The Maximum Entropy Method for Reconstructing Density Distributions

Tony Saad*
University of Utah, Salt Lake City, UT 84112, USA

I. Introduction

The method of moments is a very useful approach in the solution of transport equations for density distributions. This procedure however smoothes much of the information contained in the continuous distribution. The resulting moments are useful for inferring integral information about a distribution such as the mean number of particles and their mean size. However, it is sometimes desirable to calculate the continuous distribution. For example, when calculating the death rate of particles, one must know the number of particles at a specified particle size. In other situations, it is quite useful to look at the evolution of the density distribution instead of the moments.

The maximum entropy method provides an elegant means of reconstructing a density distribution given a finite number of moments. This is commonly known as the *moments problem*. The precise statement of this problem is as follows: given a finite number of moments, is it possible to find a unique distribution that gives rise to these moments? In general, there are infinitely many continuous distributions whose moments match the known moments. Additional constraints are then required to guide the process of finding a continuous distribution that fits the known moments. The maximum entropy method is one such such constraint.

II. Maximum Entropy

The maximum entropy method is based on the concept that *the distribution that maximizes the information entropy is the one that is statistically most likely to occur*. In the context of information theory, the information entropy S, of a distribution $p(x)$, is given by the integral

$$S = -\int_{\Omega} p(x) \ln \frac{p(x)}{M(x)} \mathrm{d}x \tag{1}$$

where Ω is the support of the distribution and $M(x)$ is a prior distribution expressing our prior knowledge. Then, one looks for a distribution $p(x)$ that maximizes S subject to the known moments. This is then formulated as a variational problem whose solution may be easily obtained.

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

$$\int_{\Omega} x^k p(x) \mathrm{d}x = \mu_k; \quad k = 0, 1, \ldots, N \tag{2}$$

*Research Associate, Institute for Clean and Secure Energy.

where $(N+1)$ is the number of known moments. Note that μ_k is known for $0 \leq k \leq N$. Then, by introducing Lagrangian multipliers λ_k, we define the entropy functional

$$H \equiv S + \sum \lambda_k \left(\int_\Omega x^k p(x) \mathrm{d}x - \mu_k \right) = - \int_\Omega p(x) \ln \frac{p(x)}{M(x)} \mathrm{d}x + \sum_{k=0}^{N} \lambda_k \left(\int_\Omega x^k p(x) \mathrm{d}x - \mu_k \right) \tag{3}$$

This functional is a maximum when

$$\frac{\partial H}{\partial \lambda_k} = 0 \tag{4}$$

$$\frac{\partial H}{\partial p(x)} = 0 \tag{5}$$

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

$$\frac{\partial H}{\partial p(x)} = - \int_\Omega \left[\ln \frac{p(x)}{M(x)} + 1 \right] \mathrm{d}x + \sum_{k=0}^{N} \lambda_k \int_\Omega x^k \mathrm{d}x = 0 \tag{6}$$

Since the above integrals must be valid on an arbitrary domain Ω, the integrand must be zero. At the outset, we get

$$- \ln \frac{p(x)}{M(x)} - 1 + \sum_{k=0}^{N} \lambda_k x^k = 0 \tag{7}$$

whose general solution is

$$p(x) = M(x)\, e^{\sum_{k=0}^{N} \lambda_k x^k} \tag{8}$$

Note that we have combined the (-1) term in Eq. (7) with λ_0, i.e. $\lambda_0 = \lambda_0 - 1$.

III. Solution

To find the maximum entropy solution, we have to solve for the Lagrangian multipliers λ_k. For this, we must solve the following nonlinear system of equations

$$\begin{aligned} \int_\Omega M(x) e^{\lambda_0 + \lambda_1 x + \cdots + \lambda_k x^k} \mathrm{d}x &= \mu_0 \\ \int_\Omega x\, M(x) e^{\lambda_0 + \lambda_1 x + \cdots + \lambda_k x^k} \mathrm{d}x &= \mu_1 \\ &\vdots \\ \int_\Omega x^k M(x) e^{\lambda_0 + \lambda_1 x + \cdots + \lambda_k x^k} \mathrm{d}x &= \mu_k \end{aligned} \tag{9}$$

A globally convergent Newton solver may be used to calculate λ_k.

The Jacobian can also be easily calculated. If we denote the moments based on the maximum entropy solution as

$$\widetilde{\mu}_k \equiv \int_\Omega x^k M(x) e^{\lambda_0 + \lambda_1 x + \cdots + \lambda_k x^k} \mathrm{d}x \tag{10}$$

then, the Jacobian is given by

$$
\boldsymbol{J} \equiv \begin{bmatrix} \frac{\partial\widetilde{\mu}_0}{\partial\lambda_0} & \frac{\partial\widetilde{\mu}_0}{\partial\lambda_1} & \cdots & \frac{\partial\widetilde{\mu}_0}{\partial\lambda_k} \\ \frac{\partial\widetilde{\mu}_1}{\partial\lambda_0} & \frac{\partial\widetilde{\mu}_1}{\partial\lambda_1} & \cdots & \frac{\partial\widetilde{\mu}_1}{\partial\lambda_k} \\ \frac{\partial\widetilde{\mu}_2}{\partial\lambda_0} & \frac{\partial\widetilde{\mu}_2}{\partial\lambda_1} & \cdots & \frac{\partial\widetilde{\mu}_2}{\partial\lambda_k} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial\widetilde{\mu}_k}{\partial\lambda_0} & \frac{\partial\widetilde{\mu}_k}{\partial\lambda_1} & \cdots & \frac{\partial\widetilde{\mu}_k}{\partial\lambda_k} \end{bmatrix} = \begin{bmatrix} \widetilde{\mu}_0 & \widetilde{\mu}_1 & \cdots & \widetilde{\mu}_k \\ \widetilde{\mu}_1 & \widetilde{\mu}_2 & \cdots & \widetilde{\mu}_{k+1} \\ \widetilde{\mu}_2 & \widetilde{\mu}_3 & \cdots & \widetilde{\mu}_{k+2} \\ \vdots & \vdots & \vdots & \vdots \\ \widetilde{\mu}_k & \widetilde{\mu}_{k+1} & \cdots & \widetilde{\mu}_{k+k} \end{bmatrix} \tag{11}
$$

Note that $\tilde{\mu}_k$ is the k-th moment based on the reconstructed distribution. The Newton solver will be based on finding λ_k such that $(\tilde{\mu}_k - \mu_k)$ is below a certain specified precision.

A. Initial Guesses

The Newton method can be sensitive to the initial conditions used. However, I found it to converge in all of the tests if the following initial guesses are used

$$
\begin{cases} -1 < \lambda_0^{\text{initial}} < 0 & \\ \lambda_i^{\text{initial}} = 0 & i > 0 \end{cases} \tag{12}
$$

I used $\lambda_0^{\text{initial}} = -\ln\sqrt{2\pi}$ in my code. This value is based on a Gaussian distribution with zero mean and unit variance ($\mu = 0$, $\sigma = 1$).

IV. Implementation and Results

I implemented the maximum entropy method in C using the GNU Scientific Library using their multidimensional nonlinear solver and their quadrature numerical integration routines. I implemented two interfaces for the solver, one that uses the analytical expression for the Jacobian (fdf_maxent_solve), and one that uses finite differencing to estimate the Jacobian (f_maxent_solve).

To test the code, I generated a number of moments using known distributions. I used a tolerance of 10^{-10} for the Newton solver.

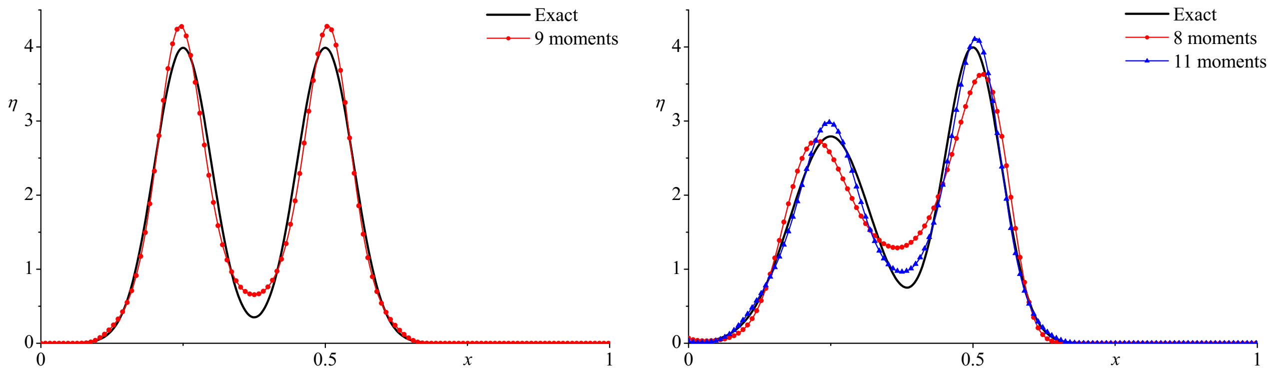
A. Distributions with Finite Support

1. Gaussian Distributions

For gaussian distributions we consider double and triple Gaussians respectively. Starting with a double Gaussian, the PDF is given by

$$
f(x) = \frac{1}{2\sigma_0\sqrt{2\pi}}\text{Exp}\left[-\frac{(x-\mu_0)^2}{2\sigma_0^2}\right] + \frac{1}{2\sigma_1\sqrt{2\pi}}\text{Exp}\left[-\frac{(x-\mu_1)^2}{2{\sigma_1}^2}\right] \tag{13}
$$

To construct a double Gaussian with peaks of equal size, one sets $\sigma_0 = \sigma_1$. This is shown in Fig. 1a where the exact distribution corresponds to the solid black line while the maximum entropy reconstruction is shown in red. The reconstructed PDF overestimates both peaks as well as the trough, but captures the overall trend of the exact distribution.



(a) Double Gaussian distribution with $\mu_0 = 1/4,\ \mu_1 = 3/4,\ \sigma_0 = \sigma_1 = 1/20$.

(b) Double Gaussian distribution with $\mu_0 = 1/4,\ \mu_1 = 2/4,\ \sigma_0 = 1/14,\ \sigma_1 = 1/20$.

Figure 1. Maximum entropy reconstruction for bimodal Gaussian distributions, (a) with peaks of equal heights and (b) different sized peaks.

Figure (1b) shows a comparison between a bimodal Gaussian with peaks of different heights and the maximum entropy reconstruction using 8 and 11 moments, respectively. Evidently, the larger the number of known moments, the more precise is the reconstruction. In this case, however, the Newton solver may not converge to the desired accuracy.

For a trimodal Gaussian distribution, we use

$$f(x) = \frac{1}{3\sigma_0\sqrt{2\pi}}\mathrm{Exp}\left[-\frac{(x-\mu_0)^2}{2\sigma_0^2}\right] + \frac{1}{3\sigma_1\sqrt{2\pi}}\mathrm{Exp}\left[-\frac{(x-\mu_1)^2}{2{\sigma_1}^2}\right] + \frac{1}{3\sigma_2\sqrt{2\pi}}\mathrm{Exp}\left[-\frac{(x-\mu_2)^2}{2{\sigma_2}^2}\right] \quad (14)$$

The reconstructed PDF is shown in Fig. 2 where 8 and 13 moments are used to reconstruct the exact distribution. In this case, one needs higher order moments to capture the changes in convexity of this distribution.

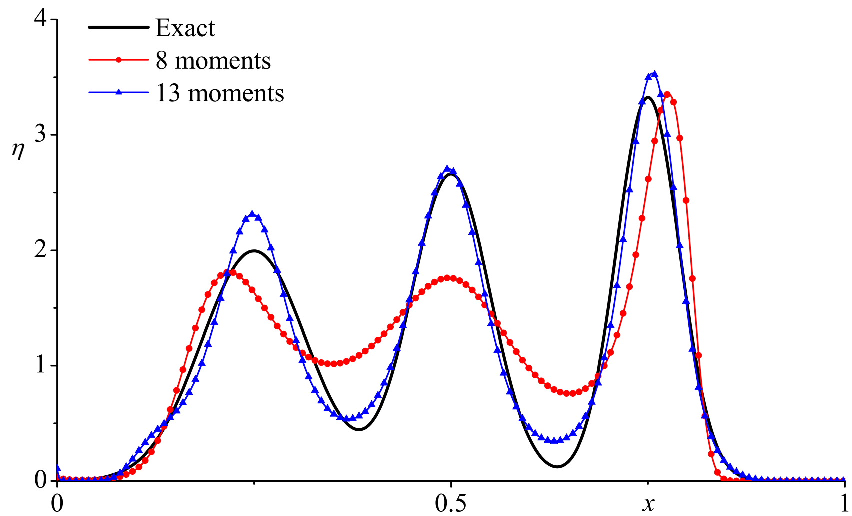
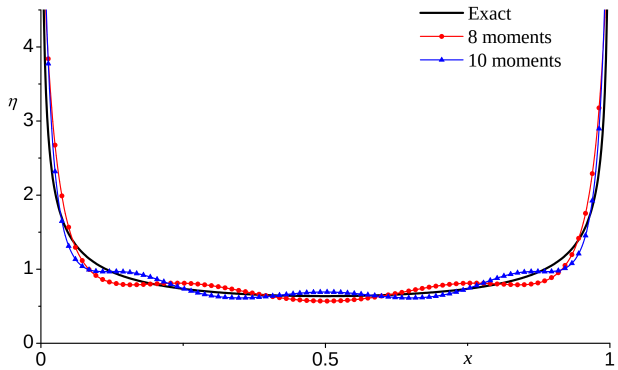
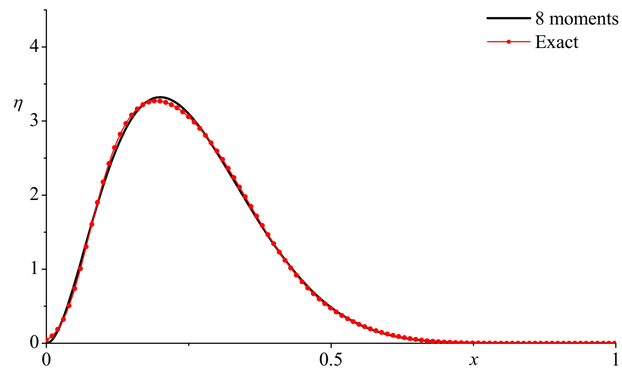


Figure 2. Maximum entropy reconstruction for a trimodal Gaussian distribution with $\mu_0 = 1/4,\ \mu_1 = 2/4,\ \mu_2 = 3/4,\ \sigma_0 = 1/15\,\sigma_1 = 1/20,\ \sigma_2 = 1/25$**.**



(a) Beta distribution with $\alpha = \beta = 0.5$.



(b) Beta distribution with $\alpha = 3$ and $\beta = 9$.

Figure 3. Maximum entropy reconstruction the beta distribution.

2. Beta Distribution

Results for the beta distribution are shown in Fig. 3. The maximum entropy reconstruction is able to capture the key features of this distribution especially for the case with singularities at the domain boundaries.

B. Distributions with Semi-Infinite Support

I ran into trouble with the numerical integrator when integrating over an infinite or a semi-infinite support. This behaviour is observed especially with distributions that slowly decay towards infinity such as a log-normal distribution. However, if integration is carried out on a finite support, one is able to still recover a reasonable approximation to the distribution. As an example of this, a log-normal distribution was used with numerical integration performed over the interval $[0, 10]$. The resulting reconstrubtion is shown in Fig. 4.

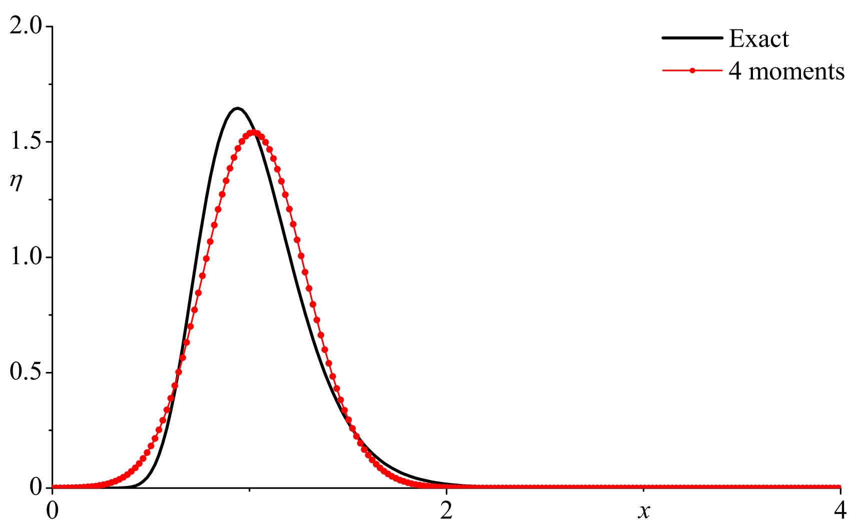


Figure 4. log-normal distribution with $\mu = 0$ and $\sigma = 1/4$.

Bibliography

[1]Bandyopadhyay, K., Bhattacharya, A. K., Biswas, P., and Drabold, D. A., "Maximum entropy and the problem of moments: A stable algorithm," *Physical Review E*, Vol. 71, No. 5, May 2005, pp. 057701.

[2]Erdogmus, D., Hild, K. E., Rao, Y. N., and Príncipe, J. C., "Minimax Mutual Information Approach for Independent Component Analysis," *Neural Computation*, Vol. 16, No. 6, June 2011, pp. 1235–1252.

[3]Pope, S. B., "A Rational Method of Determining Probability Distributions in Turbulent Reacting Flows," *Journal of Non-Equilibrium Thermodynamics*, Vol. 4, No. 5, 1979, pp. 309–320.

[4]Pope, S. B., "The Statistical Theory of Turbulent Flames," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 291, No. 1384, July 1979, pp. 529 –568.

[5]Tagliani, A., "On the application of maximum entropy to the moments problem," *Journal of Mathematical Physics*, Vol. 34, No. 1, 1993, pp. 326.

[6]Wu, Z., Phillips, G. N., Tapia, R., and Zhang, Y., "A Fast Newton Algorithm for Entropy Maximization in Phase Determination," *SIAM Review*, Vol. 43, No. 4, 2001, pp. 623.

[7]"Landau : Maximum entropy and the moment problem," *Bulletin of the American Mathematical Society*, Vol. 16, No. 1, 1987, pp. 47—77.