Reconstruction of the UniVariate & BiVariate Probability Distributions via Maximum Entropy Method

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

December 2014

Global Overview

- Purpose
- Methodology
- Findings

Purpose

- Establishing an accurate and efficient numerical-probabilistic algorithm based on Newton's technique and Maximum Entropy (ME) method.
- Determining the important univariate and bivariate distributions which are very effective in Medical Sciences, Industrial and Engineering fields, Environment and Renewable Energy System and in Computer Science especially in Cybernetics and Internet systems.

Methodology

- The design of all papers is to apply the new proposed algorithm involving the combined use of the Newton method and ME method, to find the unique solution of an optimization problem which occurs when maximizing Shannon's Entropy.

Maximum Entropy Method

- When approximating an unknown probability distribution, an important question is, "what is the best approximation?" In recent years, maximum entropy method has been used to obtain unknown distribution via solving optimization problems. Jaynes (1957) has introduced a principle that the best approximation of an unknown distribution subject to certain constraints should have maximum entropy. This is known as the maximum-entropy principle.
- ME method is one of the most effective way of using limited available information to make an appropriate probability distribution. This well-known method is very useful in many situations in which statistical data communicated with the random variable are either partially available or not available at all. ME allows to obtain maximum possible information from given measurement data of limited accuracy.

Shannon's Entropy

- **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.

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:

Maximum Entropy Method

- ME 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 Method

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 Distribution

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:

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

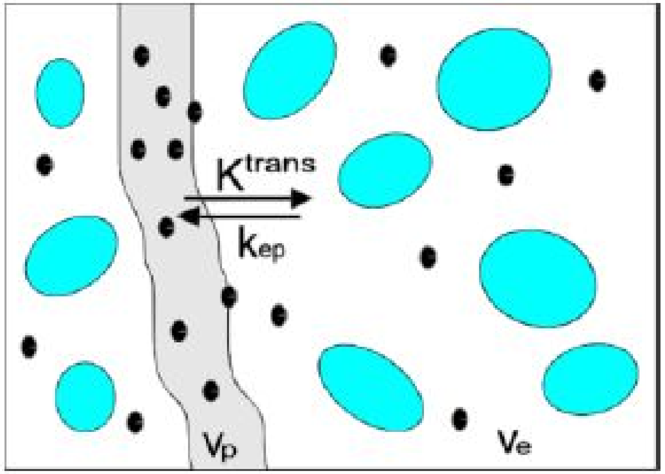
Findings

1. Univariate Maximum Entropy Distribution via a Computational Algorithm (Accepted, 2013)
2. Two-Dimensional Maximum Entropy Distribution via Numerical Algorithm (2014,Under Review)
3. Maximum Entropy Method and its Applications in Renewable Energy Systems (2014-Under Review)
4. Shannon Entropy via Hybrid Numerical Algorithm for Solving an Inverse Problem (2014-Under Review)
5. Maximum Entropy Method for Dynamic Contrast-Enhanced Magnetic Resonance Imaging (2014-Under Review)
6. Bivariate Inverse problem & Maximum Entropy Method (2015)
7. Colocalization & Maximum Entropy Method (2015)

Maximum Entropy Method for Dynamic Contrast-Enhanced Magnetic Resonance Imaging

The purpose of the present study is:

- To introduce the Maximum Entropy (ME) method as a powerful tool for reconstructing images from many types of data.
- To investigate the relation between Maximum A posteriori Baysian (MAP) method and ME method to estimate the Kinetic parameteres of dynamic contrast-enhanced magnetic resonanceimaging (DCE-MRI) and to determine an the prior probability distribution of CA in Plasma.



Compartmental model of kinetics in the tissue. Tracer is delivered via the vascular space, perfuses into the extracellular extravascular space with rate K^{trans}, perfuses back to vascular space with rate k_{ep}, and is eventually washed out

Tracer-Kinetic Model

- Kinetic models are always an approximation to the true physiological processes in the tissue. This model assumes that the contrast agent (CA) in this model abides and exchanges between two sections: in the tissue and the vascular space.
- When the kinetic behavior of the CA in the tissue of interest is considered, we use this model in the form of differential equation system:

Kinetic Model

$$\frac{dp_{C_t}(t)}{dt} = K_1 p_{C_p}(t) K_2 p_{C_t}(t), \qquad (1)$$

where $p_{C_p}(t)$ and $p_{C_t}(t)$ are the concentrations of the contrast agent at time t in the tissue of interest, that is, in the EES, and plasma, respectively. K_1 and K_2 are the rate constants for the exchanges of contrast agent between plasma and EES. Under $p_{C_p(0)} = 0$, Eqn.(1) can be solved and leads to

$$p_{C_t}(t) = K_1. \int_0^t p_{C_p}(t) e^{-K_2(t-u)} du. \qquad (2)$$

A different approach on Eqn.(1) was presented by Murase (2004): We reformulate (1) to

$$p_{C_t}(t) = K_1 \int_0^t p_{C_p}(u) du - K_2 \int_0^t p_{C_t}(u) du. \qquad (3)$$

This can be written in matrix form

$$\boldsymbol{C} = \boldsymbol{A} \times \boldsymbol{B}, \qquad (4)$$

ME & MAP Approaches

in which

$$A(i) = \left(\int_0^{t_i} p_{C_p}(u)du, \int_0^{t_i} p_{C_t}(u)du \right) \text{ for } i = 1, 2, ..., n,$$

$$B = (K_1, K_2)^T,$$

and

$$C = \begin{pmatrix} P_{Ct}(t_1) \\ P_{Ct}(t_2) \\ \vdots \\ P_{Ct}(t_n) \end{pmatrix},$$

when $p_{C_t}(t_i)$ and $p_{C_p}(t_i)$ are known for $i = 1, 2, ..., n.$. We can rewrite Eq.(3) in this form:

$$p_{C_t}(t) = A(i)B + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2). \tag{5}$$

Maximum a Posterior Estimation

Assume that we want to estimate an unobserved population parameter θ on the basis of observations x.

$$\theta \mapsto f(x|\theta)$$

is known as the likelihood function and the estimate:

$$\hat{\theta}_{\mathrm{ML}}(x) = \arg\max_{\theta} f(x|\theta)$$

Now assume that a prior distribution g over θ exists. This allows us to treat θ as a random variable as in Bayesian statistics. Then the posterior distribution of θ is as follows:

$$\theta \mapsto f(\theta|x) = \frac{f(x|\theta)\, g(\theta)}{\displaystyle\int_{\vartheta \in \Theta} f(x|\vartheta)\, g(\vartheta)\, d\vartheta}$$

where g is density function of θ, Θ is the domain of g. This is a straightforward application of Bayes' theorem. The method of maximum a posterior estimation then estimates θ as the mode of the posterior distribution of this random variable:

$$\hat{\theta}_{\mathrm{MAP}}(x) = \arg\max_{\theta} \frac{f(x|\theta)\, g(\theta)}{\displaystyle\int_{\vartheta} f(x|\vartheta)\, g(\vartheta)\, d\vartheta} = \arg\max_{\theta} f(x|\theta)\, g(\theta).$$

Mutual Information

- Mutual information measures the information that X and Y share: it measures how much knowing one of these variables reduces uncertainty about the other. For example, if X and Y are independent, then knowing X does not give any information about Y and vice versa, so their mutual information is zero.
- At the other extreme, if X is a deterministic function of Y and Y is a deterministic function of X then all information conveyed by X is shared with Y: knowing X determines the value of Y and vice versa. As a result, in this case the mutual information is the same as the uncertainty contained in Y (or X) alone, namely the entropy of Y (or X). Moreover, this mutual information is the same as the entropy of X and as the entropy of Y. (A very special case of this is when X and Y are the same random variable.)

$$I(X;Y) = H(X) - H(X|Y)$$

ME & MAP Approaches

ME density is usually obtained by maximizing Shannon's entropy, (Thomas and cover, 2006):

$$H(K) = -\int p(k) log p(k) dk. \tag{5}$$

subject to $\quad E(\phi_i(k)) = \int \phi_i(k) p(k) dx = \mu_i, i = 0, ..., N$

where $\phi_1, ..., \phi_m$ are $N+1$ known functions, and $\mu_i, i = 0, ..., N$ are the given expectation data. The solution of the ME problem is given by

$$p(k) = e^{-\sum_{i=0}^{N} \lambda_i \phi_i(k)} \qquad x \in S, \tag{6}$$

ME & MAP Approaches

λ_i should be chosen such that $p(k)$ satisfies the known moment constraints (Djafari, 2011). Now, if we determine $p_{C_t}(t|k)$ and $p(k)$, we can solve the following system:

$$\widehat{K} = argmax_{x>0}\ p_{C_t}(k|t) = argmax_{x>0}\{p_{C_t}(t|k)p(k)\}.$$

We use MEM to find a probability density $p_{C_t}(t|k)$ in this way:

$$p_{C_t}(t|k) = \exp[-T(K)] \text{ with } T(K) = \frac{[C_{tis} - AK]^t[C_{tis} - AK]}{\sigma^2}. \quad (8)$$

Empirical Application

- This dataset involves vectors of $C_p(t_i)$, t and $C_{tis}(t_i)$ for $i = 1,...,46$. We have established a MATLAB code for the new algorithm. The first step should be to find the probability density of $C_p(t)$ Consider these constraints:

$$\begin{cases} \int_t p_{C_p}(t)dt & = 1 \\ \int_t t p_{C_p}(t)dt & = 0.6963625369 \end{cases}$$

and

$$\begin{cases} Normalization \quad \phi_0(t) & = 1 \\ \phi_1(t) & = t \end{cases}$$

Prior Model for CA in Plasma

Then, the density which satisfied these constraints and maximize the Shannon entropy is in the form of:

$$p_{C_p}(t) = e^{-0.10655061745-1.26042322763t} + 0.5556.$$

This is the primary model for data $C_p(t)$ changing the inverse problem to forward. So, the mean of absolute error between exact data and estimated model is 0.071. Since k is strongly depends on dataset $C_p(t)$, so we can use $p_{C_p}(t)$ as prior information for estimating K, then we put it $p(k)$.

Modeling the CA in Tissue

Here, we use ME method again to determine $p_{C_t}(t)$. Suppose that:

$$\begin{cases} \int_t p_{C_t}(t)dt & = 1 \\ \int_t t p_{C_t}(t)dt & = 0.7136921 \\ \int_x Log(t) p_{C_t}(t)dt & = -0.78612099 \end{cases}$$

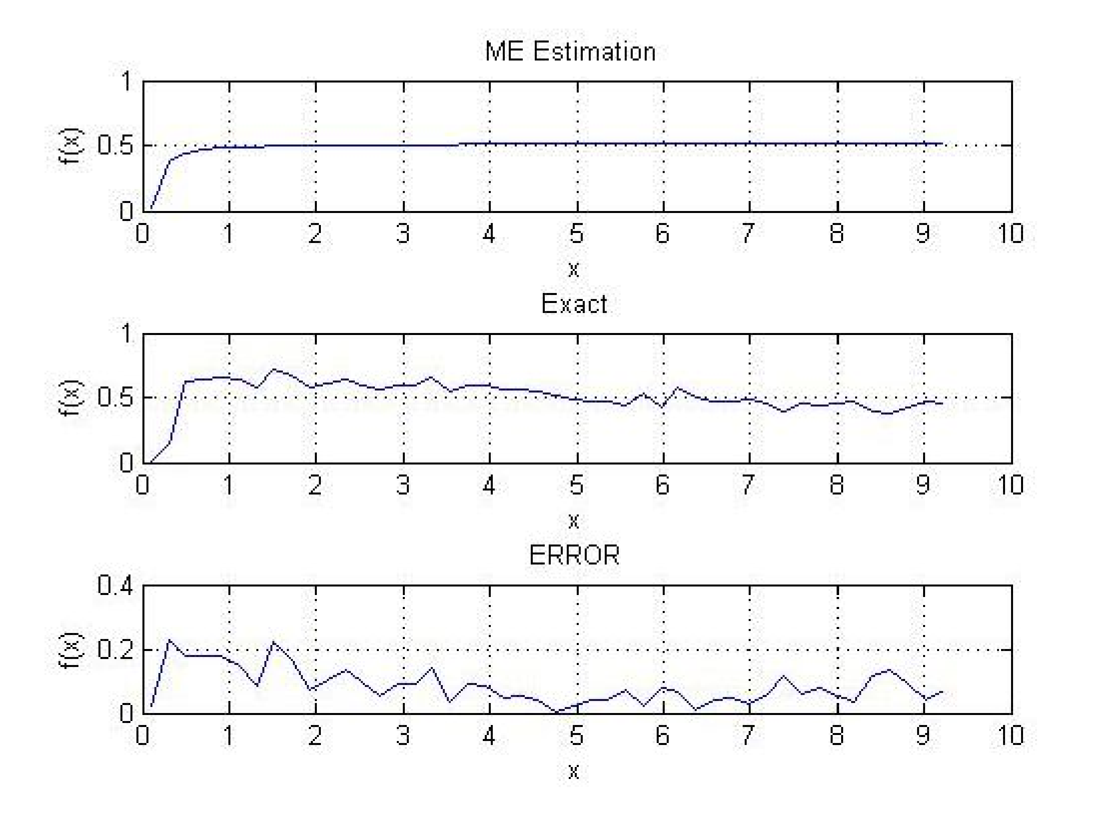
and

$$\begin{cases} Normalization \quad \phi_0(C_p) & = 1 \\ \phi_1(t) & = t \\ \phi_2(t) & = Log(t) \end{cases}$$

The estimated probability distribution subject to these constraints is:

$$p_{C_t(t)} = -e^{-3.42-0.07145t-1.1759Log(t)} + 0.5171.$$

CA in Tissue



ME Estimation
f(x)
1
0.5
0
0 1 2 3 4 5 6 7 8 9 10
x
Exact
f(x)
1
0.5
0
0 1 2 3 4 5 6 7 8 9 10
x
ERROR
f(x)
0.4
0.2
0
0 1 2 3 4 5 6 7 8 9 10
x

Estimation of Kinetic Parameters

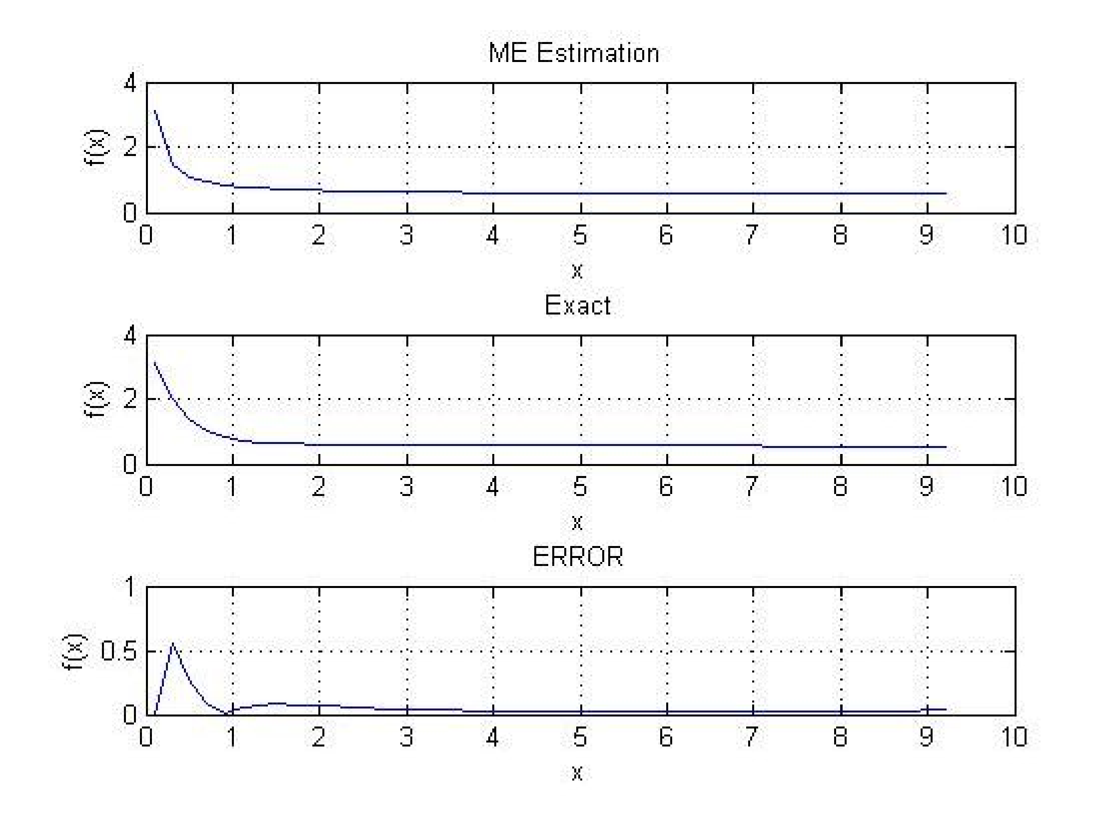
$$\begin{cases} \widehat{K_1} = 0.965822 \\ \widehat{K_2} = 1.146418 \end{cases}$$

In this step, for solving inverse problem to find $p_{C_p}(t)$, we can put $p_{C_t}(t)$, K_1 and K_2 in Eq.(2), therefore we have final for of $p_{C_p}(t)$:

$$p_{C_p(t)} = -e^{-3.2486-0.07145t-1.1759Log(t)} + 0.5398 + e^{-1.1473-Log(t)},$$

and the mean of absolute error for the final $p_{C_p}(t)$ is 0.0515. Figure (1) shows a good agreement between the exact data and ME estimated model:

Final model of CA in Plasma



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.

[8] A. M.-Djafari , G. Demoment (1990), Estimating Priors In Maximum Entropy Image Processing in Proceedings of IEEE ICASSP. pp:2069-2072.

[9] V. J. Schmid (2010), Kinetic Models for Cancer Imaging,Advances in Experimental Medicine and Biology Volume 680, 2010, pp 549-557

The End