A moments based distribution bounding method*

Sándor Rácz, Árpád Tari, Miklós Telek
Budapest University of Technology and Economics
Magyar tudósok krt. 2, 1117 Budapest, Hungary,
e-mail: {raczs,arpi,telek}@webspn.hit.bme.hu

Abstract

We present a numerical procedure to bound distributions based on their finite number of moments. Numerical examples demonstrate the properties of the proposed method.
Keywords: Estimation of distribution, Moment problem.

1. Introduction

There are large discrete state stochastic models in which the numerical analysis of the distribution of the measure of interest is computationally infeasible, but the first some moments of this measure are computable [1]. Based on these moments one can approximate the distribution (with an unknown error) or calculate upper and lower bounds of the cumulative distribution function (cdf). [2, 3] presents a numerical procedure for moments based distribution bounding, but the proof of the correctness of this procedure was not available. In the present paper, we prove the moments based bounds and show their properties via numerical examples. There are excellent monographs summarizing the theory of moments [4, 5, 6]. Our results are largely based on the fundamental book of Akhiezer [6].

The approach of this paper differs from the traditional statistician approach [7], because we assume that the exact value of the first n moments are known and we look for the minimal and maximal value of the cdf of all distributions with these moments. Indeed, we evaluate how much information is carried by the first n moments about a distribution.

The rest of this paper is organized as follows. Section 2 collects the basic properties of moments applied in our distribution estimation method. Section 3 provides a high level description of the proposed method, while symbolic and numerical results are provided in Section 4 and 5, respectively. Section 6 concludes the paper.

2. Basic properties of moments

Following [6] we introduce a set of notations.

*This work was partially supported by Hungarian Scientific Research Fund (OTKA) under Grant No. T-34972.

$$\mathfrak{M}_{2n} = \text{the set of distributions with the same } 0,1,\ldots,2n \text{ moments.} \quad (1)$$

$$\sigma(\cdot) = \text{a non-decreasing function} \quad (\sigma(x_1) \leq \sigma(x_2) \text{ if } x_1 \leq x_2). \quad (2)$$

$$\mu_i = \int_{-\infty}^{\infty} x^i \, d\sigma(x) \quad (i = 0,1,2,\ldots), \quad (3)$$

"the ith moment" of $\sigma(\cdot)$.

$$D_n = \begin{vmatrix} \mu_0 & \mu_1 & \ldots & \mu_n \\ \mu_1 & \mu_2 & \ldots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_{n+1} & \ldots & \mu_{2n} \end{vmatrix}, \quad (4)$$

the Hankel determinant of order n.

$$P_0(x) = 1 \quad \text{and} \quad P_n(x) = \frac{1}{\sqrt{D_{n-1}D_n}} \begin{vmatrix} \mu_0 & \mu_1 & \ldots & \mu_n \\ \mu_1 & \mu_2 & \ldots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n-1} & \mu_n & \ldots & \mu_{2n-1} \\ 1 & x & \ldots & x^n \end{vmatrix}, \quad (n = 1,2,\ldots), \quad (5)$$

orthonormal polynomials composed by the μ_i sequence.

$$\rho_n(x) = \frac{1}{\sum_{k=0}^{n} |P_k(x)|^2}, \quad (6)$$

the radius of the Hellinger circle [6].

Definition 1. *The μ_i $(i = 0,1,\ldots,2n)$ sequence is a* positive sequence *if $D_k > 0$ $(k = 0,1,\ldots,n)$.*

Definition 2. *The μ_i $(i = 0,1,\ldots,2n)$ sequence is a* determined sequence *if $D_k > 0$ $(k = 0,1,\ldots,n-1)$ and $D_n = 0$. (In this case $D_k = 0$ $(\forall k \geq n)$ [8].)*

2.1. The Hamburger moment problem and its solvability

The Hamburger moment problem [9] plays an essential role in the theory of moments [4, 7]. It can be formulated as follows. Given a sequence of numbers μ_i $(i = 0,1,2,\ldots)$, under what conditions is it possible to find a positive bounded non-decreasing function $\sigma(\cdot)$ such that

$$\mu_i = \int_a^b x^i d\sigma(x) \quad , \qquad \text{for } a = -\infty, \; b = \infty, \; i = 0,1,2,\ldots \quad .$$

The cases, when $a = 0$, $b = \infty$, and $a = 0$, $b = 1$, are referred to Stieltjes [8] and Hausdorff [10] moment problem, respectively.

Theorem 1. *[6] Let $\mu_0, \mu_1, \mu_2, \ldots$ be a sequence of real numbers. The Hamburger moment problem has a solution if and only if $D_n \geq 0$, $n = 0,1,\ldots$.*

Theorem 2. *[11] The solution of the Hamburger moment problem consists of infinite points of increase if and only if $D_n > 0$, $\quad n = 0,1,\ldots$.*

Theorem 3. *[11] The solution of the Hamburger moment problem consists of exactly n distinct points of increase if and only if $D_0 > 0, D_1 > 0 \ldots D_{n-1} > 0, D_n = D_{n+1} = \ldots = 0$. The moment problem is determined in this case.*

An immediate consequence of Theorem 2 and 3 is that if μ_i are the moments of a distribution and $D_n = 0$ then all the higher Hankel determinants equal to 0, as well ($D_k = 0$ for all $k > n$).

2.2. Finite number of moments

Theorems 1 - 3 are about the infinite series μ_i and D_n, but in this paper, we have a finite number of moments to deal with. To bound a distribution based on its first $2n+1$ moments[1] we need to find the extreme members of the $\mathfrak{M}_{2n}$ class. At an arbitrary point C, the $\sigma(\cdot)$ distribution with positive sequence of moments $\mu_0, \ldots, \mu_{2n}$ is bounded by $\min_{\breve{\sigma} \in \mathfrak{M}_{2n}} \breve{\sigma}(C) \le \sigma(C) \le \max_{\hat{\sigma} \in \mathfrak{M}_{2n}} \hat{\sigma}(C)$. In the rest of this paper we investigate $\min_{\breve{\sigma} \in \mathfrak{M}_{2n}} \breve{\sigma}(C)$ and $\max_{\hat{\sigma} \in \mathfrak{M}_{2n}} \hat{\sigma}(C)$ in two steps. The first step is to determine the maximum mass the members of $\mathfrak{M}_{2n}$ can have at C, and the second step is to construct a distribution having this maximal mass at C. It will be shown that there is only one distribution composed by $n+1$ discrete points (including the one at C) with maximal mass in C and this distribution characterizes both the lower and the upper bound of the $\mathfrak{M}_{2n}$ class at C.

To simplify the discussion, we always study the bounds at point 0 with a proper transformation of moments. If the original point of interest is C then the moments of the distribution whose evaluated point is shifted to 0 are:

$$\mu'_i = \int_{-\infty}^{\infty} (x-C)^i d\sigma(x) = \int_{-\infty}^{\infty} \sum_{k=0}^{i} \binom{i}{k} x^k (-C)^{i-k} d\sigma(x) = \sum_{k=0}^{i} \binom{i}{k} (-C)^{i-k} \mu_k. \tag{7}$$

Without loss of generality, from now on we assume that the point of interest is 0.

2.3. Maximum mass concentrated at 0

Theorem 4. *[6] If $\mu_0, \mu_1, \ldots, \mu_{2n}$ is a positive sequence (i.e., $D_k > 0$, $(k = 0, \ldots, n)$) and x is an arbitrary real number then,*

$$\max_{\sigma(\cdot) \in \mathfrak{M}_{2n}} (\sigma(x^+) - \sigma(x^-)) \le \rho_n(x) \ . \tag{8}$$

Theorem 4 gives the meaning of the introduction of $\rho_n(x)$. Indeed, $\rho_n(x)$ defines the maximal mass that can be located at point x given the first $2n+1$ moments. Following a completely different way of thinking than the one in [6], we obtained a different and computationally more effective way to determine maximal mass.

Theorem 5. *[12, 4] If $\mu_0 = 1, \mu_1, \mu_2, \ldots, \mu_{2n}$ is a positive sequence of moments of $\sigma(\cdot)$ the*

[1] Throughout this paper the first k moments mean the $\mu_0, \mu_1, \ldots, \mu_{k-1}$ sequence.

maximal mass of $\sigma(\cdot)$ at 0 is

$$p = \frac{\begin{vmatrix} \mu_0 & \mu_1 & \dots & \mu_n \\ \mu_1 & \mu_2 & \dots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_{n+1} & \dots & \mu_{2n} \end{vmatrix}}{\begin{vmatrix} \mu_2 & \mu_3 & \dots & \mu_{n+1} \\ \mu_3 & \mu_4 & \dots & \mu_{n+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n+1} & \mu_{n+2} & \dots & \mu_{2n} \end{vmatrix}}, \tag{9}$$

which means that

$$\rho_n(0) = p. \tag{10}$$

Appendix A presents a proof of this theorem different from the ones in [12, 4].

Our way to obtain p was rather intuitive. A mass at point 0 does contribute to μ_0, but does not contribute to any $\mu_i, i > 0$. We locate a mass at 0 such that the Hankel determinant of the $\mu_0 - p, \mu_1, \dots, \mu_{2n}$ sequence is just on the limit of positivity. Using $2n+1$ moments the limit of positivity is reached at

$$\begin{vmatrix} \mu_0 - p & \mu_1 & \dots & \mu_n \\ \mu_1 & \mu_2 & \dots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_{n+1} & \dots & \mu_{2n} \end{vmatrix} = 0, \tag{11}$$

whose solution is given by Theorem 5.

Theorem 6. *If $\mu_0, \mu_1, \mu_2, \dots, \mu_{2n}$ is a positive sequence of moments (i.e., $D_k > 0$ ($k = 0, \dots, n$)) then the $\mu_0 - p, \mu_1, \dots, \mu_{2n}$ sequence represents a determined moment problem (i.e., $D'_k > 0$ ($k = 0, \dots, n-1$) and $D'_n = 0$).*

Proof: Since p is the solution of (11) the Hankel determinant of order n associated with the $\mu_0 - p, \mu_1, \mu_2, \dots, \mu_{2n}$ sequence equals to zero and using Theorem 3 it implies Theorem 6. ■

2.4. Maximum difference of the distribution bounds

The following theorem underlines the importance of $\rho_n(0)$, (and p), by giving an additional meaning to them.

Theorem 7. *[6] If $\mu_0, \mu_1, \mu_2, \dots, \mu_{2n}$ is a positive sequence, $\sigma_1(\cdot) \in \mathfrak{M}_{2n}$ and $\sigma_2(\cdot) \in \mathfrak{M}_{2n}$ then, for arbitrary real x we have:*

$$\left| \int_{-\infty}^{x^+} d\sigma_1(u) - \int_{-\infty}^{x^-} d\sigma_2(u) \right| \le \rho_n(x). \tag{12}$$

Theorem 7 provides the maximum difference of any two members of $\mathfrak{M}_{2n}$ at x. A direct consequence of Theorem 7 is that the difference of $\min_{\breve{\sigma} \in \mathfrak{M}_{2n}} \breve{\sigma}(x)$ and $\max_{\hat{\sigma} \in \mathfrak{M}_{2n}} \hat{\sigma}(x)$ cannot be larger than $\rho_n(x)$.

Having the difference between the lower and upper bounds it is enough to find one of them. The following theorem suggests a way to place the bounds.

Theorem 8. *If $\mu_0, \mu_1, \mu_2, \ldots, \mu_{2n}$ is a positive sequence, $\sigma(\cdot) \in \mathfrak{M}_{2n}$, $\sigma^*(\cdot) \in \mathfrak{M}_{2n}$ and $\sigma^*(\cdot)$ has the mass $\rho_n(0) = p$ at 0*

$$\int_{-\infty}^{0^+} d\sigma^*(u) - \int_{-\infty}^{0^-} d\sigma^*(u) = \rho_n(0) = p,$$

then,

$$\begin{aligned} \int_{-\infty}^{0^-} d\sigma(u) \ &\geq \ \int_{-\infty}^{0^-} d\sigma^*(u), \qquad (13) \\ \int_{-\infty}^{0^+} d\sigma(u) \ &\leq \ \int_{-\infty}^{0^-} d\sigma^*(u) + p. \qquad (14) \end{aligned}$$

The proof of Theorem 8 is provided in Appendix B.

2.5. Construction of a reference distribution

According to Theorem 8 we have the bounds of the $\mathfrak{M}_{2n}$ class of distributions at 0 if we can obtain a reference distribution $\sigma^*(\cdot)$. $\sigma^*(\cdot)$ has the mass of size p at 0 and $\mu_0 - p, \mu_1, \ldots, \mu_{2n}$ is a determined sequence, i.e., $D_k > 0$ $(k = 0, 1, \ldots, n-1)$ and $D_n = 0$. Since $\mu_0 - p, \mu_1, \ldots, \mu_{2n}$ is a determined sequence $\sigma^*(\cdot)$ is unique and it has exactly n further points of increase (Theorem 3 and 6).

Let x_i and p_i $(i = 1, \ldots, n)$ denote the points and the associated value of increase of $\sigma^*(\cdot)$ excluding the one at 0, respectively. x_i and p_i are defined by the moments:

$$\mu_0 - p = \sum_{i=1}^{n} p_i, \qquad \mu_k = \sum_{i=1}^{n} x_i^k \, p_i, \qquad (k = 1, 2, \ldots, 2n-1). \qquad (15)$$

These $2n$ equations can be solved in 2 steps.

Theorem 9. *[6] x_i $(i = 1, \ldots, n)$ are the roots of the polynomial $P_n(x)$ defined by the sequence $\mu_0 - p, \mu_1, \ldots, \mu_{2n}$, (see [5]).*

Having the points $x_1, x_2, \ldots, x_n$ the associated p_i values can be obtained from equation (15) with $k = 0, 1, \ldots, n-1$. In matrix form it can be given as

$$\begin{pmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \\ x_1^2 & x_2^2 & \ldots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \ldots & x_n^{n-1} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} \mu_0 - p \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{n-1} \end{pmatrix}, \qquad (16)$$

which is a Vandermonde system and can be solved efficiently using the algorithm provided, e.g., in [13].

2.6. Special cases, roots with the same sign

If all the roots of $P_n(x)$, $x_i < 0$ $(i = 1, 2, \ldots, n)$ or $x_i > 0$ $(i = 1, 2, \ldots, n)$ we can bound the $\mathfrak{M}_{2n}$ class without calculating the unknown x_is and p_is. This property can be checked without finding the roots of $P_n(x)$ by the *Liènard–Chipart criterion.*

Theorem 10. *[14] **(Liènard–Chipart)** Let $f(x)$ be a polynomial of order n,*

$$f(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n, \tag{17}$$

and T_i be the following sequence of determinants:

$$T_0 = a_0, \qquad T_1 = a_1, \qquad T_2 = \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix}, \qquad T_3 = \begin{vmatrix} a_1 & a_0 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{vmatrix}, \qquad \ldots,$$

$$T_i = \begin{vmatrix} a_1 & a_0 & 0 & 0 & 0 & \ldots & 0 \\ a_3 & a_2 & a_1 & a_0 & 0 & \ldots & 0 \\ a_5 & a_4 & a_3 & a_2 & a_1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{2i-1} & a_{2i-2} & a_{2i-3} & a_{2i-4} & a_{2i-5} & \ldots & a_i \end{vmatrix}. \tag{18}$$

If $a_0 > 0$ then the real part of the roots of $f(x)$ are all negative if and only if $T_i > 0$ ($i = 0, 1, \ldots, n$).

We can also check if all the roots of $f(x)$ have positive real parts using Theorem 10. Let $\hat{f}(x) = f(-x)$. The coefficients of $\hat{f}(x)$ can be expressed by the ones of $f(x)$:

$$\hat{a}_i = (-1)^{n-i} a_i \ , \qquad i = 0, 1, \ldots, n \ . \tag{19}$$

If all roots of $\hat{f}(x)$ have negative real parts then all roots of $f(x)$ have positive real part.

In the case when all roots of $P_n(x)$ are negative ($x_i < 0$ ($i = 1, 2, \ldots, n$)) the bounds are

$$\min_{\breve{\sigma} \in \mathfrak{M}_{2n}} \breve{\sigma}(x) = \mu_0 - p \ , \qquad \max_{\hat{\sigma} \in \mathfrak{M}_{2n}} \hat{\sigma}(x) = \mu_0 \ , \tag{20}$$

and when all roots of $P_n(x)$ ($x_1, x_2, \ldots, x_n$) are positive the bounds are

$$\min_{\breve{\sigma} \in \mathfrak{M}_{2n}} \breve{\sigma}(x) = 0 \ , \qquad \max_{\hat{\sigma} \in \mathfrak{M}_{2n}} \hat{\sigma}(x) = p \ . \tag{21}$$

3. The algorithm

Based on the above mentioned general rules of moments we construct a numerical method in this section. The method provides an upper and a lower bound of a distribution at a given point C based on the first $2n+1$ moments, $\mu_0, \mu_1, \mu_2, \ldots, \mu_{2n}$, assuming $\mu_0 = 1$. The method can be used with any positive μ_0 as well, but in this case $\rho_n(0)$ must be calculated by (6).

The main steps of the method are:

1. Checking the number of moments: we need an odd number of moments greater than 1 (including the 0th one).

2. Checking the D_k ($k = 0, 1, \ldots, n$) sequence:

   - If there is a k such that $D_k < 0$ then the μ_is can not be the moments of a non-decreasing distribution function.

- If there is a k such that $D_j = 0$ ($\forall j \geq k$) then the μ_is define a unique discrete distribution of k points and the discrete construction step of the procedure generate this distribution. (I.e., in this case the method provides the exact value of the distribution function.)
- If all D_k are positive (i.e., μ_i is a positive sequence) then there exists a set of distributions having these first $2n+1$ moments and we calculate the lower and upper bounds of this set at point C in the following steps of the algorithm.

3. Transforming the moments such that the point of interest is moved to 0 by (7).
4. Determining the maximum mass concentrated at 0 by (9).
5. Checking if 0 is the leftmost or rightmost point of the reference discrete distribution via Theorem 10.
   - If 0 is the rightmost or leftmost point the bounds are given by (20) or (21), respectively.
   - If $P_n(x)$ has positive and negative roots as well, then the procedure is continued with the following steps.
6. Determining the roots of $P_n(x)$ which are the points of the reference discrete distribution by Theorem 9.
7. Calculating the weights of the reference discrete distribution by (16).
8. Determining the lower and upper limits of the distribution at point 0 based on the sum of weights associated with negative roots and the maximum mass at point 0.

A block diagram of the method is presented in Figure 1.

4. Symbolic results

The symbolic bounds are not available for more than 9 moments because does not exist general symbolic solution for the roots of 5th and higher order polynomials. One can calculate the symbolic bounds for 7 moments but they are too complex to be presented here.

In this section we assume that the moments are transformed such that the point of interest is 0.

4.1. Estimation based on 3 moments

Having the following sequence of moments: $\mu_0 = 1, \mu_1, \mu_2$ we distinguish two legal cases. When $D_1 = 0$ the moments define a unique distribution (see Theorem 6). When $D_1 > 0$ we can bound the limits of all distributions having the same first 2 moments. The case when $D_1 < 0$ can not be obtained by the moments of a real distribution.
Case 1) determined sequence of moments: If $D_1 = \mu_2 - \mu_1^2 = 0$ then the moments determine a discrete distribution with only one point (indeed, a deterministic distribution):

$$x_1 = \mu_1, \qquad p_1 = 1. \tag{22}$$

Case 2) positive sequence of moments: If $D_1 = \mu_2 - \mu_1^2 > 0$ then we evaluate the bounds based on a discrete distribution with 2 points. One point is at 0 (where we need to bound the

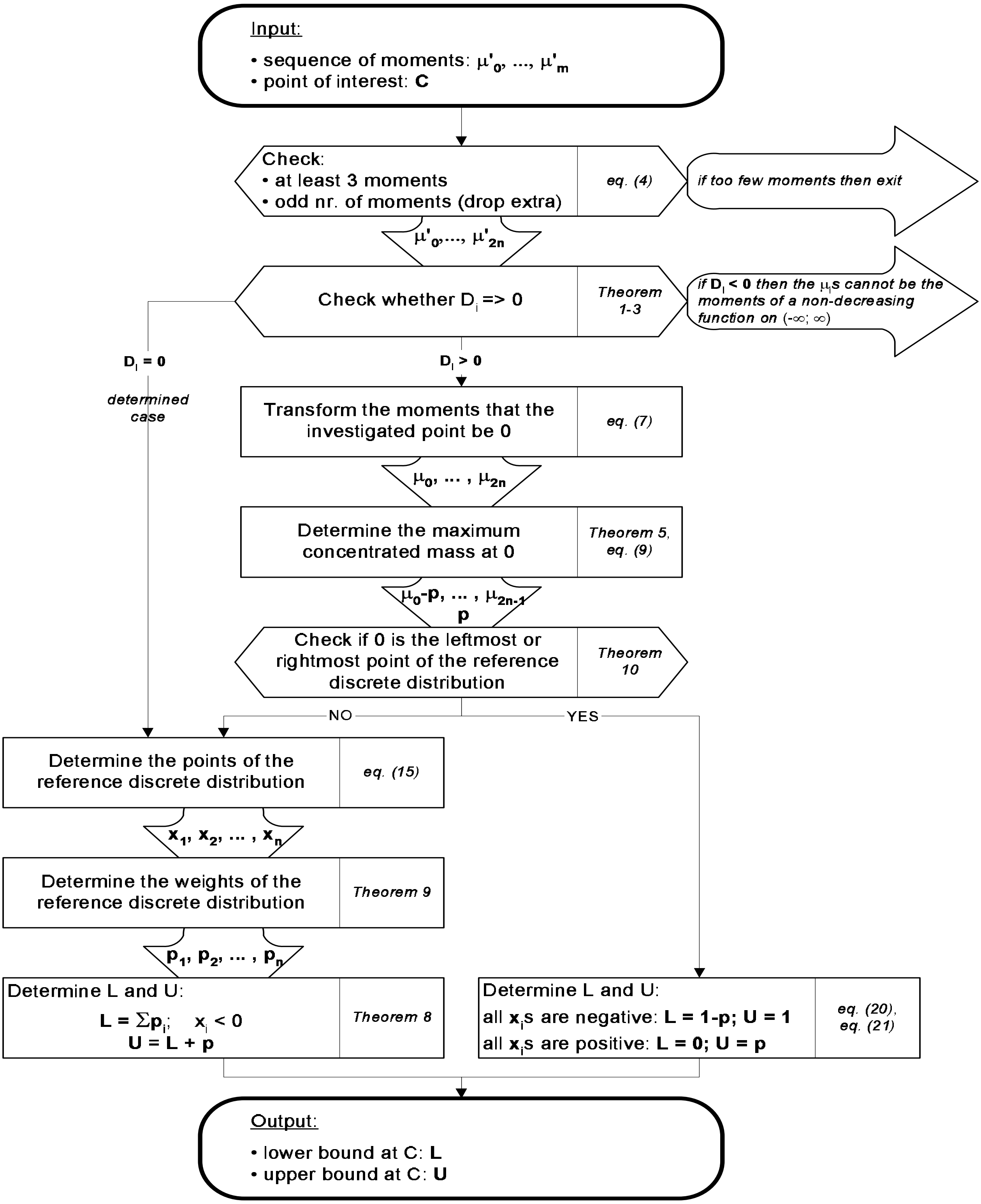


Figure 1: The block structure of the numerical procedure.

distribution) and the other one (x_1) is calculated based on (15) together with the associated probability masses (p and p_1, respectively).

$$p = \frac{\mu_2 - \mu_1^2}{\mu_2}, \tag{23}$$

$$x_1 = \frac{\mu_2}{\mu_1}, \qquad p_1 = \frac{\mu_1^2}{\mu_2}. \tag{24}$$

Note that, $\mu_2 > 0$ because $\mu_2 - \mu_1^2 > 0$, thus the sign of x_1 is the same as the sign of μ_1. A degenerate case arises when $\mu_1 = 0$. In this case $p = 1$ and x_1 becomes irrelevant since the associated mass is $p_1 = 0$.

Finally the lower and upper bounds are:

$$\boxed{\begin{array}{lll} L = \frac{\mu_2-\mu_1^2}{\mu_2}, & U = 1, & \text{if } \mu_1 < 0, \\ L = 0, & U = 1, & \text{if } \mu_1 = 0, \\ L = 0, & U = \frac{\mu_2-\mu_1^2}{\mu_2}, & \text{if } \mu_1 > 0. \end{array}} \tag{25}$$

4.2. Estimation based on 5 moments

Having the moments: $\mu_0 = 1, \mu_1, \mu_2, \mu_3, \mu_4$, we similarly discuss the two meaningful cases.

Case 1) determined sequence of moments: In this case $D_1 > 0$ and $D_2 = 0$, that is

$$D_1 = \begin{vmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{vmatrix} > 0, \quad \text{and} \quad D_2 = \begin{vmatrix} 1 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{vmatrix} = 0. \tag{26}$$

According to Theorem 3, this determined sequence of moments defines a discrete distribution with 2 points. The points and weights can be calculated based on (15). Let $r = \sqrt{-3\mu_1^2\mu_2^2 + 4\mu_2^3 + 4\mu_1^3\mu_3 - 6\mu_1\mu_2\mu_3 + \mu_3^2}$.

$$x_1 = \frac{\mu_1\mu_2 - \mu_3 + r}{2\mu_1^2 - 2\mu_2}, \tag{27}$$

$$p_1 = \frac{1}{2} + \frac{2\mu_1^3 - 3\mu_1\mu_2 + \mu_3}{2r}, \tag{28}$$

$$x_2 = \frac{\mu_1\mu_2 - \mu_3 - r}{2\mu_1^2 - 2\mu_2}, \tag{29}$$

$$p_2 = \frac{1}{2} - \frac{2\mu_1^3 - 3\mu_1\mu_2 + \mu_3}{2r}. \tag{30}$$

Case 2) positive sequence of moments: In this case $D_1 > 0$ and $D_2 > 0$. Applying the procedure of Figure 1 results in the following parameters.

$$p = \frac{-\mu_2^3 + 2\mu_1\mu_2\mu_3 - \mu_3^2 - \mu_1^2\mu_4 + \mu_2\mu_4}{\mu_2\mu_4 - \mu_3^2}, \tag{31}$$

$$x_1 = \frac{\mu_2\mu_3 - \mu_1\mu_4 - q}{2(\mu_2^2 - \mu_1\mu_3)}, \tag{32}$$

$$p_1 = \frac{-{\mu_2}^4\mu_3 + 2{\mu_1}^2{\mu_3}^3 + 3\mu_1{\mu_2}^3\mu_4 - 5\mu_1^2\mu_2\mu_3\mu_4 + {\mu_1}^3{\mu_4}^2 - q(\mu_2^3 - 2\mu_1\mu_2\mu_3 + \mu_1^2\mu_4)}{2q(\mu_3^2 - \mu_2\mu_4)}, \tag{33}$$

$$x_2 = \frac{\mu_2\mu_3 - \mu_1\mu_4 + q}{2(\mu_2^2 - \mu_1\mu_3)}, \tag{34}$$

$$p_2 = -\frac{\mu_2^2 - \mu_1\mu_3}{q}\left(-\mu_1 - \frac{(\mu_2^3 - 2\mu_1\mu_2\mu_3 + \mu_1^2\mu_4)(-\mu_2\mu_3 + \mu_1\mu_4 + q)}{2(\mu_2^2 - \mu_1\mu_3)(-\mu_3^2 + \mu_2\mu_4)}\right), \tag{35}$$

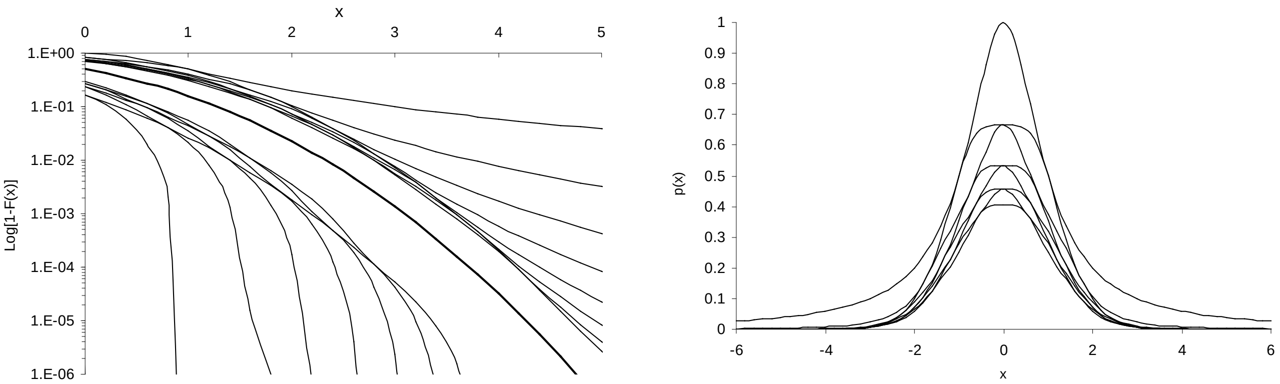
where $q = \sqrt{(-\mu_2\mu_3 + \mu_1\mu_4)^2 - 4(\mu_2^2 - \mu_1\mu_3)(\mu_3^2 - \mu_2\mu_4)}$. Note that, $p + p_1 + p_2 = 1$.

With 5 moments, it is hard to formulate the bounds conditioned directly on the moments, but using the calculated points and weights the bounds are:

$$\boxed{\begin{array}{lll} L = p_1 + p_2, & U = 1, & \text{if } x_1 < 0, x_2 < 0, \\ L = p_1, & U = p_1 + p, & \text{if } x_1 < 0, x_2 > 0, \\ L = 0, & U = p, & \text{if } x_1 > 0, x_2 > 0. \end{array}} \tag{36}$$

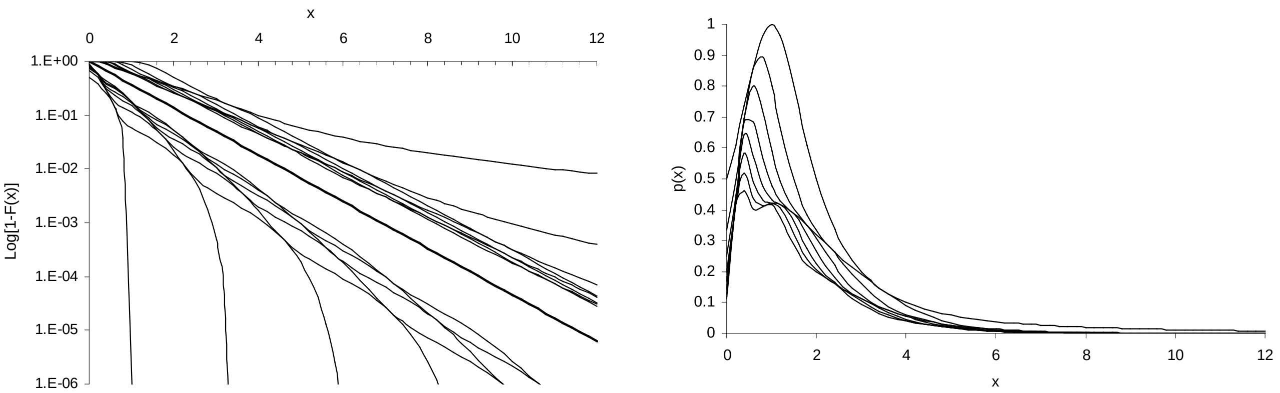
5. Numerical results

The application of the distribution estimation method for the analysis of large Markov reward models (MRMs) is demonstrated in [3]. In those examples the exact distributions are not known, hence the bounds are not related with any reference distribution. In this section we apply the distribution estimation method for known distributions which allows us to relate the moments based bounds with the exact values of a given distribution function.



a) Bounds of the ccdf. b) Distance of upper and lower bounds (p).

Figure 2: Bounding the standard normal distribution based on $3, 5, \ldots, 17$ moments.

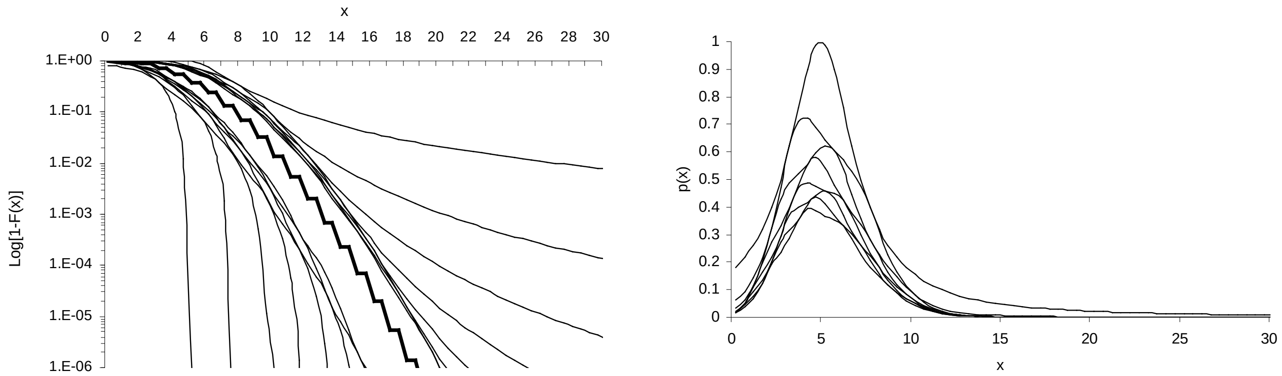


a) Bounds of the ccdf. b) Distance of upper and lower bounds (p).

Figure 3: Bounding the exponential distribution with mean 1 based on $3, 5, \ldots, 17$ moments.

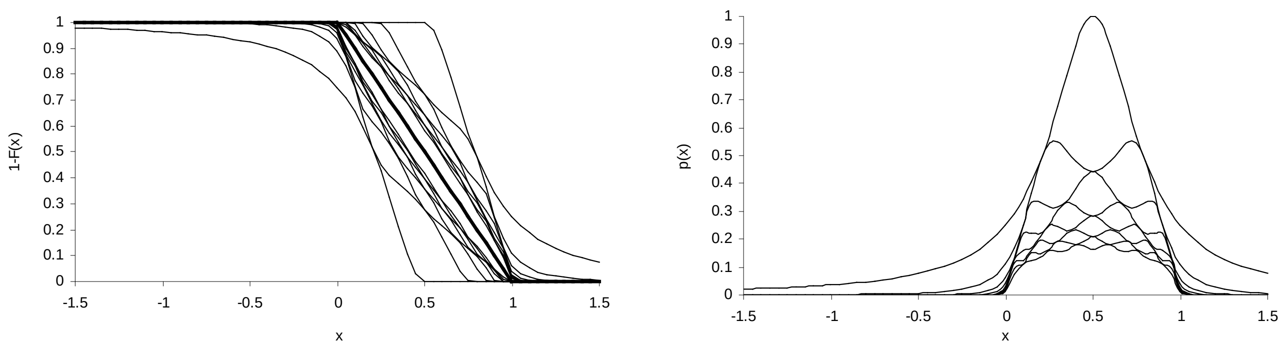
We chose 5 distributions, the standard normal ($\mu_i = 0 (i = 1, 3, \ldots)$, $\mu_i = \prod_{k=1}^{i/2} 2k - 1 (i = 0, 2, \ldots)$), the exponential ($\mu_i = i!$), the Poisson ($\mu_i = \sum_{k=0}^{\infty} k^i \lambda^k e^{-\lambda}/k!, \lambda = 5$), the $(0, 1)$ rectengular ($\mu_i = 1/(i + 1)$) and the chi-squared distribution ($\mu_i = \prod_{k=1}^{i} r + 2k - 2, r = 5$), and estimate them at several points based on $3, 5, \ldots, 17$ moments. Figures 2a, 3a, 4a and 6a present the complementary cummulative density functions (ccdf) and the distribution bounds using logarithmic scale. While the upper and lower bounds should tend to the same limit (0, in case of ccdf) as $x \to \infty$, the figures indicate a “visually” non-decreasing region between the bounds. This misleading pictures mean that the difference of the bounds (the error of the estimation) decreases exponentially. In the figures, we applied logarithmic scale to emphasize another important property of the estimation. The relative error of the estimation remains more or less constant in a wide range, where the width of the range depends on the number of moments. I.e., at the extreme values of the distribution increasing the number of moments does not improve the bounds significantly, but extends the range where the bounds maintain the given level of relative error.

Figure 5a depicts the ccdf of the rectangular distribution and its bounds with linear scale. This figure demonstrates the ability of our (Hamburger moment problem based) estimation in bounding distributions with finite support. The bounds based on more than 3 moments vanishes quickly at the limits of the distribution.



a) Bounds of the ccdf. b) Distance of upper and lower bounds (p).

Figure 4: Bounding the Poisson distribution with mean 5 based on $3, 5, \ldots, 17$ moments.

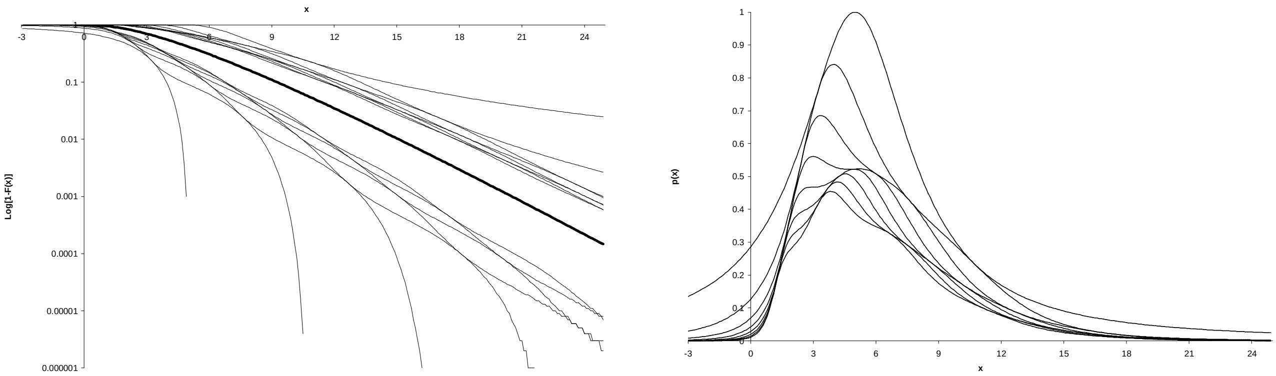


a) Bounds of the ccdf. b) Distance of upper and lower bounds (p).

Figure 5: Bounding the $(0, 1)$ continuous rectangular distribution based on $3, 5, \ldots, 17$ moments.

The lower bound of the ccdf reaches 0 at the point where all roots of (15) (with the moments of the shifted distribution) are negative. That is where the lower bounds break down in the figures with logarithmic scale. Beyond this limit the upper bound of the ccdf is p. As it is visible in the figures with logarithmic scale the upper bounds tend to an exponentially decreasing asymptotic limit which is a function of the number of moments. Indeed, the asymptotic limit is x^{-2n} as it is reported in [15].

Figures 2b, 3b, 4b, 5b and 6b depict the difference of the bounds (i.e., the maximal mass p) as a function of x. Every $p(x)$ function has a peak around the mean of the distribution



a) Bounds of the ccdf. b) Distance of upper and lower bounds (p).

Figure 6: Bounding the chi-squared distribution based on $3, 5, \ldots, 17$ moments.

and decreases towards the extreme values in both directions. The figures obviously verify the middle line of (25) since the 3-moment curves reach 1 at the mean in all cases.

All figures indicate points where the bounds or the $p(x)$ curves calculated based on $2n-1$ and $2n+1$ moments coincide. Indeed, the number of coinciding points is n. The reason of this property is associated with the layout of the reference discrete distribution with respect to the point of interest. When we bound extreme low (high) values of a distribution all roots of the reference distributions are on the right (left) of the point of interest. Between the extreme values there is a continuous transition of roots from one side to the other of the point of interest. The bounds based on $2n+1$ moments are calculated from a discrete distribution of n points (different from the point of interest). During the continuous transition of the points of the discrete distribution from one side to the other of the reference point there are n cases when a root coincide with the point of interest. In these cases the last two moments do not contain additional information for bounding the distribution and so the bound curves and the $p(x)$ curves coincide. Figures 2b, 3b, 4b and 5b also indicate that the maximal mass at the point of interest is a decreasing function of the number of moments and equality can occur only in the mentioned extreme points.

A crucial issue of the proposed distribution estimation method is the numerical stability. It is important to emphasize that all steps of the numerical procedure depends essentially on the accuracy of the calculation. For example, the standard floating point precision of Mathematica provides negative D_{17} for the positive sequence $\mu_i = i!$ (which are the moments of the exponential distribution) due to floating point errors. This example indicates that even the simplest step of the procedure can fail with “theoretically” positive series of moments. Unfortunately, we do not always have “theoretically correct” moment sequence. If the moments are calculated by other complex computational methods like the ones for the moment analysis of large MRMs [1] the resulted moments can accumulate the numerical errors of the preceding calculations. In these cases it is always recommended to check if the obtained moment sequence is positive in advance of the calculation.

Figure 7 demonstrates the effect of distorted moments on the distribution bounds when the chi-squared distribution is approximated based on 5 moments (solid lines). It is interesting to note, that the odd and even moments have different effect on the distorted bounds (dashed lines). An increased odd moment results tighter bounds and an increased even moment results looser bounds (and vice versa), when the other moments are kept fix.

The complexity as well as the accuracy of calculating p is practically identical with the calculation of the Hankel determinant. This fact suggests an easy practical approach to check the validity of p. We consider p to be valid as long as both, the numerator and the denominator of (9) are positive.

Finding the roots of the order n polynomial (Theorem 9) and the solution of the linear system (16) is more complex and more sensitive to numerical errors then calculating p. The majority of standard numerical packages can indicate ill conditioning or near-singularity which warn the presence of numerical problems. According to Theorem 3 the $P_n(x)$ polynomial has n real roots. The presence of complex roots also suggests numerical problems. According to our experience in the majority of the cases our numerical procedure provided trustworthy results (which follow the general trends of stable results) even in ill conditioned cases, but we can not evaluate the numerical error of these cases. We never experienced a case with incorrect bound in the sense that the calculated lower (upper) bound was higher (lower) than the exact value of the distribution at the given point. In case of serious numerical problems the number of moments used for the estimation has to be reduced by two.

The most stable distribution bounds can be obtained for extreme values, when the Liènard–Chipart criterion is fulfilled (Theorem 10). The complexity and the precision of calculating the T_i determinants is similar to the one of calculating p and the search of the roots of $P_n(x)$ and the solution of the linear system is skipped in this case. Hence, bounds provided by the numerical method are most trustworthy in these cases and the only possible check of precision is the positivity of the Hankel determinants.

Appendix

A. Proof of Theorem 5

To prove the theorem we need the following lemma:

Lemma 11 *[16]* ***(Jacobi)*** *Let A_{ij} denote the order $n-1$ minors of the $n \times n$ quadratic matrix A. Let $A\begin{pmatrix} i_1 \dots i_p \\ i_1 \dots i_p \end{pmatrix}$ denote the order $n-p$ minors of A, which can be obtained by deleting the rows and columns $i_1, i_2, \dots, i_p$. For these quantities the following equation holds:*

$$\begin{vmatrix} A_{i_1 i_1} & A_{i_1 i_2} & \dots & A_{i_1 i_p} \\ A_{i_2 i_1} & A_{i_2 i_2} & \dots & A_{i_1 i_p} \\ \vdots & \vdots & \ddots & \vdots \\ A_{i_n i_1} & A_{i_n i_2} & \dots & A_{i_p i_p} \end{vmatrix} = |A|^{p-1} A \begin{pmatrix} i_1 \dots i_p \\ i_1 \dots i_p \end{pmatrix} . \tag{37}$$

Proof: We prove the theorem by induction. Let us substitute the definitions of $\rho_n(z)$ and p into (10):

$$\frac{\begin{vmatrix} \mu_0 & \mu_1 & \dots & \mu_n \\ \mu_1 & \mu_2 & \dots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_{n+1} & \dots & \mu_{2n} \end{vmatrix}}{\begin{vmatrix} \mu_2 & \mu_3 & \dots & \mu_{n+1} \\ \mu_3 & \mu_4 & \dots & \mu_{n+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n+1} & \mu_{n+2} & \dots & \mu_{2n} \end{vmatrix}} = \frac{1}{\sum_{k=0}^{n} |P_k(0)|^2} \tag{38}$$

1. For $n = 1$:
   The left hand side is

$$\frac{\begin{vmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{vmatrix}}{\mu_2} = \frac{\mu_2 - \mu_1^2}{\mu_2} \ , \tag{39}$$

   and the right hand side is

$$\begin{aligned} \rho_n(0) &= \frac{1}{1 + |P_n(0)|^2} = \frac{1}{1 + \left| \frac{1}{\sqrt{\begin{vmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{vmatrix} |1|}} \begin{vmatrix} 1 & \mu_1 \\ 1 & 0 \end{vmatrix} \right|^2} = \\ &= \frac{1}{1 + \left| \frac{1}{\sqrt{\mu_2 - \mu_1^2}} (-\mu_1) \right|^2} = \frac{1}{1 + \frac{\mu_1^2}{\mu_1 \mu_2 - \mu_1^2}} = \\ &= \frac{1}{\frac{\mu_2 - \mu_1^2 + \mu_1^2}{\mu_2 - \mu_1^2}} = \frac{\mu_2 - \mu_1^2}{\mu_2} \ . \end{aligned} \tag{40}$$

2. Let us assume that (38) is true for an arbitrary $n = k$ that is

$$\frac{\begin{vmatrix} \mu_0 & \mu_1 & \dots & \mu_k \\ \mu_1 & \mu_2 & \dots & \mu_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_k & \mu_{k+1} & \dots & \mu_{2k} \end{vmatrix}}{\begin{vmatrix} \mu_2 & \mu_3 & \dots & \mu_{k+1} \\ \mu_3 & \mu_4 & \dots & \mu_{k+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k+1} & \mu_{k+2} & \dots & \mu_{2k} \end{vmatrix}} = \frac{1}{1 + (P_1(0))^2 + \ldots + (P_k(0))^2} \ . \tag{41}$$

   We can write $(P_i(0))^2$ instead of $|P_i(0)|^2$ because it can be seen from (5) that $P_i(x)$ is real for every real x.

3. Now, we show that (38) is true for $n = k + 1$. The reciprocal of (38) is

$$\frac{\begin{vmatrix} \mu_2 & \mu_3 & \dots & \mu_{k+2} \\ \mu_3 & \mu_4 & \dots & \mu_{k+3} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k+2} & \mu_{k+3} & \dots & \mu_{2k+2} \end{vmatrix}}{\begin{vmatrix} \mu_0 & \mu_1 & \dots & \mu_{k+1} \\ \mu_1 & \mu_2 & \dots & \mu_{k+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k+1} & \mu_{k+2} & \dots & \mu_{2k+2} \end{vmatrix}} = 1 + (P_1(0))^2 + \ldots + (P_{k+1}(0))^2 \ . \tag{42}$$

Transforming the right hand side of the equation above we get:

$$
\begin{aligned}
1 \quad &+(P_1(0))^2+\ldots+(P_k(0))^2+(P_{k+1}(0))^2 = \\
= \quad &1+(P_1(0))^2+\ldots+(P_k(0))^2+\frac{1}{D_{k+1}D_k}\begin{vmatrix} \mu_1 & \mu_2 & \ldots & \mu_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_k & \mu_{k+1} & \ldots & \mu_{2k+1} \\ 1 & 0 & \ldots & 0 \end{vmatrix}^2 = \\
= \quad &1+\ldots+(P_k(0))^2+\frac{(-1)^{2k+2}\begin{vmatrix} \mu_1 & \ldots & \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} & \ldots & \mu_{2k+1} \end{vmatrix}^2}{\begin{vmatrix} 1 & \ldots & \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} & \ldots & \mu_{2k+2} \end{vmatrix}\begin{vmatrix} 1 & \ldots & \mu_k \\ \vdots & \ddots & \vdots \\ \mu_k & \ldots & \mu_{2k} \end{vmatrix}}\,. \qquad (43)
\end{aligned}
$$

We can write $\frac{1}{D_{k+1}D_k}$ instead of $\frac{1}{|D_{k+1}D_k|}$, because $\mu_0, \mu_1, \ldots, \mu_{2n}$ is a positive sequence and so $D_i > 0$ for all $i = 1, \ldots, n$. On the basis of (41) we can substitute $1+\ldots+(P_k(0))^2$ and this equals to the left hand side of (42):

$$
\frac{\begin{vmatrix} \mu_2 & \ldots & \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} & \ldots & \mu_{2k} \end{vmatrix}}{\begin{vmatrix} 1 & \ldots & \mu_k \\ \vdots & \ddots & \vdots \\ \mu_k & \ldots & \mu_{2k} \end{vmatrix}} + \frac{\begin{vmatrix} \mu_1 & \ldots & \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} & \ldots & \mu_{2k+1} \end{vmatrix}^2}{\begin{vmatrix} 1 & \ldots & \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} & \ldots & \mu_{2k+2} \end{vmatrix}\begin{vmatrix} 1 & \ldots & \mu_k \\ \vdots & \ddots & \vdots \\ \mu_k & \ldots & \mu_{2k} \end{vmatrix}} = \frac{\begin{vmatrix} \mu_2 & \ldots & \mu_{k+2} \\ \vdots & \ddots & \vdots \\ \mu_{k+2} & \ldots & \mu_{2k+2} \end{vmatrix}}{\begin{vmatrix} 1 & \ldots & \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} & \ldots & \mu_{2k+2} \end{vmatrix}}\,. \qquad (44)
$$

Hence, we have to prove that:

$$
\begin{vmatrix} \mu_2 & \ldots & \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} & \ldots & \mu_{2k} \end{vmatrix}\begin{vmatrix} 1 & \ldots & \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} & \ldots & \mu_{2k+2} \end{vmatrix} + \begin{vmatrix} \mu_1 & \ldots & \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} & \ldots & \mu_{2k+1} \end{vmatrix}^2 = \begin{vmatrix} \mu_2 & \ldots & \mu_{k+2} \\ \vdots & \ddots & \vdots \\ \mu_{k+2} & \ldots & \mu_{2k+2} \end{vmatrix}\begin{vmatrix} 1 & \ldots & \mu_k \\ \vdots & \ddots & \vdots \\ \mu_k & \ldots & \mu_{2k} \end{vmatrix}\,. \qquad (45)
$$

Let $p = 2$, $i_1 = 1$, $i_2 = k+1$ and $M = D_{k+1}$. By Lemma 11 it can be seen using (37) that

$$
\begin{vmatrix} M_{1,1} & M_{1,k+1} \\ M_{k+1,1} & M_{k+1,k+1} \end{vmatrix} = |M|\, M\begin{pmatrix} 1 & k+1 \\ 1 & k+1 \end{pmatrix}. \qquad (46)
$$

Expanding the determinant on the left hand side:

$$
M_{1,1}\, M_{k+1,k+1} - M_{k+1,1}\, M_{1,k+1} = |M|\; M\begin{pmatrix} 1 & k+1 \\ 1 & k+1 \end{pmatrix}, \qquad (47)
$$

and $M_{1,k+1} = M_{k+1,1}$ because M is a symmetric matrix. Rearranging the equation we get:

$$|M|\; M\begin{pmatrix} 1 & k+1 \\ 1 & k+1 \end{pmatrix} + (M_{k+1,1})^2 = M_{1,1}\, M_{k+1,k+1}. \tag{48}$$

Using the definitions of M_{ii} and $M\begin{pmatrix} 1 & k+1 \\ 1 & k+1 \end{pmatrix}$ equation (48) can be written as:

$$\begin{vmatrix} \mu_2 & \dots \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} \dots & \mu_{2k} \end{vmatrix} \begin{vmatrix} 1 & \dots \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} \dots & \mu_{2k+2} \end{vmatrix} + \begin{vmatrix} \mu_1 & \dots \mu_{k+1} \\ \vdots & \ddots & \vdots \\ \mu_{k+1} \dots & \mu_{2k+1} \end{vmatrix}^2 = \begin{vmatrix} \mu_2 & \dots \mu_{k+2} \\ \vdots & \ddots & \vdots \\ \mu_{k+2} \dots & \mu_{2k+2} \end{vmatrix} \begin{vmatrix} 1 & \dots \mu_{k} \\ \vdots & \ddots & \vdots \\ \mu_{k} \dots & \mu_{2k} \end{vmatrix}, \tag{49}$$

which is identical with (45) that had to be proven. ■

B. Proof of Theorem 8

Using the $\sigma_1(x) := \sigma^*(x)$ and the $\sigma_2(x) := \sigma(x)$ substitutions by Theorem 7 we have:

$$\left| \int_{-\infty}^{0^+} d\sigma^*(u) - \int_{-\infty}^{0^-} d\sigma(u) \right| \le p\ . \tag{50}$$

Resolving the absolute value sign it is:

$$\int_{-\infty}^{0^+} d\sigma^*(u) - p \le \int_{-\infty}^{0^-} d\sigma(u) \le \int_{-\infty}^{0^+} d\sigma^*(u) + p. \tag{51}$$

By the construction of $\sigma^*(x)$ it follows that

$$\int_{-\infty}^{0^+} d\sigma^*(u) = \int_{-\infty}^{0^-} d\sigma^*(u) + p\ . \tag{52}$$

Substituting it to the leftmost inequality we get:

$$\int_{-\infty}^{0^-} d\sigma^*(u) \le \int_{-\infty}^{0^-} d\sigma(u)\ , \tag{53}$$

which is (13). Now, using the $\sigma_1(x) := \sigma(x)$ and the $\sigma_2(x) := \sigma^*(x)$ substitutions Theorem 7 gives

$$\left| \int_{-\infty}^{0^+} d\sigma(u) - \int_{-\infty}^{0^-} d\sigma^*(u) \right| \le p. \tag{54}$$

Resolving the absolute value sign we have:

$$\int_{-\infty}^{0^-} d\sigma^*(u) - p \le \int_{-\infty}^{0^+} d\sigma(u) \le \int_{-\infty}^{0^-} d\sigma^*(u) + p, \tag{55}$$

whose rightmost inequality gives (14). ■

References

[1] M. Telek and S. Rácz. Numerical analysis of large Markovian reward models. *Performance Evaluation*, 36&37:95–114, Aug 1999.

[2] S. Rácz. *Numerical analysis of communication systems through Markov reward models.* PhD thesis, Technical University of Budapest, 2000.

[3] S. Rácz, Á. Tari, and M. Telek. MRMSolve: Distribution estimation of large Markov reward models. In *Tools 2002*, pages 72–81, London, England, April 2002. Springer, LNCS 2324.

[4] J. A. Shohat and D. J. Tamarkin. *The problem of moments.* Americal Mathematical Society, Providence, Rhode Island, 1946. Mathematical surveys.

[5] M. G. Krein and A. A. Nudelman. *The Markov moment problem and extremal problems*, volume 50. Americal Mathematical Society, Providence, Rhode Island, 1977. Translation of mathematical monographs.

[6] N. I. Akhiezer. *The classical moment problem and some related questions in analysis.* Hafner publishing company, New York, 1965. (translation of Н. И. Ахиезер: Классическая Проблема Моментов и Некоторые Вопросы Анализа, published by Государственное Издательство Физико-Математической Литературы, Moscow, 1961).

[7] M. G. Kendall and A. Stuart. *The advanced theory of statistics – Distribution theory*, volume 1. Hafner publishing company, New York, 3rd edition, 1969.

[8] T. Stieltjes. Reserches sur les fractions continues. *Ann. Fac. Sci. Univ. Toulouse*, 2:1–122, 1894.

[9] H. Hamburger. Ueber eine erweiterung des Stieltjes'schen momentproblems. *Mathematische Annalen*, 81:235–319, 1920.

[10] F. Hausdorff. Momentprobleme für ein endliches intervall. *Mathematische Zeitschrift*, 16:220–248, 1923.

[11] Aldo Tagliani. Existence and stability of a discrete probability distribution by maximum entropy approach. *Applied Mathematics and Computation*, 110:105–114, 2000.

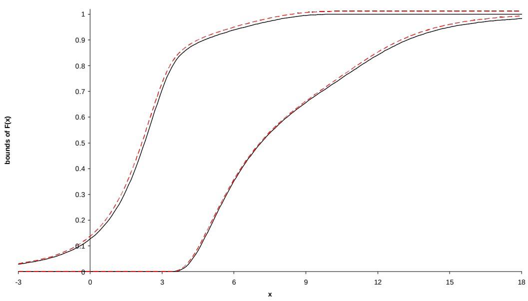
[12] M. Reisz. Sur le probleme des moments et le th'eoreme de parseval correspondant. *Acta Litt. Sci.*, 1:209–225, 1922.

[13] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1993. http://lib-www.lanl.gov/numerical/bookcpdf.html.

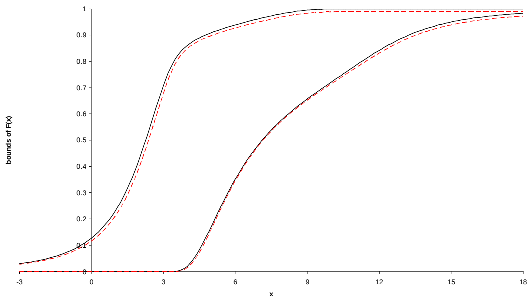
[14] G. A. Korn and T. M. Korn. *Mathematical handbook for scientists and engineers.* Dover, 2000.

[15] B. G. Lindsay and P. Bassak. Moments determine the tail of a distribution (but not much else). *The American Statistician*, 54(4):248–251, 2000.

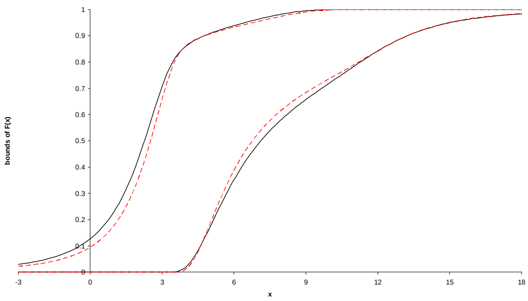
[16] P. Rózsa. *Lineális algebra és alkalmazásai.* Tankönyvkiadó, 1991. in Hungarian.



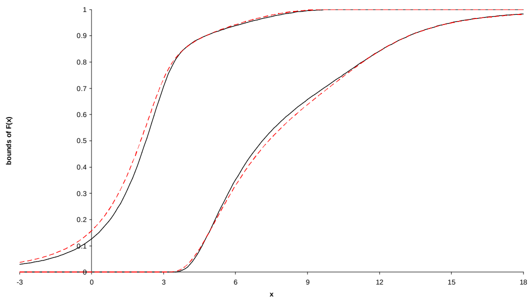
$\mu_0 + 1.1\%$ ($\mu'_0 = 1.011\mu_0$).



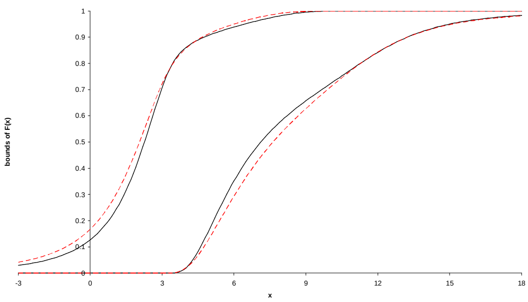
$\mu_0 - 1.1\%$ ($\mu'_0 = 0.989\mu_0$).



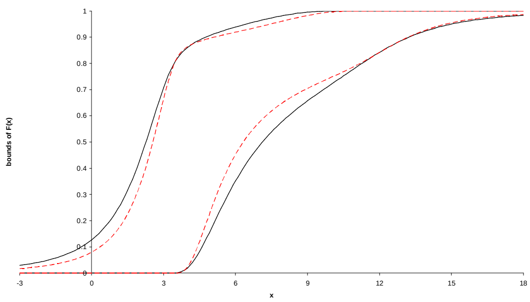
$\mu_1 + 1.1\%$ ($\mu'_1 = 1.011\mu_1$).



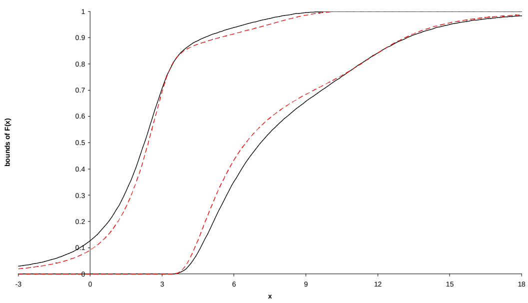
$\mu_1 - 1.1\%$ ($\mu'_1 = 0.989\mu_1$).



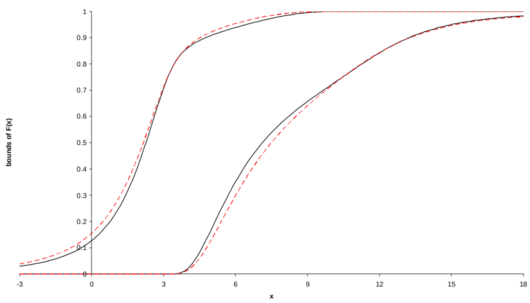
$\mu_2 + 1.1\%$ ($\mu'_2 = 1.011\mu_2$).



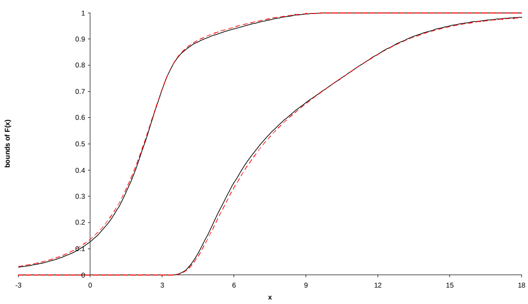
$\mu_2 - 1.1\%$ ($\mu'_2 = 0.989\mu_2$).



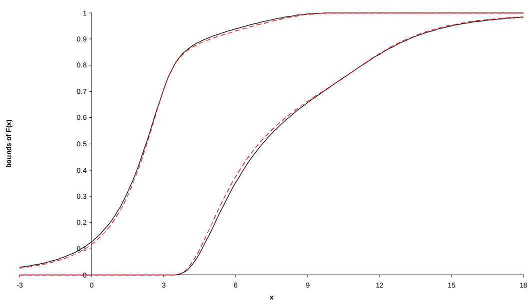
$\mu_3 + 1.1\%$ ($\mu'_3 = 1.011\mu_3$).



$\mu_3 - 1.1\%$ ($\mu'_3 = 0.989\mu_3$).



$\mu_4 + 1.1\%$ ($\mu'_4 = 1.011\mu_4$).



$\mu_4 - 1.1\%$ ($\mu'_4 = 0.989\mu_4$).

Figure 7: Effect of distorted moments on the 5-moment bounds of the chi-squared distribution.