/conference/ecea-1

Conference Proceedings Paper – Entropy

Maximum Entropy Approach for Reconstructing Bivariate Probability Distributions

Gholamhossein Yari [1], Zahra Amini Farsani [2], *

[1] School of Mathematics, Iran University of Science and Technology, Tehran, IRAN; E-Mails: Yari@iust.ac.ir
[2] School of Mathematics, Iran University of Science and Technology, Tehran, IRAN; E-Mails: z_aminifarsani@iust.ac.ir

* School of Mathematics, Iran University of Science and Technology, Tehran, IRAN; E-Mails: z_aminifarsani@iust.ac.ir;
Tel.: +98-2173225403; Fax: +9821-77240302.

Received: 14 October 2014 / Accepted: 26 October 2014 / Published: 3 November 2014

Abstract: The main purpose for this study is to provide a useful algorithm that combines the Maximum Entropy Method (MEM) and a computational method to predict the unique form of a two-dimensional maximum entropy distributions. In this paper, we present the application of MEM to determine the important bivariate distributions which are very effective in industrial and engineering fields especially in Cybernetics and internet systems. The combination of MEM and numerical method as a proposed algorithm is a useful method with the ability to model datasets which includes both missing and presence datas, The new algorithm provides reasonable estimations for reconstructing the target bivariate distributions which has maximum entropy. We examined the effectiveness of our algorithm to make predictions for two famus distributions. The resulting distributions have minimum error with respect to missing information. In fact, maximum entropy distribution is able to fit the known density given the prior knowledge of the target distribution. So, the MEM modeling procedure can be applied in present form for a variety of applications with precsence-only datasets. Possessing the simple and accurate mathematical formulation and using presence-only data, MEM has become a well-suited method for different kinds of distribution modeling.

Keywords: Maximum Entropy Method, Bivariate distribution, Shannon Entropy, Computational algorithm, Modeling, Cybernetics.

---

1. Introduction

The method of maximum entropy is a very effective procedure in the determination of the general form of a density via solving optimization problems which is introduced by Rubinstein (1999). The concept of maximum entropy method was first proposed by Jaynes (1957). It is a useful method of reconstructing a density given finite numbers of moment constraints from incomplete datasets. There are many different standard statistical methods for modeling missing/presence available data such as generalized linear and additive models (GLM and GAM). The essence of MEM is to detect the probability distribution of maximum entropy, subject to a set of constraints that represents our presence-only information about the target distribution. In other words, the MEM makes it possible to find a unique distribution that gives rise to these moments.

Shannon et al. (1948), Introduced different methods to calculate the maximum entropy distributions. In this paper, we apply the MEM to estimate two-dimensional distributions and different classes of distributions. Note that not all classes of distributions contain a maximum entropy distribution.

The case of maximum entropy bivariate distribution has been studied by Djafari (2011). He has noted that an important problem in statistics is the construction of a joint probability distribution from its marginals. He linked this problem to an important problem in Computed tomography (CT) as a reconstruction of an image from its projections. To determine $p(x, y)$ from $p_1(x)$ and $p_2(y)$, he proposed to use copula by maximum entropy method as the solution to the problem of reconstructing a bivariate density from its marginals.

In this study, we consider the determination of a two-dimensional distribution in ME problem as an optimization problem by introducing a combined algorithm. The new algorithm is able to resolve a non-linear system and calculate Lagrange multipliers to determine the class of distribution. The solution of this non-linear system requires calculation of two-dimensional integrals. Since the integrals are complex, the classical and existing numerical methods such as Simpson, etc., are not efficient and the results are not sufficiently accurate, the first step of the algorithm is a implementation of Newton method to transform this system to a system of linear equations. Results show that the proposed algorithm both advantages, of high accuracy and good agreement with exact solutions.

In the present study, various distributions such as Normal, beta etc., are considered. The rest of this paper is organized as follows: In Section 2, we present a summary of the ME method. In Section 3, we will show Some preliminary numerical experiments in various classes of distributions. Sections 4 is devoted to conclusions.

2. Results and Discussion

2.1. The Principle of Maximum Entropy

The MEM is determining the distribution that maximizes the information entropy. Jaynes' (1957) stated that maximum entropy principle provides a method for solving this problem when the available information is in the form of moment constraints. Suppose a probabilistic system in which X is a continuous random-variate with probability $p(x)$, then the information entropy S, of a distribution $p(x)$, is named Shannon entropy introduced by Shanoon et al. in 1948:

$$S = -\int_{\Omega} p(x) ln p(x) dx, \quad (1)$$

where â„¦ is the support of the distribution and the information available is in the form of the normal constraint of probability

$$\int_{\Omega} p(x) dx = 1$$

Our purpose is to find $p(x)$ that maximizes the information entropy S given in Eq.(1) subject to

$$\int_{\Omega} g_k(x) p(x) dx = a_k; \quad k = 0,...,N,$$

(2)

where g_k and a_k are known constants for $k = 0,...,N$.

ME method solves this problem by maximizing the Shannon entropy subject to the given known constraints. The problem of Shannon entropy maximization can be solved by Lagrange multiplier method (Fletcher, 1991). Where $(N+1)$ is the number of known moments. Then , we define the entropy functional by introducing Lagrangian multipliers λ_k

$$H \equiv S + \sum \lambda_k \left(\int_{\Omega} g_k(x) p(x) dx - a_k\right)$$
$$= -\int_{\Omega} p(x) ln(p(x)) dx + \sum \lambda_k \left(\int_{\Omega} g_k(x) p(x) dx - a_k\right).$$

(3)

This functional is a maximum when

$$\frac{\partial H}{\partial \lambda_k} = 0,$$

(4)

$$\frac{\partial H}{\partial p(x)} = 0.$$

(5)

Eq. (4) gives us the constraints defined in Eq. (2) and Eq. (5) leads to

$$\frac{\partial H}{\partial p(x)} = -ln p(x) - 1 + \lambda_0 + \sum_{k=1}^{N} \lambda_k g_k(x)$$
$$= 0,$$

(6)

which general solution is

$$p(x) = e^{-\sum_{k=0}^{N} \lambda_k g_k(x)}, \qquad x \in \Omega$$

(7)

where $\lambda_0,...,\lambda_N$ are chosen so that $g(x)$ satisfies the constraints. The Lagrange multipliers can be given by

$$\lambda_0 = ln(\int_\Omega exp(\sum_{k=0}^{N}\lambda_k g_k(x))dx) \tag{8}$$

The Lagrange multipliers $\lambda = [\lambda_1,...,\lambda_N]$ should be calculated to determine the class of maximum entropy distributions. Our goal is determination of $N+1$ Lagrange multiplier from a set of nonlinear equations system, N data constraints and normalization constraint. To obtain the $N+1$ Lagrange multipliers, the following set of $N+1$ nonlinear equations system should be solved:

$$\begin{aligned}
&\int_\Omega e^{\lambda_0+\lambda_1 g_1(x)+...+\lambda_k g_k(x)} dx = a_0, \\
&\int_\Omega g_1(x) e^{\lambda_0+\lambda_1 g_1(x)+...+\lambda_k g_k(x)} dx = a_1, \\
&\vdots \\
&\int_\Omega g_k(x) e^{\lambda_0+\lambda_1 g_1(x)+...+\lambda_k g_k(x)} dx = a_k,
\end{aligned} \tag{9}$$

and a globally convergent Newton solver may be used to calculate λ_k:

$$\tilde{a} \equiv \int_\Omega g_k(x) e^{\lambda_0+\lambda_1 g_1(x)+...+\lambda_k g_k(x)} dx. \tag{10}$$

3. Implementation and Results

As MEM is not available in standard statistical packages, we applied the MEM by writting code in MATLAB. In this code, we have shown the performance of Newton's method for solving two-dimensional nonlinear equations and their quadrature numerical integration. We implemented the combination of two methods for the solver, one that uses the probabilistic expression for the constraints, and one that uses numerical technique to estimate Lagrange multipliers. To test the code, a number of moments were generated by the use of known distributions. We used a finite support to carry out the integrations for solving nonlinear system and the results have a reasonable approximation to the target distributions. The following examples include the new ME characterizations of many bivariate distributions.

2.1. Bivariate Normal Distribution

For normal distribution, we consider these constraints*:*

$$\begin{cases} Normalization \quad g_0(x,y) = 1, \\ g_1(x,y) \qquad\qquad = x+y, \\ g_2(x,y) \qquad\qquad = x^2+y^2, \end{cases}$$

and $a_0 = 1, \quad a_1 = 1, \quad a_2 = \frac{3}{2}$. Consider $\Omega = (-\infty, \infty)$ as support of X and Y. Starting with a bivariate normal distribution, the PDF is given by

$$p(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_x}{\sigma_x})^2 - 2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y}) + (\frac{y-\mu_y}{\sigma_y})^2]},$$

$$-\infty < \mu < \infty \ , \rho = 0, \ \sigma_x = \sigma_y = 1.$$

Hence, the density that satisfies the constraints and also maximizes the entropy is

$$p(x,y) = e^{\lambda_0 + \lambda_1(x+y) + \lambda_2(x^2+y^2)}.$$

We apply numerical method for calculating Lagrange multipliers. The exact form of this normal distribution in terms of these conditions should be given as

$$p(x,y) = \frac{2}{\pi} e^{-(x^2+y^2)+(x+y)-\frac{1}{2}}.$$

Figure1 shows the plot of error between exact distribution and estimated bivariate normal distribution.

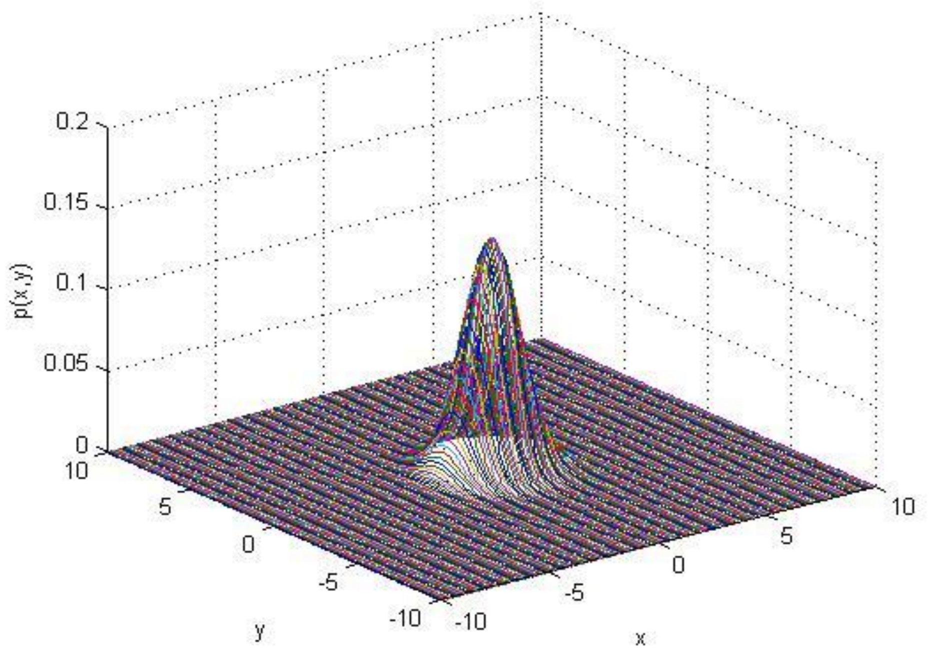

Figure 1: Errors of Maximum Entropy Reconstruction of Bivariate Normal Distribution for $\tilde{\lambda} = [0.9832, -1.6427, 0.9529]$, $\lambda_{Exact} = [.9516, -1, 1]$ *and* $\lambda_0 = [1.96, .059, .4375]$.

2.2. Bivariate Pareto Distribution

Consider the known constraints of pareto distribution $a_0 = 1,\ a_1 = \frac{3}{2}$:

$$\begin{cases} Normalization \quad g_0(x,y) = 1, \\ g_1(x,y) \qquad\qquad = log(x+y+1), \end{cases}$$

and $\Omega = [0,\infty)$, the distribution that maximized entropy is in the form of

$$p(x,y) = e^{\lambda_0 + \lambda_1 log(x+y+1)},$$

and it is pareto distribution:

$$p(x,y) = \alpha(\alpha+1)(1+x+y)^{-\alpha-2},$$

We consider $\alpha = 1$, we have

$$p(x,y) = 2(1+x+y)^{-3}.$$

See figure3 for final errors.

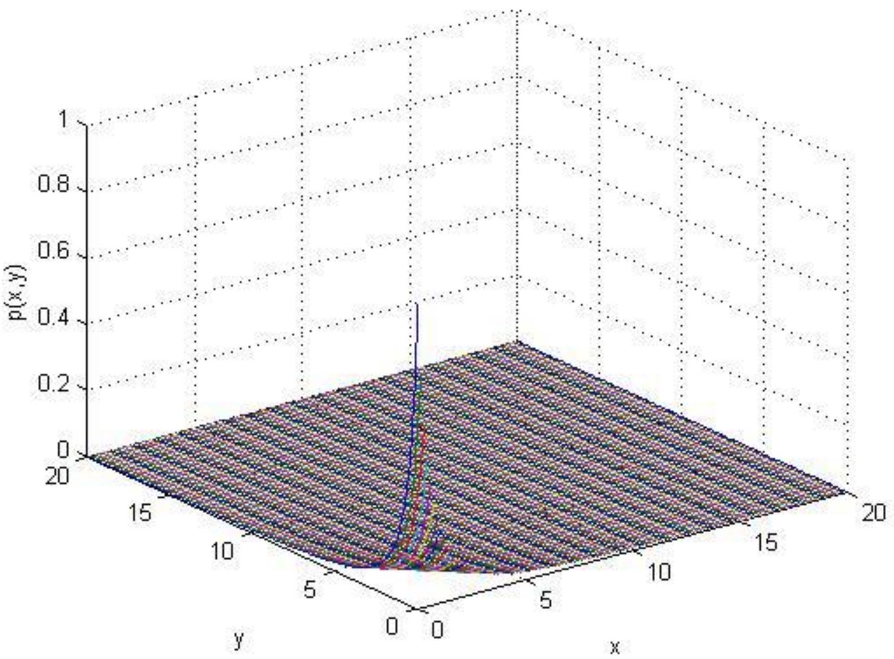

Figure 2: Errors of Maximum Entropy Reconstruction of Bivariate Pareto Distribution with $\lambda_0 = [0.5403, 1.66]$ *and* $\lambda_{exact} = [0.7696, 2.92]$.

4. Conclusions

In the present paper, an application of maximum entropy method is discussed to reconstruct bivariate distribution. A new algorithm is presented based on the standard Newton's method and probabilistic procedure. In addition, we define the critical boundary key points, which a new measure on monitoring wear condition and identifying probable wear faults are stated. The new algorithm is applied to identify distributions for cybernetics data as illustrated by four examples. By using this algorithm, the effectiveness of MEM were improved, and all kinds of bivariate distributions can be detected successfully. So the accurancy of this method is verified.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Janes, E. T.; Information Theory and Statistical Mechanics, *Journal of Physics.Reveiw* 1957, 106. 4, 620-630.
2. Rubinstein, R.Y.; The Simulated Entropy Method for Combinatorial and Continuous Optimization. *Methodology and Computing in Applied Probability,* 1999, 2, 127–190.
3. Fletcher, R. (1991), *Practical methods of optimization*, 1991,Wiley, New York, NY.
4. Shanonn, C. E. and Weaver, W.; The Mathematical Theory of Communication, 1948, Univ. of Illinois Press, Urbana.
5. Djafari, A. M.; Maximum Entropy Copulas, *AIP Conference Proceeding*, 2011, 329-339.
6. Radon, J.; ber die bestimmung von funktionen durch ihre integralwerte lings gewisser mannigfaltigkeiten, *Ber. Verh. Sïich. Akad. Wiss. Leipzig*, Math. Nat., 1917, 262-277.
7. Thomas, J. A. and Cover, T. M.; Elements of Information Theory, 2006, J. Wiley. New Jersey, NJ.