Constructing and Estimating Probability Distributions from Moments

J. Tekel and L. Cohen[1]

City University of New York, 695 Park Ave. New York, NY 10065 USA

We present two methods to obtain a probability distribution from its moments. In the first method one chooses a set of orthogonal polynomials with corresponding weighting function. In the second method one chooses a starting distribution and uses that to construct orthogonal polynomials. We give a number of examples.

I. Introduction

In previous publications we considered the problem of constructing probability distributions from its moments and have the results to a number of problems. In this paper we apply the method to some exactly solvable problems and numerically and graphically compare the exact and approximate results. Suppose we have a set of orthogonal polynomials $L_n(x)$ with weighting function $w(x)$. We define the complete set of functions by

$$u_n(x) = \frac{1}{\sqrt{N_n}} \sqrt{w(x)} L_n(x) \tag{1}$$

where N_n are normalizing numbers so that

$$\int w(x) L_n(x) L_m(x) dx = N_n \delta_{nm} \tag{2}$$

and hence

$$\int u_n(x) u_m(x) dx = \delta_{nm} \tag{3}$$

We assume that all quantities are real. Any function, $f(x)$ can be expanded as

$$f(x) = \sum_{n=0}^{\infty} c_n u_n(x) \tag{4}$$

with

$$c_n = \int f(x) u_n(x) dx \tag{5}$$

[1]Work supported by the Office of Naval Research.

Expansion of a probability density. For a probability density $P(x)$ we expand it as [3, 5]

$$P(x) = \sqrt{w(x)} \sum_{n=0}^{\infty} c_n u_n(x) \tag{6}$$

The reason for inserting $\sqrt{w(x)}$ is so that the coefficients c_n will be functions of the moments *only*, as we now show. Using Eq. (5) with $f(x) = P(x)/\sqrt{w(x)}$, the coefficients, c_n, are

$$c_n = \int \frac{P(x)}{\sqrt{w(x)}} u_n(x) dx = \frac{1}{\sqrt{N_n}} \int P(x) L_n(x) = \frac{1}{\sqrt{N_n}} \langle L_n(x) \rangle \tag{7}$$

Since $L_n(x)$ are polynomials $\langle L_n(x) \rangle$ can be constructed from the moments and hence so can the c_n. The distribution is then constructed by way of

$$P(x) = \sqrt{w(x)} \sum_{n=0}^{\infty} \frac{1}{\sqrt{N_n}} \langle L_n(x) \rangle \, u_n(x) \tag{8}$$

or

$$P(x) = w(x) \sum_{n=0}^{\infty} \frac{1}{N_n} \langle L_n(x) \rangle \, L_n(x) \tag{9}$$

II. Explicit Initial Distribution

In the above formulation the weighting function and the polynomials are chosen and fixed. If the probability distribution that we have fits the weighting function then the above expansion is fine. However suppose we have a probability distribution and we want to construct the polynomials. We now show how that can be done [5]. That will allow us to relate two probability distributions and their moments. We first describe the standard theory for obtaining orthogonal polynomials for a given weighting function. Suppose we define [1, 4]

$$m_j = \int x^j w(x) dx \tag{10}$$

then the following polynomials [1, 4]

$$K_n(x) = \begin{vmatrix} m_0 & m_1 & m_2 & \cdots & m_n \\ m_1 & m_2 & m_3 & \cdots & m_{n+1} \\ m_2 & m_3 & m_4 & \cdots & m_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1} & m_n & m_{n+1} & \cdots & m_{2n-1} \\ 1 & x & x^2 & \cdots & x^n \end{vmatrix} \tag{11}$$

will be an orthogonal set in the sense that

$$\int w(x) K_n^*(x) K_m(x) dx = N_n \delta_{nm} \tag{12}$$

where the normalizing factors N_n, are given by

$$N_n = \begin{vmatrix} m_0 & m_1 & m_2 & \cdots & m_{n-1} \\ m_1 & m_2 & m_3 & \cdots & m_n \\ m_2 & m_3 & m_4 & \cdots & m_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1} & m_n & m_{n+1} & \cdots & m_{2n-2} \end{vmatrix} \begin{vmatrix} m_0 & m_1 & m_2 & \cdots & m_n \\ m_1 & m_2 & m_3 & \cdots & m_{n+1} \\ m_2 & m_3 & m_4 & \cdots & m_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1} & m_n & m_{n+1} & \cdots & m_{2n-1} \\ m_n & m_{n+1} & m_{n+2} & \cdots & m_{2n} \end{vmatrix} \tag{13}$$

Note that the normalizing factor N_n depends on moments up to $2n$. The first few polynomials are

$$K_0(x) = m_0 \tag{14}$$

$$K_1(x) = \begin{vmatrix} m_0 & m_1 \\ 1 & x \end{vmatrix} = x - m_1 \tag{15}$$

$$K_2(x) = \begin{vmatrix} m_0 & m_1 & m_2 \\ m_1 & m_2 & m_3 \\ 1 & x & x^2 \end{vmatrix} = (m_2 - m_1^2)x^2 + (m_1 m_2 - m_3)x + m_1 m_3 - m_2^2 \tag{16}$$

Now, to apply this formulation to our aim of constructing probability distributions we take the probability $P(x)$, to be the weighting function [5]

$$w(x) = P(x) \tag{17}$$

and furthermore since we assume the probability is normalized we have that

$$m_0 = 1 \tag{18}$$

and the rest of the m's are then the moments of $P(x)$.

Now suppose we have another probability distribution $P'(x)$ with moments m_j' . As in the previous section we can write

$$P'(x) = P(x) \sum_{n=0}^{\infty} \frac{1}{N_n} \langle K_n(x) \rangle' \; K_n(x) \tag{19}$$

In Eq. (19) $\langle K_n(x) \rangle'$ is given by

$$\langle K_n(x) \rangle' = \int dx \, K_n(x) P'(x) \tag{20}$$

or explicitly

$$\langle K_n(x) \rangle' = \begin{vmatrix} 1 & m_1 & m_2 & \cdots & m_n \\ m_1 & m_2 & m_3 & \cdots & m_{n+1} \\ m_2 & m_3 & m_4 & \cdots & m_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1} & m_n & m_{n+1} & \cdots & m_{2n-1} \\ \int P'(x)dx & \int xP'(x)dx & \int x^2 P'(x)dx & \cdots & \int x^n P'(x)dx \end{vmatrix} \tag{21}$$

$$= \begin{vmatrix} 1 & m_1 & m_2 & \cdots & m_n \\ m_1 & m_2 & m_3 & \cdots & m_{n+1} \\ m_2 & m_3 & m_4 & \cdots & m_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1} & m_n & m_{n+1} & \cdots & m_{2n-1} \\ 1 & m'_1 & m'_2 & \cdots & m'_n \end{vmatrix} \tag{22}$$

Note that $\langle K_n(x) \rangle'$ involves the moments of both distributions. Therefore we have,

$$P'(x,t) = P(x) \sum_{n=0}^{\infty} \frac{1}{N_n} \begin{vmatrix} 1 & m_1 & m_2 & \cdots & m_n \\ m_1 & m_2 & m_3 & \cdots & m_{n+1} \\ m_2 & m_3 & m_4 & \cdots & m_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1} & m_n & m_{n+1} & \cdots & m_{2n-1} \\ 1 & m'_1 & m'_2 & \cdots & m'_n \end{vmatrix} \begin{vmatrix} 1 & m_1 & m_2 & \cdots & m_n \\ m_1 & m_2 & m_3 & \cdots & m_{n+1} \\ m_2 & m_3 & m_4 & \cdots & m_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1} & m_n & m_{n+1} & \cdots & m_{2n-1} \\ 1 & x & x^2 & \cdots & x^n \end{vmatrix} \tag{23}$$

Explicitly, the first few terms are

$$P'(x) = P(x) \left(1 + \frac{1}{N_1} \begin{vmatrix} 1 & m_1 \\ 1 & m'_1 \end{vmatrix} K_1(x) + \frac{1}{N_2} \begin{vmatrix} 1 & m_1 & m_2 \\ m_1 & m_2 & m_3 \\ 1 & m'_1 & m'_2 \end{vmatrix} K_2(x) ... \right) \tag{24}$$

Note that if the moments are the same then indeed $P'(x) = P(x)$ because only the first term survives. That is the case since in the determinant the first and last rows are the same.

The above formulation relates two probability distributions and their moments. There are other ways one can relate probability distributions, such as the Gram-Charlier or Edgeworth series and their generalization [2]. The comparison of the method presented here with those other methods will be discussed elsewhere. The advantage of this formulation is that it expands the probabilities in terms of orthogonal polynomials and explicitly brings in the moments in a relatively simple way. We now define the approximate probability of order M by

$$\tilde{P}'_M(x) = P(x) \sum_{n=0}^{M} \frac{1}{N_n} \langle K_n(x) \rangle' \ K_n(x) \tag{25}$$

III. Examples

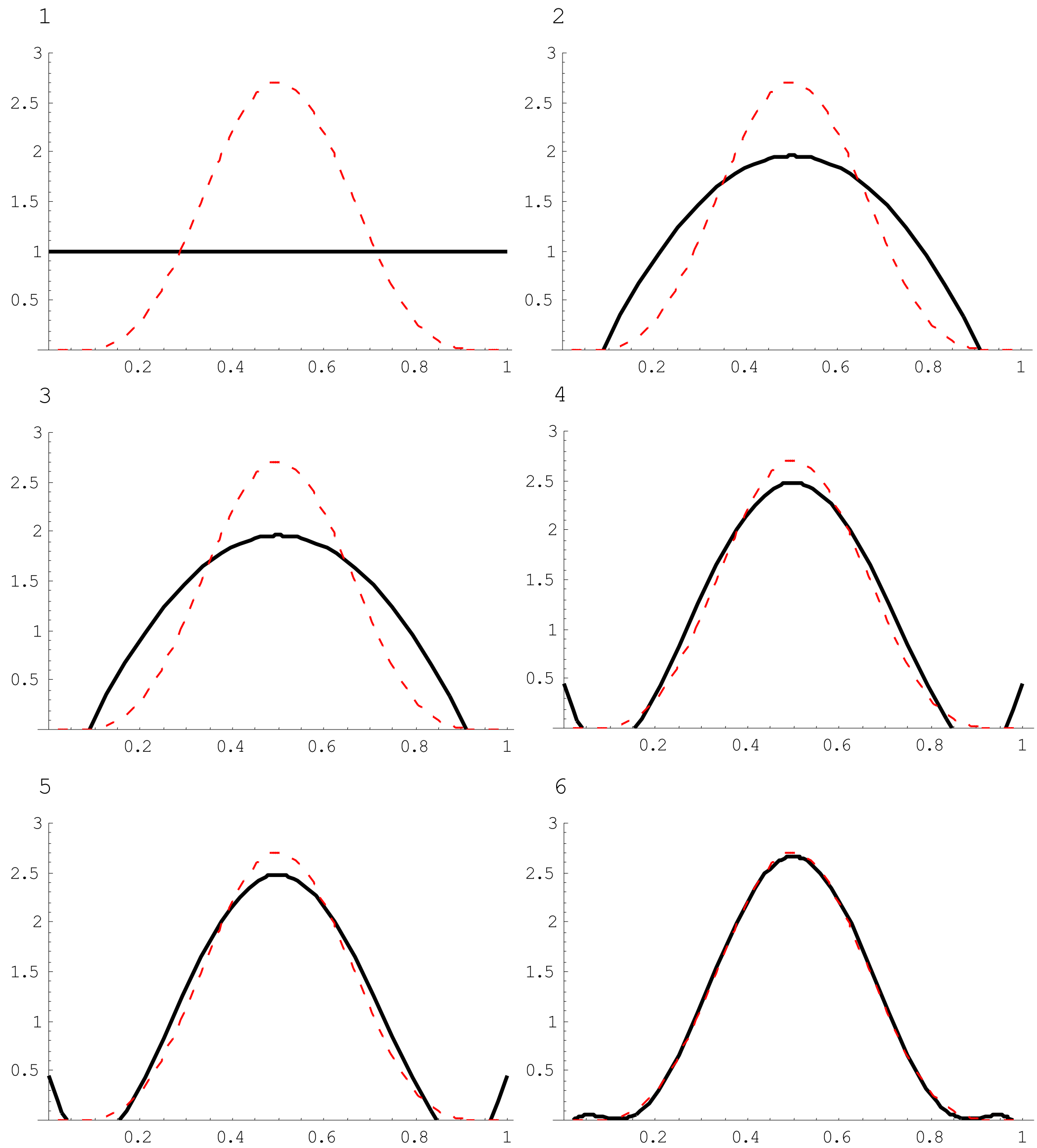
We now consider a number of examples comparing $\tilde{P}'_M(x)$ with the exact solution and examine how it depends on the value of M. In all the cases $P(x)$ and $P'(x)$ are normalized to one. In the figures, the number above the figure is the M value. The bold line is the approximation, $\tilde{P}_M(x)$, and the exact solution, $P'(x)$, is the dotted line.

Example 1. We aim to approximate

$$P'(x) = 2772x^5(1-x)^5 \qquad , \qquad 0 \le x \le 1 \tag{26}$$

from the starting distribution

$$P(x) = 1 \qquad , \qquad 0 \le x \le 1 \tag{27}$$



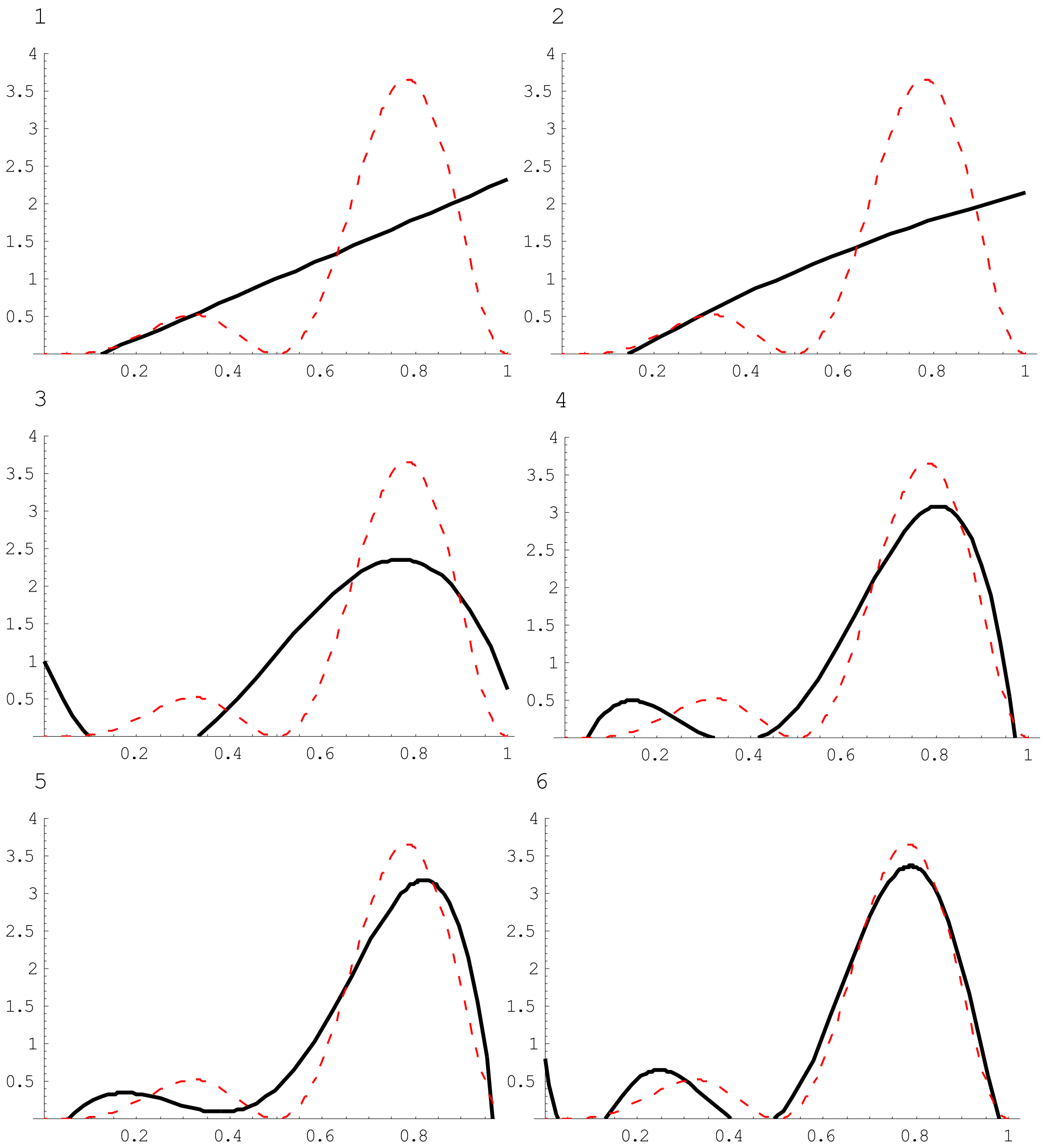
We clearly see that after a few iterations we have found the exact formula and the graphs overlap indicating an excellent approximation.

Example 2. For this example we take

$$P'(x) = \frac{48\pi^2}{8\pi^2 - 3} x^2 \sin^2(2\pi x) \qquad , \qquad 0 \le x \le 1$$

and the starting distribution is

$$P(x) = 1 \qquad , \qquad 0 \le x \le 1 \tag{28}$$

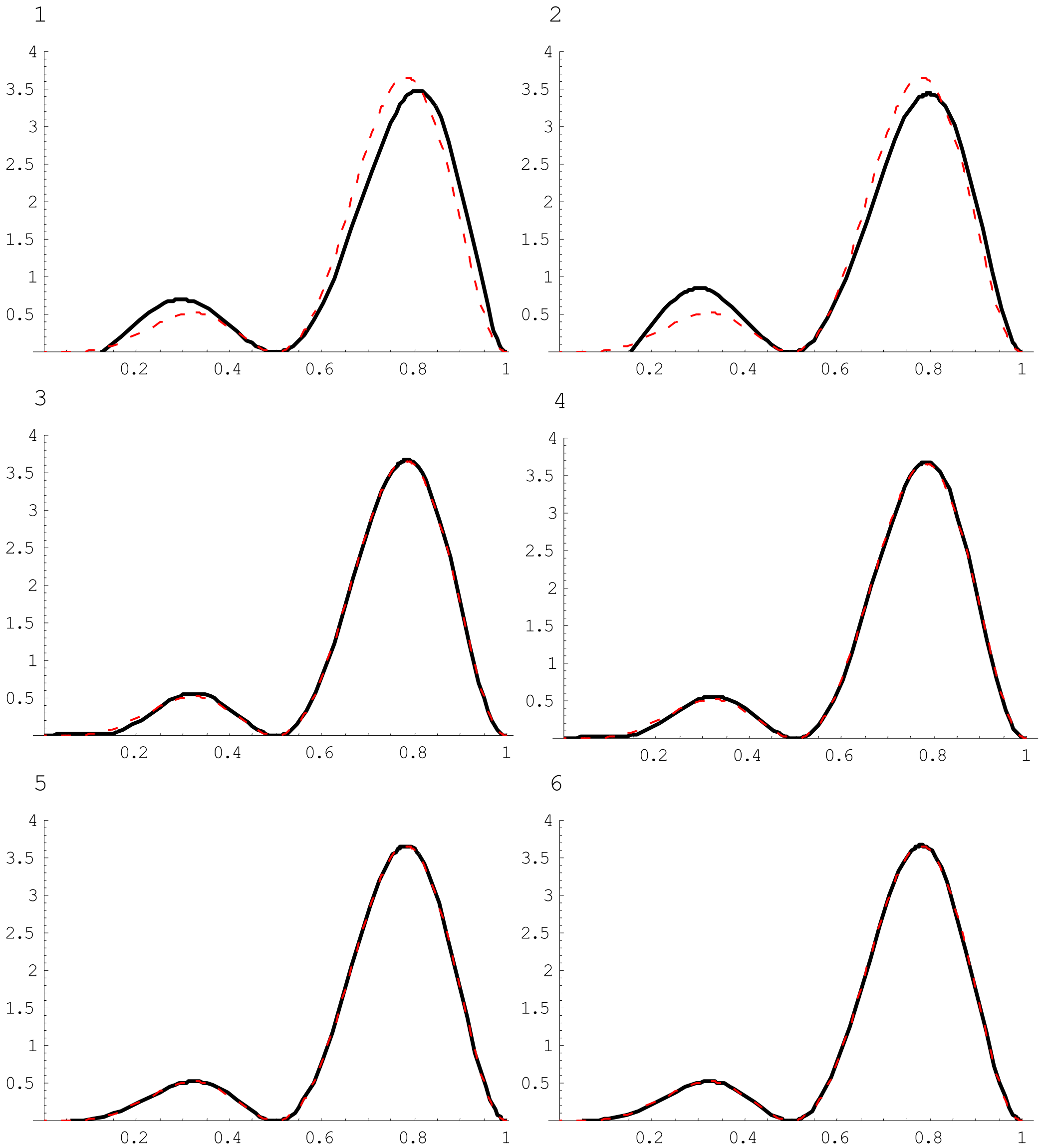


Example 3. The exact is

$$P'(x) = \frac{48\pi^2}{8\pi^2 - 3} x^2 \sin^2(2\pi x) \qquad , \qquad 0 \le x \le 1$$

and the starting distribution is

$$P(x) = 840\,[x(1/2 - x)(1 - x)]^2 \qquad , \qquad 0 \le x \le 1 \tag{29}$$

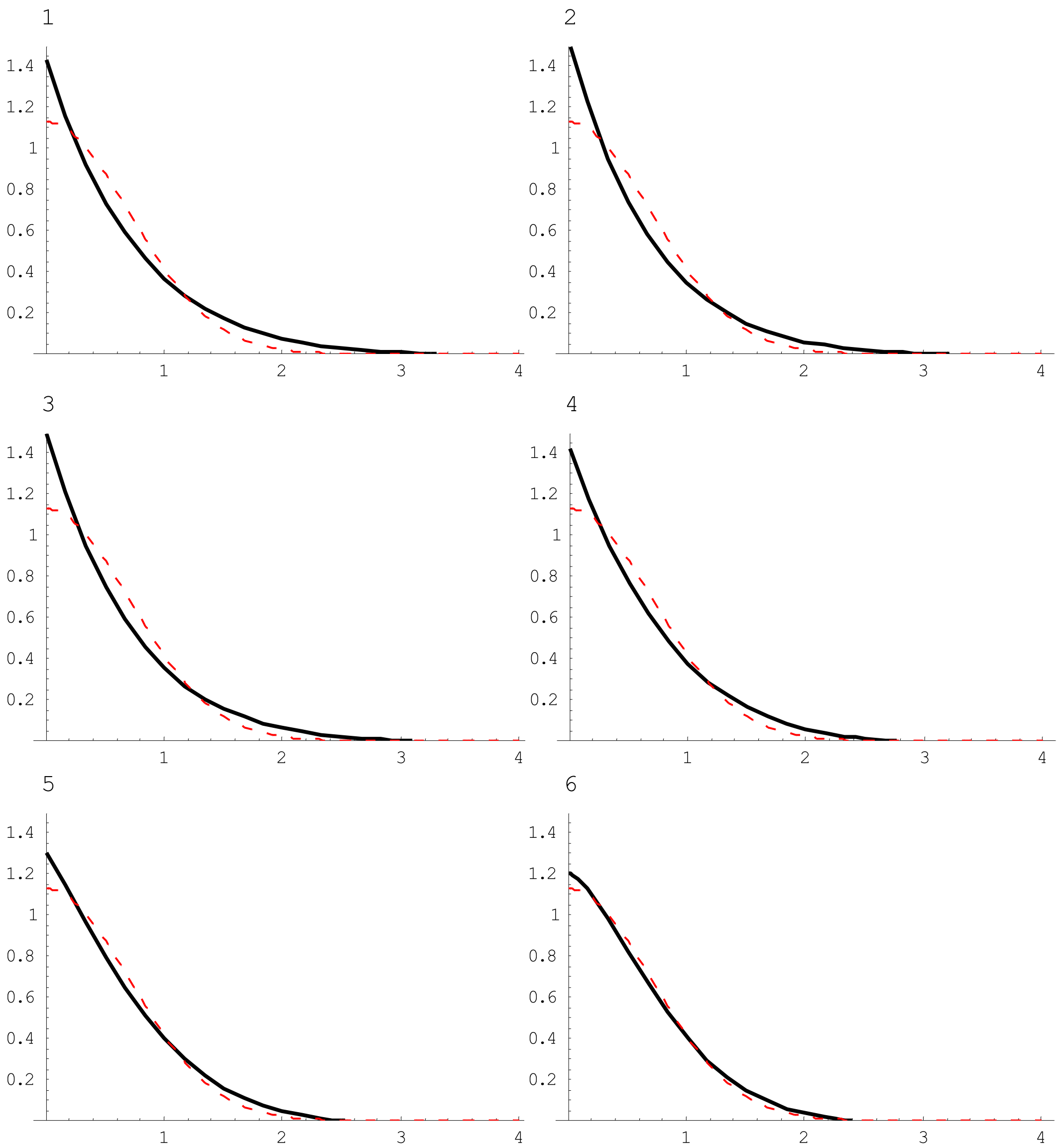


Example 4 In this example the interval is the positive x axis. For the exact we take

$$P'(x) = \frac{1}{2\sqrt{\pi}} e^{-x^2} \qquad , \qquad 0 \le x \le \infty$$

and the starting distribution is taken to be

$$P(x) = xe^{-x} \qquad , \qquad 0 \le x \le \infty \tag{30}$$

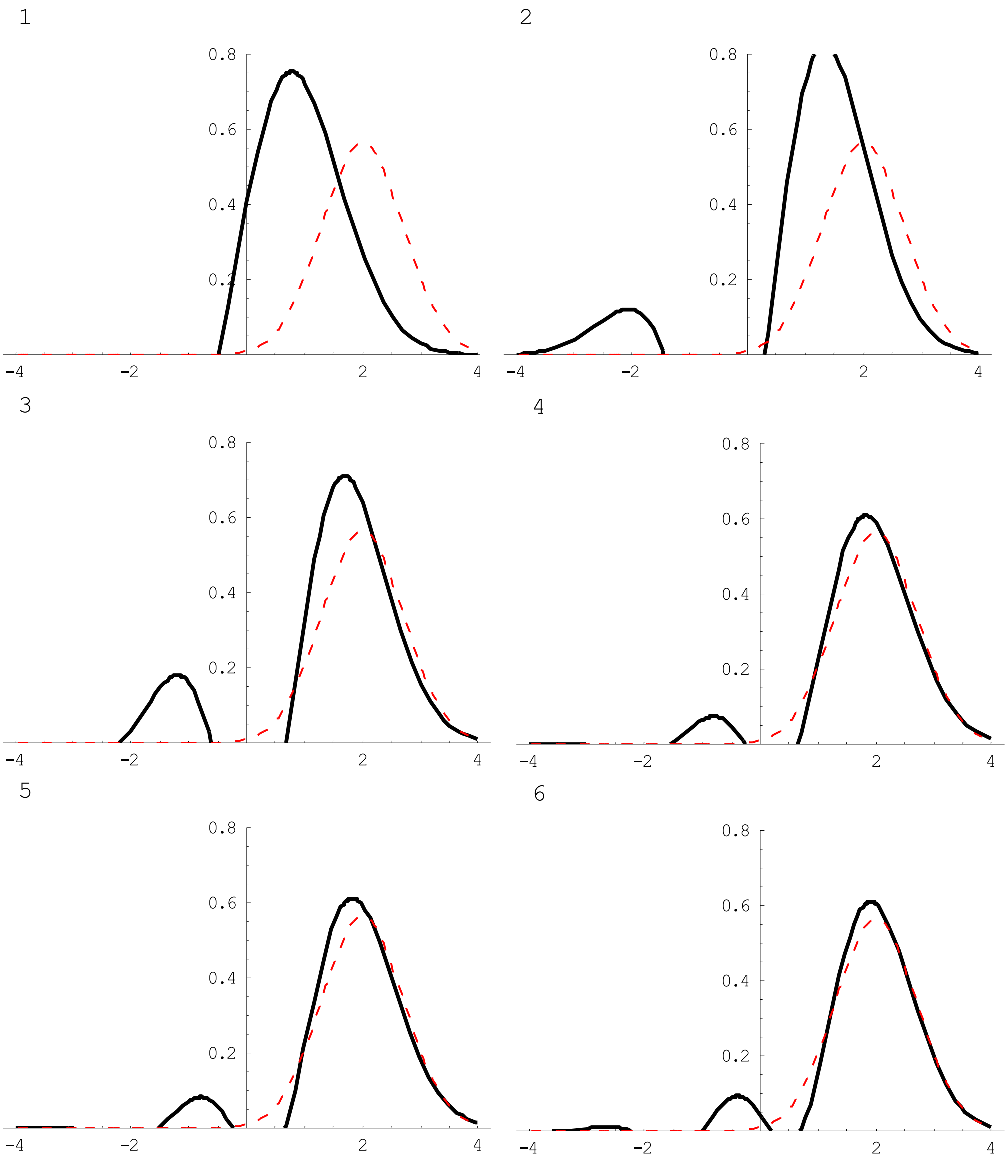


Example 5. We consider two Gaussian distributions where one is a shifted version of the other. For the exact we take

$$P'(x) = \frac{1}{\sqrt{\pi}} e^{-(x-2)^2} \qquad , \qquad \infty \leq x \leq \infty$$

and the starting distribution is

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad , \qquad 0 \leq x \leq \infty \tag{31}$$



IV. Conclusion

We have presented two methods for obtaining a probability distribution from its moments. In the first method one chooses a set of orthogonal polynomials with corresponding weighting function. In the second method one chooses a starting distribution and uses that to construct the orthogonal polynomials. We have presented a number of examples to show that the method works very well.

References

[1] G. E. Andrews, R. Askey, R. Roy, *Special Functions*, Cambridge University Press, 2000.

[2] L. Cohen, "Construction and transformation of probability densities", Integrated Computer-Aided Engineering, Volume 12, Number 1, pp. 129 - 134, 2005.

[3] L. Cohen, K. Davidson and P. Loughlin, "Construction of Time-Frequency Representations From Moments" Proc. SPIE, Vol. 6697, pp B1-B8, 2007.

[4] G. Szego, *Orthogonal Polynomials*, American Mathematical Society, 1975.

[5] J. Tekel and L. Cohen, "Transformation of time dependent probability distributions" , Proc. SPIE 7335, 051-058, 2009.