Maximum Entropy Approach for Reconstructing Bivariate Probability Distributions

Zahra Amini Farsani
Iran University of Science and Technology, Tehran, Iran,
&
Ludwig-Maximilian Universität –Munich, Germany

October 2014

Global Overview

- Purpose
- Methodology
- Findings
- Originality

Purpose

- Establishing an accurate and efficient numerical-probabilistic algorithm based on Newton's technique and Maximum Entropy (ME) method.
- Determining the important bivariate distributions which are very effective in industrial and engineering fields especially in Cybernetics and internet systems.

Methodology

- The design of this paper is to construct a new algorithm involving the combined use of a classical numerical approach, Newton method and a probabilistic method, ME, to find the unique solution of an optimization problem which occurs when maximizing Shannon's Entropy.

Findings

- Conducting different simulation studies for determining different classes of bivariate maximum entropy distributions to check the reliability of the proposed algorithm.

Originality

- Quantifying a method which deals with how to construct a probability distribution using incomplete set of information.
- Maximum entropy method is the only way to choose the distribution based on a finite number of expectation of known functions.
- This method will provide you the unique solution to find a probability distribution based on given information. This is principle of maximum entropy (Jaynes, 1957).
- MATLAB code for univariate & Bivariate cases.

Contents of Entropy

- Introduction & Review of Shannon Entropy
- Maximum Entropy Method

Shannon's Entropy

The ME density is usually obtained by maximizing Shannon's entropy (Shanonn, 1948):

$$h(f) = -\int f(x) log f(x) dx,$$

Where density function $f(x)$ should satisfy in the following constraints:

Introduction

- **Maximum Entropy Probability Distribution** is a probability distribution whose entropy is at least as great as that of all other members of a specified class of distributions.
- Jaynes (1957) has introduced the best approach to ensure that the approximation satisfies any known constraints on the unknown distribution and subject to those constraints, the distribution should have maximum entropy. This is known as the principle of maximum-entropy.

Maximum Entropy

- Maximum Entropy Distribution

Consider the following problem: Maximize the entropy $h(f)$ over all probability densities f satisfying

1. $f(x) \geq 0$, with equality outside the support set S
2. $\int_S f(x)\,dx = 1$
3. $\int_S f(x) r_i(x)\,dx = \alpha_i \quad$ for $1 \leq i \leq m$.

Thus, f is a density on support set S meeting certain moment constraints $\alpha_1, \alpha_2, \ldots, \alpha_m$.

Maximum Entropy

We form the functional

$$J(f) = -\int f \ln f + \lambda_0 \int f + \sum_{i=1}^{m} \lambda_i \int f r_i$$

and "differentiate" with respect to $f(x)$, the xth component of f, to obtain

$$\frac{\partial J}{\partial f(x)} = -\ln f(x) - 1 + \lambda_0 + \sum_{i=1}^{m} \lambda_i r_i(x).$$

Setting this equal to zero, we obtain the form of the maximizing density

$$f(x) = e^{\lambda_0 - 1 + \sum_{i=1}^{m} \lambda_i r_i(x)}, \qquad x \in S,$$

where $\lambda_0, \lambda_1, \ldots, \lambda_m$ are chosen so that f satisfies the constraints.

Maximum Entropy

Example *(One-dimensional gas with a temperature constraint)*
Let the constraints be $EX = 0$ and $EX^2 = \sigma^2$. Then the form of the maximizing distribution is

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

To find the appropriate constants, we first recognize that this distribution has the same form as a normal distribution. Hence, the density that satisfies the constraints and also maximizes the entropy is the $\mathcal{N}(0, \sigma^2)$ distribution:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

Example1: *Bivariate Normal Distribution*

For normal distribution, we consider these constraints:

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

and $a_0 = 1, \quad a_1 = 1, \quad a_2 = \frac{3}{2}.$ Consider $\Omega = (-\infty, \infty)$ as support of X and Y.

Bivariate Normal Distribution

2

Starting with a bivariate normal distribution, the PDF is given by

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

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

Bivariate Normal Distribution

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

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

Bivariate Normal Distribution

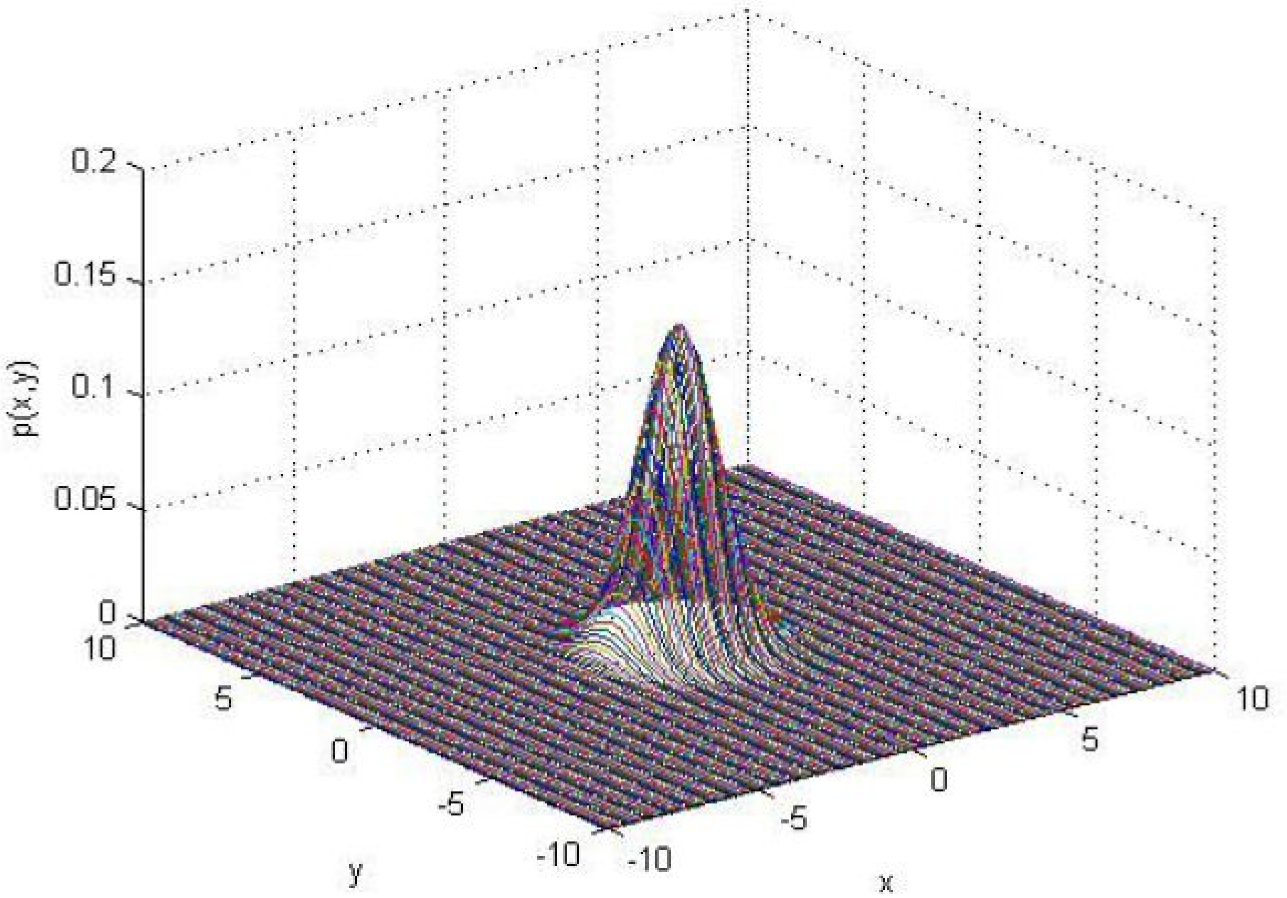

Figure 1: Errors of Maximum Entropy Reconstruction of Bivariate Normal Distribution for

$\tilde{\lambda} = [0.9832, -1.6427, 0.9529], \lambda_{Exact} = [.9516, -1, 1]$ *and* $\lambda_0 = [1.96, .059, .4375]$.

Example2: ***Bivariate Pareto Distribution***

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

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

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

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

Bivariate Pareto Distribution

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 figure 2 for final errors.

Bivariate Pareto Distribution

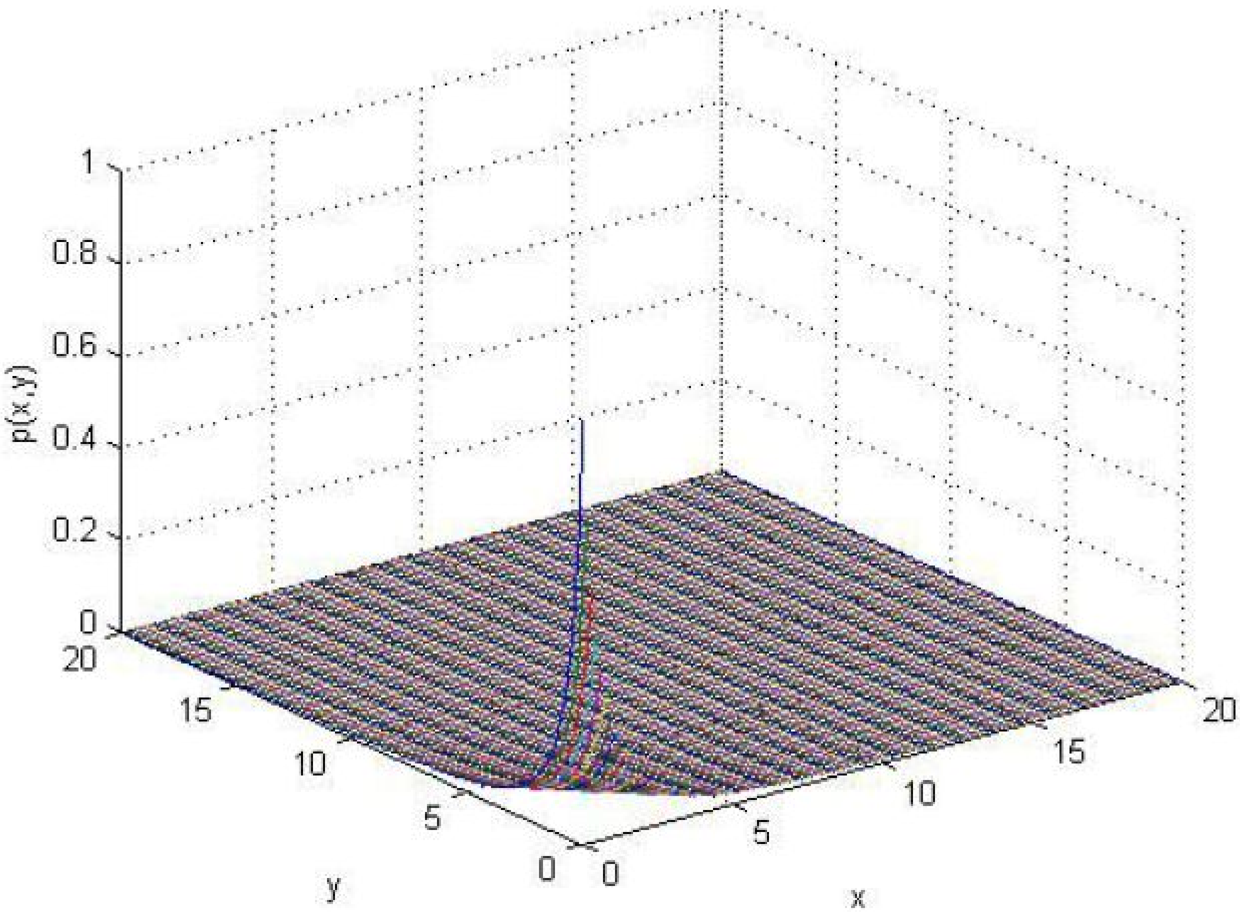

Figure 2 Errors of Maximum Entropy Reconstruction of Bivariate Pareto Distribution with $\lambda_0 = [0.5403, 1.66]$ *and* $\lambda_{exact} = [0.7696, 2.92]$.

References

[1] Fletcher, R. (1991), Practical methods of optimization, Wiley, New York, NY.

[2] Rubinstein, R.Y. (1999). “The Simulated Entropy Method for Combinatorial and Continuous Optimization.” *Methodology and Computing in Applied Probability* 2, 127–190.

[3] E. T.Jaynes, Information Theory and Statistical Mechanics, *Journal of Physics.Reveiw.* 106 (1957) 620-630.

[4] C. E. Shanonn, A Mathematical Theory of Communication, *Bell System Technical Journal*, 27 (1948) 379–423.

[5] J. A. Thomas, T. M. Cover, , Elements of Information Theory, John Wiley, New Jersey, 2006.

[6] Djafari, A. M. (2011), “Maximum Entropy Copulas”, *AIP Conference Proceeding*, pp. 329-339.

[7] Radon, J. (1917), ”ber die bestimmung von funktionen durch ihre integralwerte lings gewisser mannigfaltigkeiten”, *Ber. Verh. Sïich. Akad. Wiss. Leipzig*, Math. Nat. Kl69, pp.262-277.

The End