# A unified approach to the moments based distribution estimation – Unbounded support

Árpád Tari[1], Miklós Telek[2], and Peter Buchholz[1]

[1] Universität Dortmund, Germany
[2] Budapest University of Technology and Economics, Hungary
arpad@sch.bme.hu, telek@hit.bme.hu, peter.buchholz@cs.uni-dortmund.de

**Abstract.** The problem of moments has been studied for more than a century. This paper discusses a practical issue related to the problem of moments namely the bounding of a distribution based on a given number of moments. The presented approach is unified in the sense that all measures of interests are provided as a quadratic expression of the same Hankel-matrix.

Application examples indicate the importance of the presented approach.

**Keywords**: reduced moment problem, moments based distribution bounding, Hankel matrix.

## 1 Introduction

The aim of making stochastic models of real systems is usually to evaluate some performance parameters of the system. Since the parameters of interest are random variables in stochastic models the goal of the analysis is often to characterize these random variables by their distribution. Real-life problems often require the solution of huge models and/or the evaluation of time demanding numerical procedures. One way of avoiding these problems is to introduce certain simplifications that result in the reduction of the complexity, another way is to approximate the measure. The approach proposed in this paper is along the second line. Instead of calculating the distribution of the measure of interest we calculate some of its first moments and approximate the distribution based on them as it is depicted in Figure 1. The thickness of the arrows indicates the usual complexity of the algorithms available in the literature for the different type calculations. Having a complex stochastic model it is usually faster to calculate the moments of the measure of interest than its distribution and the calculation of an approximation or bounds of the distribution from the moments can be done very efficiently.

The drawback of this approach is that the computation yields only an approximation of the distribution, not its exact values. Fortunately, one can have an idea on the error of this approach because it is also possible to bound the distribution based on known moments. As a consequence, it becomes possible to bound performance measures of real systems where moments are also computable, but a direct analysis of the distribution is infeasible.
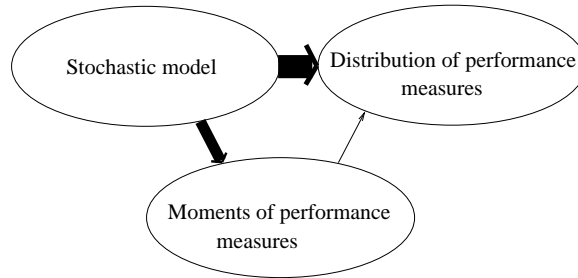
**Fig. 1.** Using moments to estimate performance measures

The task of moments-based distribution estimation can be formulated as follows. We have an unknown distribution function $\sigma(x)$ with support on $[a, b]$ and we know the following quantities:

$$\mu_i = \int_a^b x^i \, d\sigma(x), \quad i = 0, 1, 2, \ldots, 2n, \tag{1}$$

where $\mu_i$ is called the $i^{th}$ moment of the distribution $\sigma(x)$. We look for a procedure to estimate $\sigma(x)$ at some point $x = C$.

Many algorithms exist that are able to fit certain types of distributions to some given moments. For example, it is usually obvious to fit distributions with one parameter to $\mu_1$ and with two parameters to $\mu_1$ and $\mu_2$. Fitting of $\mu_1 - \mu_3$ with acyclic phase type distribution is presented in [1]. Appie van de Liefvoort provided a method to fit arbitrary number of moments with matrix exponential distribution in [2]. Any of these methods provides a particular $\tilde{\sigma}(C)$ value assuming the considered class of distributions. Unfortunately, the error of this estimate it is not known if the performance measure does not belong to the considered class of distributions.

To overcome this difficulty we look for minimal and maximal estimates of $\sigma(C)$ over the class of all valid distribution functions. In this way we define a lower and an upper limit, which bound all distribution functions in point $C$ having moments $\mu_i, (i = 0, 1, 2, \ldots, 2n)$. No distribution function has smaller and no has larger value at $C$, than the limits we present. The bounds are strict in the sense that there exist always a distribution that reaches these values.

The problem to determine the distribution based on its moments is called "*the reduced moment problem*" (where the word reduced means that only a finite number of moments is known). The term was introduced by Stieltjes who did the first extensive study on the subject in [3], though Chebyshev solved a particular case of the problem as early as 1873. Markov, Hamburger, Hausdorff, Nevanlinna, M. Riesz, Carleman and Stone wrote the most important articles about the moment problem. One can find a good historical overview in [4]. The case when the limits of the integral in (1) are $(-\infty, \infty)$ is referred to as Hamburger moment problem. It means that the considered set of distributions has a support in the whole $(-\infty, \infty)$ interval (*unbounded support*), which is the most general case.

The mathematical tools involved in the solution of this problem were continued fractions, approximate quadratures of integrals, singular integral equations, orthogonal polynomials and operators in Hilbert space [4]. Our approach is based mainly on matrix operations, though we also rely on the theory of orthogonal polynomials.

The method presented here is basically the same as the one in [5], but with major improvements: we eliminated the moments transformation step of that method and deduced simpler formulas which resulted in notable simplification of the algorithm. In contrast with [5] here we present a unified approach where all related quantities are expressed with quadratic matrix expressions of the same matrix ($\boldsymbol{M}^{-1}$).

We intend to present this paper in a way that is simple and easy to understand and does not require any special mathematical knowledge from the reader and we also provide references to the related mathematical background. Furthermore we present details about our implementation of the procedure, which is not always obvious due to potential numerical problems.

The paper is organized as follows: Section 2 defines the conditions that have to be fulfilled in order to use our estimation algorithm. Sections 3 and 4 summarize the theoretical background and the bounding procedure of the moment based estimation. In Section 5 a step-by-step instruction to evaluate the algorithm can be found. Section 6 provides an overview of the applicability of the proposed approach and gives insights into the numerical issues. An example is analyzed in Section 7 focused on the strengths and applicabilities of our approach. Section 8 concludes the paper.

## 2 Necessary conditions

Before calculating an estimate of $\sigma(\cdot)$ based on $\mu_i, (i = 0, 1, 2, \ldots, 2n)$ we need to check if a non-decreasing function exists whose moments are $\mu_i, (i = 0, 1, 2, \ldots, 2n)$. This can be checked using the theorem of Hausdorff [6, p. 30]:

**Theorem 1.** *[6] Let $\mu_0, \mu_1, \mu_2, \ldots, \mu_{2n}$ be a sequence of real numbers. These numbers can be a moments of a distribution function with support $(-\infty, \infty)$ if and only if*

$$|\boldsymbol{M}_k| \geq 0, \qquad k = 0, 1, \ldots, n, \tag{2}$$

*where*

$$\boldsymbol{M}_k = \begin{pmatrix} \mu_0 & \mu_1 & \ldots & \mu_k \\ \mu_1 & \mu_2 & \ldots & \mu_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_k & \mu_{k+1} & \ldots & \mu_{2k} \end{pmatrix}, \qquad k = 0, 1, \ldots, n \tag{3}$$

*is the so-called Hankel-matrix of dimension $(k+1) \times (k+1)$.*

**Definition 2.** *[7] The $n \times n$ real matrix $\boldsymbol{A}$ is called* positive definite, *if*

$$\boldsymbol{\xi}^T \boldsymbol{A} \boldsymbol{\xi} > 0 \tag{4}$$

*for any nonzero real vector $\boldsymbol{\xi} \in \mathbb{R}^n$. This is equivalent to the requirement that the determinants associated with all upper-left submatrices of $\boldsymbol{A}$ are positive.*

According to this definition the matrix $\boldsymbol{M}$ is positive definite. The following lemma states the same about its inverse.

**Lemma 3.** *[7] If $\boldsymbol{A}$ is a $n \times n$ positive definite matrix, then $\boldsymbol{A}^{-1}$ is also positive definite.*

It follows from Theorem 1 that we have constraints only on an odd number of moments (including $\mu_0$). Indeed, if $\mu_i, (i = 0, 1, 2, \ldots, 2n)$ is a valid moment sequence then $\mu_{2n+1}$ can take any value in $(-\infty, \infty)$ and $\mu_i, (i = 0, 1, 2, \ldots, 2n + 1)$ remains to be a valid moments sequence. As a consequence $\mu_{2n+1}$ does not carry any information about the possible limits of $\sigma(\cdot)$, hence we simply discard it if $\mu_{2n+2}$ is not known.

**Theorem 4.** *[8] $\sigma(\cdot)$ consists of exactly $n$ distinct points of increase if and only if*

$$|\boldsymbol{M}_0| > 0, |\boldsymbol{M}_1| > 0, \ldots, |\boldsymbol{M}_{n-1}| > 0, |\boldsymbol{M}_n| = 0. \tag{5}$$

*In this case $|\boldsymbol{M}_{n+1}| = |\boldsymbol{M}_{n+2}| = \ldots = 0$ and the moment problem is said to be determined.*

In the special case when the moment problem is determined there is exactly one discrete distribution (with $n$ points) with the given moments and upper and lower bounds are identical.

From now on we assume non-determined moment problem and denote by $\boldsymbol{M}$ the largest possible Hankel-matrix, that can be formed from the known moments: $\boldsymbol{M} = \boldsymbol{M}_n$.

## 3   Discrete reference distribution

We construct a discrete distribution from the given moments. This distribution has an interesting extremal property: among the distribution functions having the $\mu_i, \, i = 0, 1, \ldots, 2n$ moments it has the largest concentrated mass at point $C$.

The construction of this discrete distribution can be considered as the solution of a system of equations. We have the $\mu_i, \, i = 0, 1, \ldots, 2n$ moments and point $C$ and we search for a discrete distribution with moments $\mu_i, \, i = 0, 1, \ldots, 2n$ and a maximum mass at $C$. The following lemma gives some more information about this distribution.

**Lemma 5.** *[4, p. 42] A distribution function with exactly $n+1$ points of increase is uniquely determined by a single point of them. To each real value $C$ there corresponds one and only one distribution function with $n + 1$ or $n$ points of*

*increase, which contains $C$ among its points. The distribution function has $n+1$ points of increase, if $C$ is not a root of the following polynomial:*

$$P(x) = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n-1} & \mu_n & \cdots & \mu_{2n-1} \\ 1 & x & \cdots & x^n \end{vmatrix} \tag{6}$$

*and has $n$ points otherwise.*

If there is a discrete distribution supported exactly on $n+1$ points (including $C$), then this discrete distribution can be constructed from the definition of moments:

$$\mu_i = p\,C^i + \sum_{j=1}^{n} p_j\,x_j^i \ . \tag{7}$$

This task can also be formulated as finding the appropriate values of $x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_n, p$ so that

$$\boldsymbol{M} - \boldsymbol{R}\boldsymbol{S}^T = \boldsymbol{0}, \tag{8}$$

where

$$\boldsymbol{R} = \begin{pmatrix} p & p_1 & \cdots & p_n \\ pC & p_1 x_1 & \cdots & p_n x_n \\ pC^2 & p_1 x_1^2 & \cdots & p_n x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ pC^n & p_1 x_1^n & \cdots & p_n x_n^n \end{pmatrix} = (p\,\boldsymbol{c}, p_1\,\boldsymbol{x_1}, \ldots, p_n\,\boldsymbol{x_n}) \tag{9}$$

$$\boldsymbol{S} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ C & x_1 & \cdots & x_n \\ C^2 & x_1^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ C^n & x_1^n & \cdots & x_n^n \end{pmatrix} = (\boldsymbol{c}, \boldsymbol{x_1}, \ldots, \boldsymbol{x_n}) \tag{10}$$

and $\boldsymbol{x_i}$ and $\boldsymbol{c}$ are vectors of order $n$ formed by the powers of $x_i$ and $C$, respectively:

$$\boldsymbol{x_i}^T = \left(1, x_i, x_i^2, \ldots, x_i^n\right)^T, \qquad \boldsymbol{c}^T = \left(1, C, C^2, \ldots, C^n\right)^T, \tag{11}$$

$\boldsymbol{0}$ is an $(n+1) \times (n+1)$ zero matrix.

**Lemma 6.** *If $\boldsymbol{A}$ is a $n \times n$ nonsingular matrix, $\boldsymbol{U}$ and $\boldsymbol{V}$ are $n \times n$ nonsingular matrices such that*

$$\boldsymbol{A} - \boldsymbol{U}\boldsymbol{V}^T = \boldsymbol{0}, \tag{12}$$

*where $\boldsymbol{0}$ is a zero matrix of order $n$, then*

$$\boldsymbol{V}^T \boldsymbol{A}^{-1} \boldsymbol{U} = \boldsymbol{I}, \tag{13}$$

*where $\boldsymbol{I}$ is identity matrix of order $n$.*

*Proof.* Since $\boldsymbol{U}$ and $\boldsymbol{V}$ are of order $n$ and their inverses exist we have

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{V}^T,$$

$$\boldsymbol{U}^{-1}\boldsymbol{A}\left(\boldsymbol{V}^T\right)^{-1} = \boldsymbol{I},$$

$$\boldsymbol{V}^T\boldsymbol{A}^{-1}\boldsymbol{U} = \boldsymbol{I}^{-1} = \boldsymbol{I},$$

where we took the inverse of both sides in the last step. $\square$

According to Lemma 6 we can write:

$$\boldsymbol{I} = \boldsymbol{S}^T\boldsymbol{M}^{-1}\boldsymbol{R}$$

$$\boldsymbol{I} = \begin{pmatrix} \boldsymbol{c}^T \\ \boldsymbol{x_1}^T \\ \vdots \\ \boldsymbol{x_n}^T \end{pmatrix} \boldsymbol{M}^{-1} \ (p\,\boldsymbol{c}, p_1\,\boldsymbol{x_1}, \dots, p_n\,\boldsymbol{x_n})$$

$$\begin{pmatrix} 1\,0\dots0 \\ 0\,1\dots0 \\ \vdots\,\vdots\,\ddots\,\vdots \\ 0\,0\dots1 \end{pmatrix} = \begin{pmatrix} p\,\boldsymbol{c}^T\boldsymbol{M}^{-1}\boldsymbol{c} & p_1\,\boldsymbol{c}^T\boldsymbol{M}^{-1}\boldsymbol{x_1} & \dots & p_n\,\boldsymbol{c}^T\boldsymbol{M}^{-1}\boldsymbol{x_n} \\ p\,\boldsymbol{x_1}^T\boldsymbol{M}^{-1}\boldsymbol{c} & p_1\,\boldsymbol{x_1}^T\boldsymbol{M}^{-1}\boldsymbol{x_1} & \dots & p_n\,\boldsymbol{x_1}^T\boldsymbol{M}^{-1}\boldsymbol{x_n} \\ \vdots & \vdots & \ddots & \vdots \\ p\,\boldsymbol{x_n}^T\boldsymbol{M}^{-1}\boldsymbol{c} & p_1\,\boldsymbol{x_n}^T\boldsymbol{M}^{-1}\boldsymbol{x_1} & \dots & p_n\,\boldsymbol{x_n}^T\boldsymbol{M}^{-1}\boldsymbol{x_n} \end{pmatrix} \quad (14)$$

Note that $x_1, \dots, x_n$ and $C$ must be all different, otherwise $\boldsymbol{S}$ and $\boldsymbol{R}$ are singular.

The element $[1,1]$ of this matrix equation gives:

$$p = \frac{1}{\boldsymbol{c}^T\,\boldsymbol{M}^{-1}\,\boldsymbol{c}} \ . \quad (15)$$

This $p$ has an important extremal property.

**Theorem 7.** *$p$ is the maximum possible mass at point $C$ that a distribution may have whose first $2n$ moments are $\{\mu_0, \dots, \mu_{2n}\}$.*

*The difference of any 2 distributions ($\sigma_1(x)$ and $\sigma_2(x)$) with first $2n$ moments equal to $\{\mu_0, \dots, \mu_{2n}\}$ cannot be greater than $p$:*

$$\left| \int_{-\infty}^{C+0} \mathrm{d}\sigma_1(x) - \int_{-\infty}^{C-0} \mathrm{d}\sigma_2(x) \right| \le p \ . \quad (16)$$

For a proof see Appendix A.

Let us consider the first columns of the matrices in (14) below $[1,1]$:

$$p\,\boldsymbol{x_i}^T\boldsymbol{M}^{-1}\boldsymbol{c} = 0, \qquad i = 1, 2, \dots, n \ . \quad (17)$$

According to Lemma 3 $\boldsymbol{M}^{-1}$ is a positive definite matrix and it follows that $p > 0$. So the left hand side of (17) equals to 0 if $x_i$ $(i = 1, 2, \dots, n)$ is the root of the polynomial

$$w(x) = \boldsymbol{x}^T\boldsymbol{M}^{-1}\boldsymbol{c}, \quad (18)$$

where $\boldsymbol{x}$ is a vector consisting of the powers of the unknown $x$. This polynomial is of order $n$, and the roots $x_1, \ldots, x_n$ must be real and different [9], which involves that all the roots of $w(x)$ appear in (14). So the points of the discrete distribution are actually the roots of $w(x)$ and they are real and distinct according to the theory of orthogonal polynomials [9].

The diagonal elements of the matrix equation results the weights of the discrete distribution:

$$p_i = \frac{1}{{\boldsymbol{x_i}}^T \boldsymbol{M}^{-1} \boldsymbol{x_i}} \qquad i = 1, 2, \ldots, n \ , \tag{19}$$

since $\boldsymbol{M}^{-1}$ is a positive definite matrix, it follows that ${\boldsymbol{x_i}}^T \boldsymbol{M}^{-1} \boldsymbol{x_i} > 0$.

If $C$ coincides with one of the $x_i$ roots, then the discrete distribution would have only $n$ points which means that we should have $\boldsymbol{M} := \boldsymbol{M}_{n-1}$ as starting point in creating a discrete distribution instead of $\boldsymbol{M}_n$, otherwise all other steps of the process are the same.

$\sigma(x)$ denotes the discrete distribution supported on the $x_1, x_2, \ldots, x_n, C$ points with weights $p_1, p_2, \ldots, p_n, p$ .

## 4    Lower and upper limits

The following theorem is the base for finding the minimum and maximum values of all the functions having the same $\{\mu_0, \ldots, \mu_{2n}\}$ moments.

Let $\sigma^*(x)$ be a distribution satisfying (2) whose moments are $\mu_0, \ldots, \mu_{2n}$, but different from $\sigma(x)$.

**Theorem 8.** *The following relations hold for $\sigma(x)$ and $\sigma^*(x)$:*

$$\int_{-\infty}^{C-0} \mathrm{d}\sigma^*(x) \geq \int_{-\infty}^{C-0} \mathrm{d}\sigma(x), \tag{20}$$

$$\int_{-\infty}^{C+0} \mathrm{d}\sigma^*(x) \leq \int_{-\infty}^{C-0} \mathrm{d}\sigma(x) + p \ . \tag{21}$$

The proof of Theorem 8 is provided in Appendix B.

According to this theorem no function has either smaller and greater value at $C$ than $\sigma(x)|_{x=C-0}$ and $\sigma(x)|_{x=C+0}$ have. $\sigma(x)|_{x=C-0}$ and $\sigma(x)|_{x=C+0}$ can be calculated as

$$\int_{-\infty}^{C-0} \mathrm{d}\sigma(x) = \sum_{i:x_i<C} p_i = L, \qquad \int_{-\infty}^{C+0} \mathrm{d}\sigma(x) = \sum_{i:x_i<C} p_i + p = U \ . \tag{22}$$

The lower limit of the distribution is obtained as the sum of the weights of the points smaller than $C$, i.e. $L$, and the upper limit is the sum of the lower limit and the maximum mass at $C$, i.e. $U$.

The $x_i$ roots of (18) depend on $C$, but their location can be characterized by a series independent of $C$ according to the following theorem. Let us denote by $u_1, \ldots, u_n$ the roots of $P(x)$ defined in (6). These roots are also real and different [9].

**Theorem 9.** *If $C$ is such that $u_{j-1} < C < u_j$ and the $x_1, x_2, \ldots, x_n$ and $u_1, u_2, \ldots, u_n$ roots are increasingly ordered then the $x_1, x_2, \ldots, x_n, C$ and the $u_1, u_2, \ldots, u_n$ numbers are mutually separated as*

$$x_1 < u_1 < x_2 < u_2 < \ldots < u_{j-1} < C < u_j < x_j < u_{j+1} < \ldots < u_n < x_n. \quad (23)$$

The proof is provided in Appendix C.

According to Theorem 9 the number of $x_i$ roots smaller (greater) than $C$ equals to the number of $u_i$ roots smaller (greater) than $C$. In this way the $u_i$ roots define the number of points in the summation in (22). Based on Theorem 9 it is sufficient to calculate only the $x_i$ roots smaller than $C$, (or alternatively the $x_i$ roots greater than $C$).

Another consequence of Theorem 9 is that once the roots of $P(x)$ are calculated (this has to be done only once independent of $C$), we only need to calculate the roots of $w(x)$ in either the $(u_1, C)$ or the $(C, u_n)$ interval – depending on which contains less $u_i$ roots. In this way we can reduce the number of roots to compute (it is no more than $\left\lfloor \frac{n}{2} \right\rfloor$) and we know the intervals where the roots are located, which allows the use of sophisticated numerical algorithms.

## 5    Steps of the algorithm

**Given**: $\mu_0, \mu_1, \ldots, \mu_m$ and a set of $C$ values where we need to bound the distribution.

1. Test if the moments satisfy the

$$|\boldsymbol{M}_k| \geq 0 \qquad k = 0, 1, \ldots n \quad (24)$$

   inequalities. We denote the number of applicable moments (for which the (24) inequalities hold) by $2n + 1$ ($\mu_0, \ldots, \mu_{2n}$).

2. Find the roots of the polynomial $P(x)$:

$$P(x) = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k-1} & \mu_k & \cdots & \mu_{2n-1} \\ 1 & x & \cdots & x^k \end{vmatrix} . \quad (25)$$

   The roots are called $u_1 < u_2 < \ldots < u_n$ .

3. Do for each $C$ point of interest
   (a) If $C = u_i$ for some $i$ then $\boldsymbol{M} := \boldsymbol{M}_{n-1}$ else $\boldsymbol{M} := \boldsymbol{M}_n$ .
   (b) Calculate the largest possible $p$:

$$p = \frac{1}{\boldsymbol{c}^T \, \boldsymbol{M}^{-1} \, \boldsymbol{c}} \quad . \tag{26}$$

   (c) Calculate the points of the reference discrete distribution: if $(u_1, C)$ contains less $u_i$ than $(C, u_n)$, then find all the roots of the following polynomial that are *smaller* than $C$ :

$$\boldsymbol{c} \, \boldsymbol{M}^{-1} \, \boldsymbol{x}^T = 0, \tag{27}$$

   where

$$\boldsymbol{x}^T = \left(1, x, x^2, \ldots, x^n\right)^T \quad . \tag{28}$$

   Else find all the roots of the same polynomial that are *greater* than $C$ . The results are the $x_i$, $i = 1, \ldots, r$ points, where $r$ denotes the number of roots that had to be calculated: $r \leq \left\lfloor \frac{n}{2} \right\rfloor$ .
   (d) Calculate the weights of the reference discrete distribution:

$$p_i = \frac{1}{\boldsymbol{x_i}^T \, \boldsymbol{M}^{-1} \, \boldsymbol{x_i}}, \qquad i = 1, 2, \ldots, r \quad . \tag{29}$$

   (e) $L$ and $U$ are given by the following formula, if the interval $(u_1, C)$ contains less $u_i$ than $(C, u_n)$:

$$L = \sum_{i:x_i < C} p_i, \qquad U = L + p, \tag{30}$$

   else

$$L = U - p, \qquad U = 1 - \sum_{i:x_i > C} p_i \quad . \tag{31}$$

## 6   Implementation notes

The presented algorithm involves tasks which are numerically hard and unstable in general (calculating determinants and matrix inverse, finding the roots of a polynomial), but the matrices and the polynomials we consider have special properties that commonly eliminate these numerical difficulties.

First of all we calculate determinants of symmetric matrices. We use the LU-decomposition [10, p. 43 – 50] for the calculation. We experienced numerical problems with matrices larger than $11 \times 11$, and this limits the applicability to less than 23 moments using standard floating point numbers.

To invert a positive definite symmetric matrix we use *Cholesky decomposition* with backsubstitution (see [10, p. 96–98]). The Cholesky decomposition is

extremely stable numerically and approximately two times faster than the alternative methods for solving linear equations. The algorithm fails only when the matrix is not positive definite.

For finding all the roots of $P(x)$ we use the *Laguerre′s method* [10, p. 371 – 374]. In general it is not so easy to find the roots of a polynomial when we know nothing about the location of the roots. But this algorithm works well, if all the roots of a polynomial are real, because then it is theoretically guaranteed that the method converges to a root from any starting point, and fortunately this is the case for $P(x)$. Technically it requires complex arithmetic even while converging to real roots.

Finding the roots of the polynomial $w(x)$ is an easier task (remember that not all of them is needed). According to Theorem 9 a single real root of $w(x)$ lies in $(u_{i-1}, u_i)$ (or in $(u_i, u_{i+1})$ if $u_i > C$). Finding a root in a bounded interval is much easier, than in the case when we know nothing about the position. For this task we use the *bisection algorithm* [10, p. 350 – 354].

The overall algorithm is neither CPU, nor memory intensive. In order to estimate a distribution in $N$ points we perform

- $\lfloor \frac{n+1}{2} \rfloor$ times – calculation of determinants;
- 1 time – finding $n$ roots of $P(x)$;
- 1 time – inversion of an $(n+1) \times (n+1)$ matrix ;
- $N$ times – findings of max. $\lfloor \frac{n}{2} \rfloor$ roots of $w(x)$;
- $2N$ times – vector-matrix multiplications of size $(n+1) \times (n+1)$;
- $2N$ times – scalar product of vectors of size $(n+1)$.

## 7 Example of application

This section presents an example where the two-step performance analysis process (model → moments → measure) presented in Figure 1 has advantage compared to the direct computation of the value of a distribution function (model → measure).

A telecommunication system example was introduced in [11]. The authors considered a bandwidth-sharing strategy on a single link for the following 3 traffic classes:

- *rigid*: always requires peak bandwidth allocation;
- *adaptive*: it has a peak and minimum bandwidth requirement, and the actual transfer rate depends on the link utilization (for example live video transfer or voice conversation, where quality degradation is allowed to a certain degree, but high delay variance is not preferred);
- *elastic*: similar to the adaptive class, but these flows are in the system until a given amount of data is transferred (e.g. ftp-connections, where delay is allowed but data loss is not).

The number of ongoing rigid, adaptive and elastic traffic flows $(n_{\text{rigid}}, n_{\text{adaptive}}, n_{\text{elastic}})$ represents the system state in a given instant due to

the applied memoryless assumptions. Figure 2(a) shows the part of the state space, where $n_{rigid} = 1$. The states, where flows cannot get their maximum bandwidth are filled with grey. The numbers below the state identifier show the actual bandwidth of the adaptive and elastic flows expressed in fraction of the maximum required bandwidth.
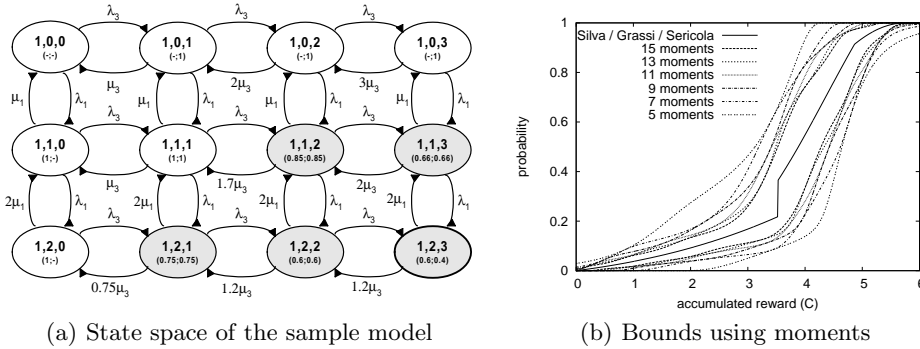


(a) State space of the sample model

(b) Bounds using moments

**Fig. 2.** Analyzing the sample model

Using this model one can evaluate the call blocking probabilities and average throughput of adaptive and elastic flows, but it is also possible to check the *throughput threshold constraint*, which is a constraint on the probability that the user-perceived throughput during the transfer of a file falls below a certain prescribed level. The calculation of the last measure requires evaluation of the MRM.

We used the MRMSolve 2.0 tool [12] to compare the complexity of the direct and the moments based analysis. The methods of Nabli and Sericola [13], De Souza and Gail [14], Donatiello and Grassi [15] result directly in the distribution and the method of Rácz and Telek [16, 17] provides the moments of the performance measure of interest.

[12] compares the different MRM solution methods. Here we only demonstrate their limits of applicability. We used a dual AMD Opteron 248 (2.2 GHz) system with 6 GB of memory running Linux operating system for the calculations. The computation time was determined by the standard Unix `time` command.

Figure 2(b) shows that we get correct bounds applying our estimation method. The more moments we use the tighter the bounds are. The methods of De Souza–Gail, Donatiello–Grassi and Nabli–Sericola result the same values.

Figure 3 is a logarithmic plot of the computation time against the size of the state space. The algorithms of De Souza–Gail and Donatiello–Grassi were terminated after 5 hours with 5,600 states, the method of Nabli and Sericola became unstable (resulting negative values of probabilities) at 7,800 states.
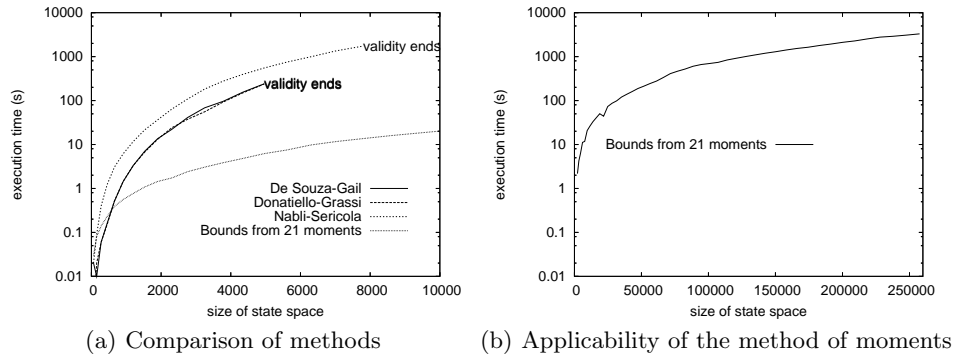
(a) Comparison of methods    (b) Applicability of the method of moments

**Fig. 3.** CPU requirements of different MRM-solver algorithms

The method of moments (Rácz and Telek) is much faster than the others, which was predictable because it yields less information about the distribution, but its main advantage is its robustness compared to the other methods: it was able to deal with a model of 260,000 states as shown in Fig. 3(b). Note that the time of the computation is dominated by the moment calculation algorithm; the estimation of the distribution based on the moments in 10,000 points required 1.44 s.

## 8    Conclusions

This paper presents an algorithm to bound a distribution based on a finite number of its moments. There are elaborated theoretical results about this problem, but the solution proposed here is different from the ones found in the literature, even though it relies on them. We also focused on the implementation and the numerical issues of the problem.

The presented example demonstrates a case when the moments based analysis is the only computationally feasible one among the solutions available in the literature.

Further research is needed to investigate the numerical behaviour of the algorithm with increased arithmetic precision. We also intend to improve the bounds based on additional information about the distribution (e.g., distributions with finite support). This kind of consideration requires refinements in the estimation procedure.

## A    Proof of Theorem 7

To prove the theorem we need the following lemma:

**Lemma 10.** *[18] If $A$ is a nonsingular $n \times n$ matrix, $u$ and $v$ are $n$ dimensional vectors and $d$ is an arbitrary real number, then*

$$\begin{vmatrix} d & v^T \\ u & A \end{vmatrix} = |A| \, (d - v^T A^{-1} u) \ .$$

(32)

The maximal mass according to [4, p. 72], can be written as:

$$p_{\max} = - \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{vmatrix} \Bigg/ \begin{vmatrix} 0 & 1 & C & \cdots & C^n \\ 1 & \mu_0 & \mu_1 & \cdots & \mu_n \\ C & \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C^n & \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{vmatrix}$$

(33)

By Lemma 10 we can write the denominator of the right hand side of (33) as $-|M| \, c^T M^{-1} c$. In addition the numerator is the determinant of $M$, hence

$$p_{\max} = - \frac{|M|}{-|M| \, c^T M^{-1} c} = \frac{1}{c^T M^{-1} c},$$

(34)

which had to be proven. The second statement of the theorem is proved in [6, p. 66].

## B    Proof of Theorem 8

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

$$\left| \int_{-\infty}^{C+0} d\sigma(x) - \int_{-\infty}^{C-0} d\sigma^*(x) \right| \leq p \ .$$

(35)

Resolving the absolute value sign this equals:

$$\int_{-\infty}^{C+0} d\sigma(x) - p \leq \int_{-\infty}^{C-0} d\sigma^*(x) \leq \int_{-\infty}^{C+0} d\sigma(x) + p.$$

(36)

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

$$\int_{-\infty}^{C+0} d\sigma(x) = \int_{-\infty}^{C-0} d\sigma(x) + p \ .$$

(37)

Substituting it to the leftmost inequality we get:

$$\int_{-\infty}^{C-0} d\sigma(x) \leq \int_{-\infty}^{C-0} d\sigma^*(x),$$

(38)

which is (20). Now using $\sigma_1(x) := \sigma^*(x)$ and $\sigma_2(x) := \sigma(x)$ substitutions Theorem 7 gives

$$\left| \int_{-\infty}^{C+0} d\sigma^*(x) - \int_{-\infty}^{C-0} d\sigma(x) \right| \leq p. \tag{39}$$

Resolving the absolute value sign we have:

$$\int_{-\infty}^{C-0} d\sigma(x) - p \leq \int_{-\infty}^{C+0} d\sigma^*(x) \leq \int_{-\infty}^{C-0} d\sigma(x) + p, \tag{40}$$

whose rightmost inequality gives (21).

## C   Proof of Theorem 9

To prove the theorem we need the following lemma.

**Lemma 11.** *[6, p. 64]) Assume that the $\{\mu_0, \ldots, \mu_{2n-1}\}$ sequence satisfies (2) and $\tilde{\sigma}(x)$ is a distribution whose first $2n$ moments are $\mu_0, \mu_1, \ldots, \mu_{2n-1}$. In this case, for $i = 1, 2, \ldots, n-1$ we have*

$$\int_{-\infty}^{u_i+0} d\tilde{\sigma}(u) \leq \rho_{n-1}(u_1) + \rho_{n-1}(u_2) + \ldots + \rho_{n-1}(u_i) \leq \int_{-\infty}^{u_{i+1}-0} d\tilde{\sigma}(u), \tag{41}$$

*where $u_1, \ldots, u_n$ are the roots of $P(x)$,*

$$\rho_n(x) = \frac{1}{\sum_{j=0}^{n} |P_j(x)|^2}, \tag{42}$$

$$P_0(x) = 1 \quad and \quad P_n(x) = \frac{1}{\sqrt{|M_{n-1}| \, |M_n|}} \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x & \cdots & x^n \end{vmatrix}. \tag{43}$$

*If in addition it is known that $\tilde{\sigma}(x)$ has more than $n$ points of increase, then the $\leq$ signs in (41) can be replaced by $<$ signs and the following inequalities hold:*

$$0 < \int_{-\infty}^{u_1-0} d\tilde{\sigma}(u), \qquad \int_{-\infty}^{u_n+0} d\tilde{\sigma}(u) < \mu_0 . \tag{44}$$

There are $n+1$ points of increase in $\sigma(\cdot)$ ($x_1, \ldots, x_n$ and $C$). Applying Lemma 11 for $\tilde{\sigma}(\cdot) = \sigma(\cdot)$ implies that the $\leq$ signs are replaced by $<$ signs in (41), i.e.

$$\int_{-\infty}^{u_i+0} d\sigma(u) < \int_{-\infty}^{u_{i+1}-0} d\sigma(u) . \tag{45}$$

This means that in any $(u_i, u_{i+1})$ ($i = 1, \ldots, n-1$) interval there must be at least one point of increase of $\sigma(x)$. Furthermore (44) implies that there is at least one point of increase in the $(-\infty, u_1)$ and at least one point of increase in the $(u_n, \infty)$ intervals.

Considering that $\sigma(\cdot)$ has $n+1$ points of increase results the theorem.

# References

1. A. Horváth M. Telek and G. Horváth. Analysis of inhomogeneous Markov reward models. *Linear algebra and its application*, 386:383–405, 2004.

2. A. van de Liefvoort. The moment problem for continuous distributions. Technical report, University of Missouri, WP-CM-1990-02, Kansas City, 1990.

3. T. Stieltjes. Reserches sur les fractions continues. *Ann. Fac. Sci. Univ. Toulouse*, 2:1–122, 1894. in French.

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

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

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. Eric W. Weisstein. Positive definite matrix. http://mathworld.wolfram.com/PositiveDefiniteMatrix.html.

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

9. G. Szegö. *Orthogonal polynomials*. American Mathematical Society, Providence, Rhode Island, 1939.

10. 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.

11. G. Fodor, S. Rácz, and M. Telek. On providing blocking probability- and throughput guarantees in a multi-service environment. *International Journal of Communication Systems*, 15:4:257–285, May 2002.

12. G. Horváth, S. Rácz, Á. Tari, and M. Telek. Evaluation of reward analysis methods with MRMSolve 2.0. In *1st International Conference on Quantitative Evaluation of Systems (QEST) 2004*, pages 165–174, Twente, The Netherlands, Sept 2004. IEEE CS Press.

13. H. Nabli and B. Sericola. Performability analysis: a new algorithm. *IEEE Transactions on Computers*, 45:491–494, 1996.

14. E. de Souza e Silva and H.R. Gail. Calculating cumulative operational time distributions of repairable computer systems. *IEEE Transactions on Computers*, C-35:322–332, 1986.

15. L. Donatiello and V. Grassi. On evaluating the cumulative performance distribution of fault-tolerant computer systems. *IEEE Transactions on Computers*, 1991.

16. S. Rácz and M. Telek. Performability analysis of Markov reward models with rate and impulse reward. In M. Silva B. Plateau, W. Stewart, editor, *Int. Conf. on Numerical solution of Markov chains*, pages 169–187, Zaragoza, Spain, 1999.

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

18. P. Rózsa. *Lineáris algebra és alkalmazásai*. Tankönyvkiadó, 1991. in Hungarian.