A moment-based estimation method for extreme probabilities

Árpád Tari[*], Miklós Telek[†] and Peter Buchholz[‡]

Abstract

The performance analysis of highly reliable and fault tolerant systems requires the investigation of events with extremely low or high probabilities. This paper presents a simplified numerical method to bound the extreme probabilities based on the moments of the distribution. This simplified method eliminates some numerically sensitive steps of the general moments based bounding procedure.

Numerical examples indicate the applicability of the proposed approach.

Keywords: reduced moment problem, moments based distribution bounding, tail distribution

1 Introduction

Performance analysis of real-life systems usually requires the evaluation of the distribution of some random variables. The direct analysis of these distributions is often infeasible due to the high computational complexity. A possible way to overcome this difficulty is to simplify the model or to calculate only an estimate of the measure of interest. Both types of simplification result in inaccuracies in calculation, but this is the price of the solvability.

In this paper we investigate the second option, the estimation of the measure of interest based on a set of its moments. There are several classes of performance analysis problems for which the analysis of the moments of a random variable is far less complex than the analysis of the distribution. For example, for the class of Markov reward models the moments of the reward measures can be computed by the effective methods presented in [12, 15], while the direct analysis of the distribution of these measures based on [8, 3, 4] is far more complex and practically infeasible for models with more than 10^4 states [7]. In these cases moments based estimation of the distribution is the only feasible solution method for large models. There are two ways of moments based estimation: to fit a certain class of distribution functions to the set of moments (e.g. [16] presents a method for fitting with matrix exponential distribution); and to calculate maximal and minimal values for the distribution among all possible distributions having the prescribed set of moments. The first approach results

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

in an unknown error if the performance measure does not belong to the considered class of distributions. To bound the error of moments based distribution approximation we apply the second approach.

Determining a distribution function based on its moments is called the *reduced moment problem* (where reduced refers to the finite number of moments). This is a well-known problem for more than 100 years and has an extensive literature. A good overview is given in [13].

We denote the i^{th} moment of a distribution function $\sigma(x)$ supported on the interval $[a, b]$ by

$$\mu_i = \int_a^b x^i \, \mathrm{d}\sigma(x), \qquad i = 0, 1, 2, \ldots, m \ . \tag{1}$$

The problem of determining a distribution whose support interval is the real axis (hence $a = -\infty$, $b = \infty$) based on its moments is called the *Hamburger* moment problem after the German mathematician who first solved this problem in 1920 [6]. We also refer to this as the *infinite case* and we discuss this problem in this paper. Other moment problems are the *Stieltjes* (when $a = 0$ and $b = \infty$) and *Hausdorff* (if $a = 0$, $b = 1$) moment problems.

The performance analysis of highly reliable or safety critical fault-tolerant systems requires the analysis very unlikely events, i.e., the distribution of a random variable at very low (close to 0) or very high (near to 1) probabilities.

In the paper we focus on the analysis of these kinds of extreme values and provide a simplified moments based estimation analysis algorithm with respect to the one that calculates lower and upper bounds for the distribution function based on a set of moments in the general case [11]. The modified algorithm is numerically stable, simple and fast.

The paper is organized as follows: Section 2 introduces the moments based estimation method. The numerical procedures involved in the solution are summarized in Section 3 and some useful expressions are deduced in Section 4. An example is analyzed in Section 5. Section 6 concludes the paper.

2 Discrete reference distribution

The method discussed here is based on the idea introduced in [10, 11]. We briefly present it here as it is the basis of our investigation.

The considered task can be formalized as follows. Find the smallest and largest values, that any distribution function $\sigma(x)$ with $\mu_0, \mu_1, \ldots, \mu_m$ moments may have at a given point C, i.e:

$$L = \min \left\{ \sigma(C) : \ \mu_i = \int_{-\infty}^{\infty} x^i \mathrm{d}\sigma(x), i = 0, \ldots, m \right\}, \tag{2}$$

$$U = \max \left\{ \sigma(C) : \ \mu_i = \int_{-\infty}^{\infty} x^i \mathrm{d}\sigma(x), i = 0, \ldots, m \right\} \ . \tag{3}$$

This means that we estimate the distribution in a single point. Estimation in an interval is only possible with a series of applications in points of the interval, but this can be done effectively repeating only parts of the algorithm.

The L and U values result from a discrete distribution that have the maximal probability mass at point C and is characterized by the $\mu_0, \mu_1, \ldots, \mu_m$ moments.

Before calculating L and U, we need to check whether the series of moments $\mu_0, \mu_1, \ldots, \mu_m$ can belong to a valid distribution function. This can be verified through the following inequalities:

$$|\boldsymbol{M}_k| \geq 0, \qquad k = 0, 1, \ldots, \left\lfloor \frac{m}{2} \right\rfloor, \tag{4}$$

where

$$\boldsymbol{M}_k = \begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_k \\ \mu_1 & \mu_2 & \cdots & \mu_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_k & \mu_{k+1} & \cdots & \mu_{2k} \end{pmatrix} . \tag{5}$$

The maximum number of moments that satisfy (4) is denoted by $2n + 1$, i.e. the considered moments are $\mu_0, \mu_1, \ldots, \mu_{2n}$, and the matrix of largest order satisfying (4) is denoted by $\boldsymbol{M} := \boldsymbol{M}_n$.

Let the roots of

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

be denoted by $u_1 < u_2 < \ldots < u_n$ in increasing order. These roots are also real and simple [14].

If $C = u_i$ for some i then the discrete distribution consists only of n points (including C) and we have to take $\boldsymbol{M} := \boldsymbol{M}_{n-1}$ [13, p. 42], so this must be checked before the calculations.

Odd number of input is needed to form the matrices $\boldsymbol{M}_k$. As a consequence if even number of moments is given, the last one (μ_{2n+1}) does not carry further information about the distribution, so this can be ignored.

The maximal probability mass that can be concentrated at C is denoted by p and calculated by [1]:

$$p = \frac{1}{\boldsymbol{c}^T \, \boldsymbol{M}^{-1} \, \boldsymbol{c}}, \tag{7}$$

where

$$\boldsymbol{c}^T = \left(1, C, C^2, \ldots, C^n\right)^T . \tag{8}$$

Furthermore, the difference between any two distribution functions with moments $\mu_0, \mu_1, \ldots, \mu_{2n}$ is not larger than p [1]. Note that the formula for p contains the inverse of $\boldsymbol{M}$, a symmetric and positive definite (due to (4)) matrix. The computation of the inverse of symmetric positive definite matrices is numerically more stable than the inversion of general matrices.

The other points of the discrete distribution are the roots of the following polynomial:

$$\chi(x) = \boldsymbol{c}^T \, \boldsymbol{M}^{-1} \, \boldsymbol{x}, \tag{9}$$

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

This is an order n polynomial and based on the theory of orthogonal polynomials [14] its roots are all real and distinct. We denote them by $x_1 < x_2 < \ldots < x_n$ in increasing order. The corresponding probability masses are

$$p_i = \frac{1}{\boldsymbol{x_i}^T \, \boldsymbol{M}^{-1} \, \boldsymbol{x_i}}, \qquad i = 1, 2, \ldots, n, \tag{10}$$

where $\boldsymbol{x_i} = \left(1, x_i, x_i^2, \ldots, x_i^n\right)^T$.

The lower limit of the distribution is obtained as the sum of the weights of the points smaller than C. The upper limit is the sum of the lower limit and the maximum mass at C:

$$L = \sum_{i:\, x_i < C} p_i, \qquad U = L + p \ . \tag{11}$$

This discrete distribution is extreme in that sense, that no other distribution function with moments μ_i has either a lower or higher value at C than L and U, respectively.

The algorithm can be simplified using the following interesting property of the points x_i. These points depend on C, but their locations can be characterized by a series independent of C. The $x_1, x_2, \ldots, x_n, C$ and the $u_1, u_2, \ldots, u_n$ roots (see (6)) are mutually separated as

$$x_1 < u_1 < x_2 < u_2 < \ldots < u_{j-1} < C < u_j < x_j < u_{j+1} < \ldots < u_n < x_n \ . \tag{12}$$

The number of points x_i which are smaller (greater) than C equals the number of points u_i that are smaller (greater) than C. As a consequence the roots $u_1, u_2, \ldots, u_n$ define the number of terms considered in (11). Therefore it is sufficient to calculate only the roots x_i smaller than C (or alternatively the roots x_i greater than C). If $C < u_1$ or $C > u_n$ we do not need to calculate the points of the discrete distribution, because in these cases the lower and upper limits are determined by p as follows:

$$L = 0, \qquad U = p, \qquad \text{if } C < u_1, \tag{13}$$

$$L = 1 - p, \qquad U = 1, \qquad \text{if } C > u_n \ . \tag{14}$$

We use these simple relations to bound the probability of extreme events and this type of estimation is called the *simplified* case. The numerical procedure is summarized in Figure 1.

3 Computational complexity

Some tasks in the proposed algorithm may involve numerical difficulties in the general case (e.g. evaluating determinants, inverting matrices, finding roots of polynomials), however the matrices and the polynomial considered here have special properties that make it possible to use numerically more stable methods to calculate them.

To calculate the determinants of symmetric matrices we use the *LU decomposition* [9, p. 43 – 50]. Testing with known distributions the maximum dimension of the matrix whose determinant could be computed correctly is 15×15, bigger matrices resulted negative determinants showing numerical instabilities in the method. Therefore the limit of the applicability is 29 moments using standard floating point arithmetic, but the maximum number of moments that satisfy (4) largely depends on the original distribution: our experiences show that in general the number of usable moments is around 20, but in some cases it is below 15.

We use *Cholesky decomposition* with backsubstitution to invert the positive definite matrix $\boldsymbol{M}$ [9, p. 96–98]. This method is known to be extremely stable

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

1. Test if the moments satisfy the
$$|\boldsymbol{M}_k| \geq 0 \qquad k = 0, 1, \ldots, \lfloor m/2 \rfloor \tag{15}$$
inequalities, where
$$\boldsymbol{M}_k = \begin{pmatrix} \mu_0 & \mu_1 & \ldots & \mu_k \\ \mu_1 & \mu_2 & \ldots & \mu_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_k & \mu_{k+1} & \ldots & \mu_{2k} \end{pmatrix} . \tag{16}$$
We denote the number of applicable moments (for which the (15) inequalities hold) by $2n + 1$ $(\mu_0, \ldots, \mu_{2n})$.
2. Find the roots of the polynomial $P(x)$:
$$P(x) = \begin{vmatrix} \mu_0 & \mu_1 & \ldots & \mu_n \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k-1} & \mu_k & \ldots & \mu_{2n-1} \\ 1 & x & \ldots & x^k \end{vmatrix} . \tag{17}$$
The roots are called $u_1 < u_2 < \ldots < u_n$.
3. Do for each $C < u_1$ or $C > u_n$ point of interest
   (a) Calculate the largest possible p:
   $$p = \frac{1}{\boldsymbol{c}^T \, \boldsymbol{M}^{-1} \, \boldsymbol{c}}, \tag{18}$$
   where
   $$\boldsymbol{c}^T = \left(1, C, C^2, \ldots, C^n\right)^T . \tag{19}$$
   (b) If $C < u_1$, then $L = 0,\ U = p$.
   If $C > u_n$, then $L = 1 - p,\ U = 1$.

Figure 1: Steps of the algorithm

numerically and approximately two times faster than the alternative methods for solving linear equations. It fails only if the matrix is not positive definite.

It is hard to find the roots of a polynomial if we do not know anything about the location of the roots. But all the roots of $P(x)$ are real, and in this case *Laguerre's method* [9, p. 371 − 374] works well as it is theoretically guaranteed that this algorithm converges to a root from any starting point.

Figure 1 shows that at different values of C only p has to be recalculated, hence the overall algorithm is neither CPU, nor memory intensive. Table 1 shows the required operations in order to estimate a distribution in N points using m moments $(\mu_0, \mu_1, \ldots, \mu_{m-1})$ out of whom $2n{+}1$ $(\mu_0, \mu_1, \ldots, \mu_{2n})$ defines a valid moment sequence.

Task	Nr. of executions
calculation of determinants	$\lfloor m/2 \rfloor + 1$
finding n roots of $P(x)$	1
inversion of an $(n+1) \times (n+1)$ matrix	1
vector-matrix multiplications of size $(n+1) \times (n+1)$	$2N$
scalar product of vectors of size $(n+1)$	$2N$
reciprocal	$2N$

Table 1: Computational cost of the simplified estimation

4 Closed-form expressions

The applicability of the simplified estimation depends on the smallest and the largest root of $P(x)$. If the degree of the polynomial $P(x)$ is less than 5, then closed form expressions can be deduced for u_i, though for degrees 3 and 4 these expressions are much too complicated and would fill several pages.

However if the degree of $P(x)$ is equal to 2 (hence we have 5 moments as input: $\mu_0, \mu_1, \mu_2, \mu_3$ and μ_4) the formulas for u_1, u_2 and even for p are quite simple. Discrete construction is needed only in the interval $[u_1, u_2]$.

$$u_{1,2} = \frac{\mu_1\mu_2 - \mu_0\mu_3 \pm \sqrt{-3\mu_1^2\mu_2^2 + 4\mu_0\mu_2^3 + 4\mu_1^3\mu_3 - 6\mu_0\mu_1\mu_2\mu_3 + \mu_0^2\mu_3^2}}{2\mu_1^2 - 2\mu_0\mu_2}, \tag{20}$$

$$\begin{aligned} \frac{1}{p} =& \frac{C^4(\mu_1^2 - \mu_0\mu_2) + C^3(-2\mu_1\mu_2 + 2\mu_0\mu_3) + C^2(3\mu_2^2 - 2\mu_1\mu_3 - \mu_0\mu_4) +}{\mu_2^3 + \mu_0\mu_3^2 + \mu_1^2\mu_4 - \mu_2(2\mu_1\mu_3 + \mu_0\mu_4)} \\ & \frac{+C(-2\mu_2\mu_3 + 2\mu_1\mu_4) + (\mu_3^2 - \mu_2\mu_4)}{\mu_2^3 + \mu_0\mu_3^2 + \mu_1^2\mu_4 - \mu_2(2\mu_1\mu_3 + \mu_0\mu_4)} \ . \end{aligned} \tag{21}$$

Having 3 input moments ($\mu_0 = 1, \mu_1$ and μ_2) the discrete reference distribution contains only 1 point: C . The only root of $P(x)$ and the maximal concentrated mass at C are the following:

$$u_1 = \mu_1, \qquad p = \frac{\mu_2 - \mu_1^2}{C^2 - 2C\mu_1 + \mu_2} \ . \tag{22}$$

The lower and upper bounding functions can be expressed by simple formulas along the whole real axis.

$$L = \begin{cases} 0 & \text{if } C < \mu_1 \ , \\ \dfrac{(C-\mu_1)^2}{C^2 - 2C\mu_1 + \mu_2} & \text{if } C \geq \mu_1 \ , \end{cases} \tag{23}$$

$$U = \begin{cases} \dfrac{\mu_2 - \mu_1^2}{C^2 - 2C\mu_1 + \mu_2} & \text{if } C < \mu_1 \ , \\ 1 & \text{if } C \geq \mu_1 \ . \end{cases} \tag{24}$$

It is easy to see that L and U are continuous functions of C .

These formulas are simple but they make only rough estimations possible. The next section shows how the increasing number of moments affects accuracy.

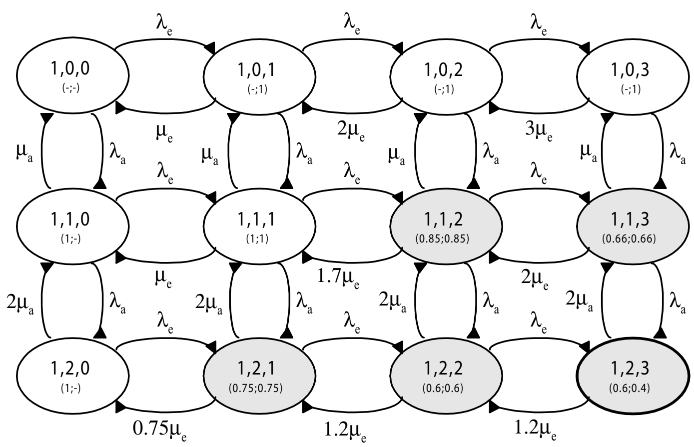

Figure 2: State space of the sample model

5 Example of application

This section demonstrates the properties of the proposed approach through an example, pointing out its strengths and weaknesses.

[5] introduced a strategy to share a telecommunication link between different traffic classes to satisfy certain pre-defined Quality of Service (QoS) constraints. Three traffic classes are defined:

- **rigid**: require constant bandwidth (b_r) allocation;
- **adaptive**: characterized by peak (b_a) and minimum bandwidth (b_a^{min}) requirements, the actual bandwidth usage depends on the link utilization (for example a video stream with adaptive compression level, where quality degradation is allowed to a certain degree, but high delay variance in not);
- **elastic**: similar to the adaptive class regarding their bandwidth requirements (b_e and b_e^{min}), but they stay in the system until a given amount of data has been transmitted (for example an ftp-session, where transfer rate changes are allowed, but data loss is not).

A Markov reward model (MRM) is used to describe system behavior. The states of the system are represented by a triple (n_r, n_a, n_e) which are the number of active flows in the system belonging to the rigid, adaptive and elastic flows, respectively. The arrival rates are $\lambda_r, \lambda_a, \lambda_e$, and the departure rates are μ_r, μ_a, μ_e. μ_e is called the *maximal* departure rate of an elastic flow experienced when maximal bandwidth is available, the *actual* departure rate is proportional to the available bandwidth, which is a function of n_r, n_a and n_e. The transition rates of the MRM are calculated from these rates, and the reward rates associated with each state are the actual bandwidth of the elastic class.

Figure 2 shows a portion of the state space in case of $n_r = 1$. The states where the elastic flows do not get the maximal bandwidth are printed in grey. The numbers below the state identifiers indicate the actual bandwidth of the adaptive and elastic flows as a fraction of their peak bandwidth.

The performance measure of our interest is the distribution of the amount of time, $T(\xi)$, required to transmit ξ amount of data by an elastic traffic flow.

We would like to ensure that the transmission completes before time t with a very high probability:

$$\Pr\left(T(\xi) < t\right) > \varepsilon, \tag{25}$$

where ε is a prescribed constant close to 1 $(0.99, \ldots, 0.99999)$. The amount of data is given and we are interested in the minimum value of t which means that the transfer of ξ amout of data will be finished during the interval $[0, t_{\min})$ with probability e.g. 0.9999 but this is not true for any $t < t_{\min}$.

This investigation requires evaluation of the MRM. We compare two different analysis approaches:

1. the moment-based method in [15] with estimation based on the moments;
2. direct analysis of the distribution of the completion time: methods of Nabli and Sericola [8], De Souza e Silva and Gail [2], Donatiello and Grassi [4].

The algorithms were implemented by their original paper. We use a dual AMD Opteron 248 (2.2 GHz) system with 6 GB of RAM running Linux for computations.

5.1 Correctness

To verify the procedures we evaluate a sample system with 105 states and calculate the whole distribution of the amount of transmitted data. The three direct methods result in the same values and the moments-based method gives real bounds as it is depicted in Figure 3. The more moments are given the tighter the bounds are. It is also observable that convergence slows down with the increasing number of moments. The bounds are the widest around the mean of the distribution. We are able to do the estimations with maximum 17 moments, because using more moments results in negative determinant while testing the necessary condition of existence (4). This is due to numerical instabilities in the procedure that calculates the determinant of a matrix.

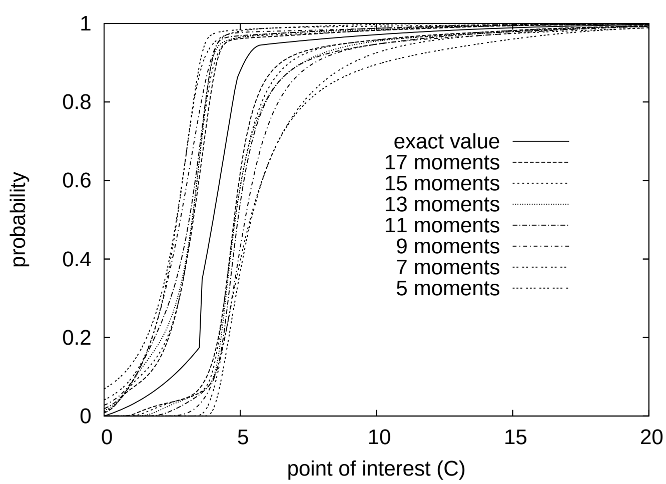

Figure 3: Distribution of the transmission time of ξ amount of elastic data

Moments	Valid from	0.9999	0.99999	0.999999
5	13.668	48.542	78.919	122.966
7	19.760	33.093	43.244	55.756
9	23.188	29.364	34.462	44.324
11	25.071	28.268	31.472	34.967
13	26.285	27.902	30.121	32.451
15	27.129	27.818	29.486	31.144
17	27.698	27.815	29.169	30.405
Exact		26.590	28.373	29.145

Table 2: Moments based bounding of the tail distribution

5.2 Numerical results

We evaluate and estimate $t_{\min}$, i.e. the minimum of t that satisfies (25). Three values of ε are considered: 0.9999, 0.99999 and 0.999999, ξ is set to 100. Fig. 4 shows the exact distribution and the bounds we get using different number of moments in case of $\varepsilon = 0.9999$. Thick black line represents this value. All the three direct analysis methods result the same values, the corresponding curve is labeled "exact" and $t_{\min}$ is the point where it reaches 0.9999 . When estimating a distribution based on its moments we get a lower and an upper bounding function. In these special cases that we investigate the upper bounding function is always equal to 1 and that's why it is omitted in the figure. The lower estimation is always smaller than the real value in any point of interest C, hence all the lower bounding functions corresponding to different number of moments are below the exact distribution function. As a consequence these functions intersect the line 0.9999 at greater values of t than $t_{\min}$.

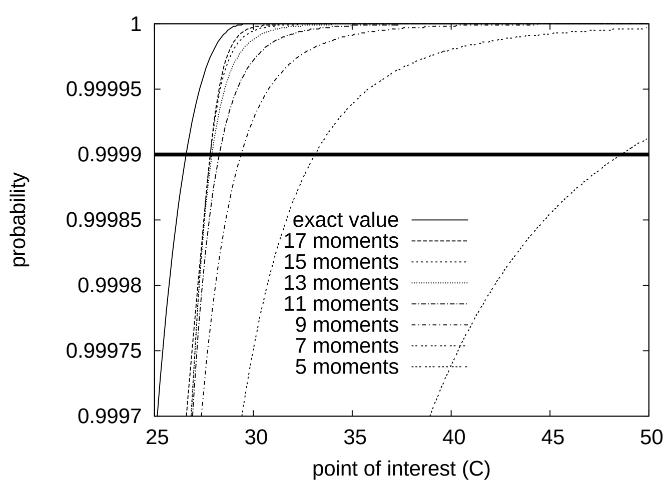

Figure 4: Lower estimation reaches to 0.9999

Table 2 presents the experiences. The 3^{rd}, 4^{th} and 5^{th} columns contain the results at different values of accuracy ε . The last row contains the "Exact" values which result from the direct distribution analysis. The other rows show the points where the moment-based estimation reaches the predefined level of accuracy. The “Valid from” column indicates u_n, i.e. from which the presented simple bounding method is applicable (see (13)) and no reference discrete distribution is needed.

The table clearly shows that more moments contain more information about the tail distribution, and the estimated value of $t_{\min}$ is closer to the real one in these cases. However convergence slows down as the number of used moments increases.

5.3 Size of the state space

We evaluated a series of runs to determine the maximum number of states which the different types of solvers are still capable to calculate. We considered a method unusable if it resulted in clearly invalid values (e.g., negative possibilities) or the running time was more than 20× of the previous configuration.

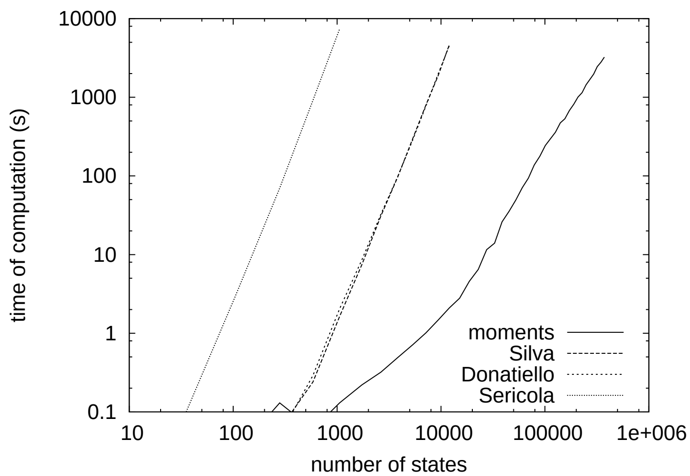

Figure 5: Evaluation time vs. state space size in logarithmic scale

Using the moments based method we could calculate the model with 370,000 states, while direct methods calculated the model with maximum 12,000 states. On the other hand the moments based approach yields less information about the distribution. The evaluation time of the estimation from the moments is 0.01s, its contribution to the overall calculation time in all considered cases is negligible.

6 Conclusion

In this paper we focus on a special use of our previously developed moments based distribution bounding method. For the computation of the distribution of extreme events the moment based analysis simplifies, because the maximal probability mass at the point of interest defines the bounds of the distribution.

We present an example where the simple bounding method is efficient and accurate compared to the results of other methods that calculate directly the values of the distribution function.

We plan to increase the accuracy of our algorithm by using extended precision arithmetic and to improve our method using additional information about the distribution functions such as finite support intervals.

References

[1] 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).

[2] 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.

[3] E. de Souza e Silva and R. Gail. An algorithm to calculate transient distributions of cummulative rate and impulse based reward. *Commun. in Statist. – Stochastic Models*, 14(3):509–536, 1998.

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

[5] 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.

[6] H. Hamburger. Über eine Erweiterung des Stieltjes'schen Momentproblems. *Mathematische Annalen*, 81:235–319, 1920.

[7] 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.

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

[9] 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.

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

[11] S. Rácz, Á. Tari, and M. Telek. A moments based distribution bounding method. 2005. to appear in Computers and mathematics with applications.

[12] 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.

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

[14] G. Szegö. *Orthogonal polynomials.* American Mathematical Society, Providence, Rhode Island, 1939.

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

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