Numerical behavior of the moment based estimation algorithms

Árpád Tari
Lehrstuhl IV Informatik
Universität Dortmund
August-Schmidt-Str. 12
D-44227 Dortmund, Germany
Email: arpad@sch.bme.hu

Miklós Telek
Department of Telecommunications
Budapest University of Technology and Economics
Magyar tudósok körútja 2.
H-1117 Budapest, Hungary
Email: telek@webspn.hit.bme.hu

***Abstract*— Moments based estimation methods give a possible solution to those problems whose direct distribution analysis is infeasible. In this paper we compare the numerical properties of 3 moments based estimation methods. The estimation methods differ in the support interval of the estimated distribution. We consider infinite support, lower bounded support and finite support.**

I. Introduction

Modeling real systems typically results in the determination of performance measures represented by random variables. In case of huge state space and/or complex model the determination of the distribution of such random variables is hard, often technically infeasible. One possible way to overcome these difficulties is to calculate only the moments of the distribution in question, then to estimate the distribution function based on its moments.

Several methods are known that result in the moments at a lower computational cost than the analysis of the distribution, e.g. transform domain analysis, randomization based analysis, Markov reward models [3].

The problem of determining the distribution based on its finite number of moments is called the *reduced moment problem* ("reduced" because of the finite number of input). A special case was solved by Chebyshev in 1873, and the first extensive treatment was done by Stieltjes [4] in 1894, who also named the problem. The contribution to the investigation of the problem of Markov, Hamburger, Hausdorff, Nevanlinna, M. Riesz, Carleman and Stone is worthy of note. A good historical overview on the subject can be found in [5]. The mathematical tools involved in the solution of the problem were continued fractions, approximate quadratures of integrals, singular integral equations, orthogonal polynomials and operators in Hilbert space [5]. The approach applied in this paper is based mainly on matrix operations, though it also relies on the theory of orthogonal polynomials.

The i^{th} moment of the distribution $\sigma(x)$ is denoted by μ_i and defined by

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

$\mu_0 = 1$ if the distribution is non-defective and μ_1 is called the *mean* of the distribution. According to the support interval of the distribution function the following possibilities are important:

- $a = -\infty, b = \infty$: **infinite case**, Hamburger moment problem;
- a finite, $b = \infty$: **positive case**;
- both a and b are finite: **bounded case**.

In line with these cases three different methods exist. The fourth case when $a = -\infty$ and b is finite is not considered as an independent case here. The problem is called *Stieltjes moment problem* if $a = 0, b = \infty$ and *Hausdorff moment problem* if $a = 0, b = 1$, but we work on the more general cases. The methods were implemented on the basis of [6] and [7].

The algorithms determine the minimum and maximum values that distribution functions can have at some point $x = C$ whose moments are μ_i, $i = 0, 1, \ldots, m$. This way we define a lower and an upper limit that are strict in that sense that there always exist distributions that reach these values.

The paper is organized as follows: Section II gives a short introduction into the distribution estimating methods whose numerical behavior is analyzed in Section III. Section IV summarizes the experiences.

II. The algorithms

The three different algorithms (according to the support interval of the distribution function) are based on the same idea: the bounds come from the discrete distribution of a minimal number of points including C, the point of interest. This distribution is unique and always exists [5].

The 3 algorithms share the main steps. To estimate the distribution in a point C we have to

1) determine whether the input $\mu_0, \mu_1, \ldots, \mu_m$ can be a series of moments of any distribution function supported on $[a, b]$;
2) calculate the maximum possible mass p at the point of interest C (also determine the maximum masses p_a and p_b at the endpoints a and b of the support interval if a and/or b are finite);

3) evaluate the points of the discrete reference distribution: $x_i, i = 1, 2, \ldots, n$;
4) calculate the weights of the discrete reference distribution: $p_i, i = 1, 2, \ldots, n$;
5) determine the lower and the upper estimation.

If calculations in several points are necessary, then they can be done efficiently repeating only parts of the algorithms which are neither CPU nor memory demanding.

Due to the nature of the positive and bounded estimations the bounds can be calculated by two different ways in each case. The mode of calculation depends on the point of interest (C). The support interval can be divided into subintervals where different formulas are valid. This behavior result in a continuous but not everywhere differentiable lower and upper limit functions. This effect can be clearly seen on the plots of the positive and bounded estimation.

III. Results

We choose known distribution functions and estimate them based on their moments.

An estimation is considered numerically incorrect if

- the bounding functions are not continuous;
- the results are not real bounds: the lower (upper) bound function must not go above (below) the known distribution function;
- any parameter of the result is invalid: e.g. the discrete distribution has negative weight(s) and/or point(s) outside of the support interval.

A. Number of used moments

We investigate how many moments can be used for the estimation, how the increasing number of input moments affects accuracy and whether the methods give correct results.

1) Infinite case: The sample distribution function was a normal distribution with mean $m = 1$ and variance $\sigma = 3$. Normal distributions are supported on the whole real axis. Due to the nature of the algorithm only odd number of moments can be used, in case of even number of moments the last one does not carry any information about the distribution and that's why it is omitted.

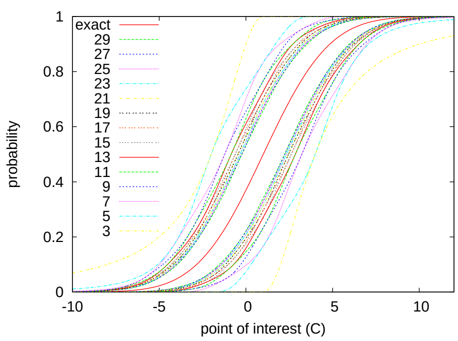

Fig. 1. Different number of moments used for estimation (infinite case)

Figure 1 shows the effect of different number of moments used for estimation (the number of moments is shown in the legend box). It is easy to see that accuracy increases with the number of moments, i.e. the lower and upper limit function pairs are closer to each other. Theoretically provided that the normal distribution is uniquely determined by its moments [1, p. 26], but convergence slows down. The algorithm could work well with as many as 29 moments, however in practice it makes no sense to do work with more than 19 moments as further moments do not improve the bounds much.

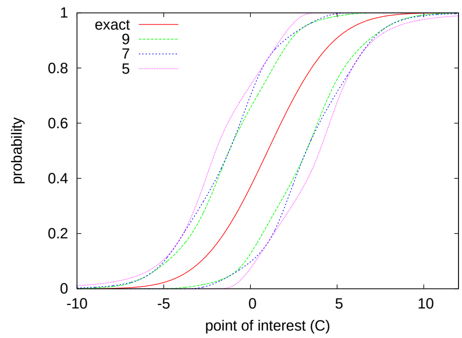

Fig. 2. Gain using more moments for estimation (infinite case)

Fig. 2 depicts the gain of using two more moments featuring the graphs of the estimations from 5, 7 and 9 moments. The consecutive estimations have points in common (which is also theoretically given [5]), and around these points there are intervals where the gain is negligible. Otherwise between two points of contact more moments give tighter bounds.

2) Positive case: First we use a distribution which has a nonzero jump at a, because the main advantage of the positive estimation over the infinite one that it can handle nonzero weight at the left endpoint of the interval, if the distribution is supported over a semi-axis.

The reference distribution is the distribution of the accumulated reward of a Markov reward model. Its formula in double transform domain (see [2]) is:

$$R(s, v) = \frac{s(9 + s(5 + s)) + 3s(3 + s)v + 2(1 + s)v^2 + 9(1 + v)}{sv(3 + s)(s(3 + s)^2 + 3(1 + s)(3 + s)v + 2(2 + s)v^2)}$$

It is supported on $[0, \infty)$ and starts with a jump $e^{-1} \approx 0.36788$ at 0.

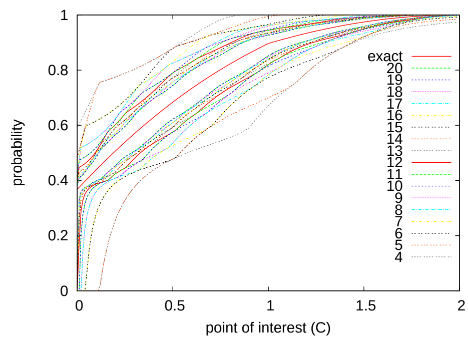

Fig. 3. Different number of moments used for estimation (positive)

Fig. 3 shows the difference between the cases with different number of input. Initially 23 moments were given but after testing the conditions whether they can form a distribution over $[0, \infty)$ only 20 of them remained because of the floating point error of the procedure that calculates determinant.

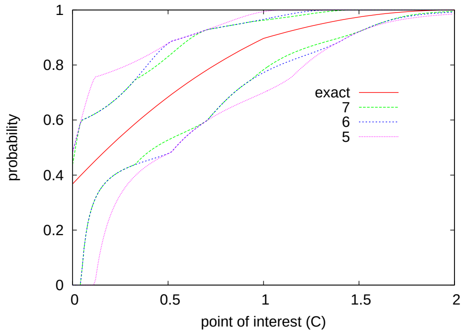

Fig. 4. Gain using more moments for estimation (positive)

More moments clearly result in tighter bounds which is even easier to see in Fig. 4. The bounding functions deflect in certain points, according to the discussion at the end of Sec. II. There are intervals where two consecutive (i.e. the number of used moments differs only by 1) estimations give the same results, though using more moments never result in broader bounds.

Set of distributions with the same moments series. The previous two examples suggest that increasing the number of moments to infinity theoretically uniquely defines the distribution. It is not always the case. The series of moments

$$\mu_i = e^{(i^2+2i)/4}, \qquad i = 0, 1, 2, \ldots$$

does not uniquely determine a single distribution as it was proven by Stieltjes [4], but the following set of probability density functions has the same moments for different values of λ:

$$w_\lambda(x) = \frac{x^{-\ln x}}{\sqrt[4]{\pi^2 e}} \left\{1 + \lambda \sin(2\pi \ln x)\right\}, \ -1 \leq \lambda \leq 1 \ .$$

Fig. 5 plots some of the possible $w_\lambda(x)$ functions along with the estimations. In this case the lower and upper bounds will not converge to each other, but to separate lower and upper limiting functions. The approximations do not significantly change after 10 moments and remain quite far from each other. The difference between them is 0.781576 at 1.96 with 18 moments.

The applied estimation method provides real upper and lower bounds also in this case. Indeed, the distance of the upper and lower bounds are determined by the information carried by the moments. In case of this moments series the moments does not carry enough information to reduce the distance of the bound to 0.

3) Bounded case: We experiment with a uniform distribution between 0 and 1.5 with weights 0.2 and 0.1 at the left and right endpoint, respectively. We choose this distribution to analyze the behavior of the estimation at the end points of the interval if the distribution function has jumps there.

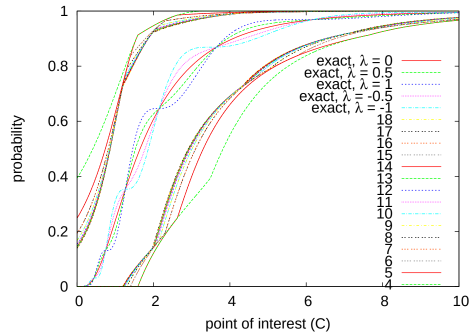

Fig. 5. Indefinite moment problem (positive)

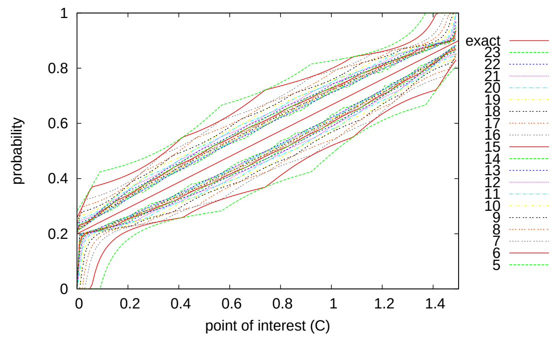

Fig. 6. Different number of moments used for estimation (bounded)

The experiences are similar to the positive case: the estimation using more moments results the same bounds in some intervals (Figure 6 and 7).

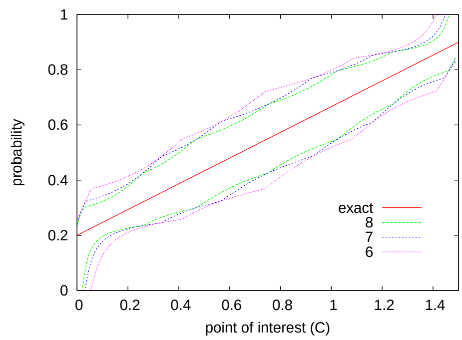

Fig. 7. Gain using more moments for estimation (bounded)

B. Comparison of the estimation methods

In this section we compare the methods when at least 2 different methods are applicable and we evaluate which one gives tighter bounds.

1) Distribution function supported on a semi-axis: Here we compare the infinite and positive estimation when they are applied to the same input moments originated from a distribution that is supported on a semi-axis. (The bounded estimation is *not* applicable in this case.) If we estimate a distribution that has no jump at a there are only slight differences between the two kinds of estimation (on some intervals the positive estimation gives tighter bounds, on others the limits are the same). The difference becomes significant when we estimate a distribution with a jump at a. For this reason we use the same sample distribution as in Section III-A.2.

Fig. 8 shows both estimations on the same graph using 19 moments. In some intervals the positive estimation gives tighter bounds, however relevant difference is only observable around 0, this is depicted in Fig. 9.

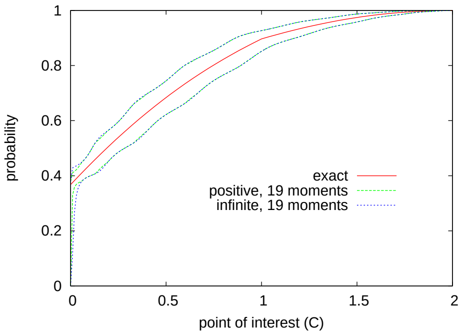

Fig. 8. Whole distribution function

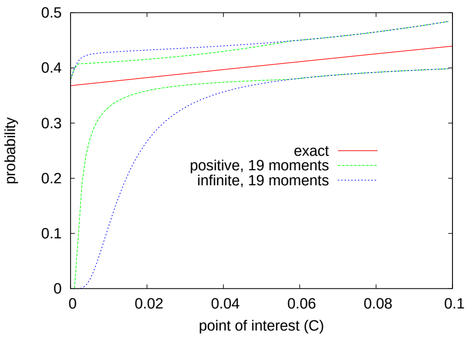

Fig. 9. Right endpoint of the interval

2) Distribution function supported on a bounded interval: We use the same distribution as in Section III-A.3. The comparison plots, Fig. 11 and 10, show only the end points of the interval, where relevant differences are observable. The bounded estimation gives the closest bounds at 1.5, because it can handle jumps at the end points of the support interval. The positive and infinite methods give the same bounds in the $[1.40186, 1.5]$ interval. At 0 both the positive and bounded estimations give better results than the infinite one, as it is expected.

IV. Conclusion

Using more moments never result in wider bounds, however the gain can be quite different along the support interval. It is also possible that we do not get closer bounds either. The applicability limit for the estimations is about 20 moments.

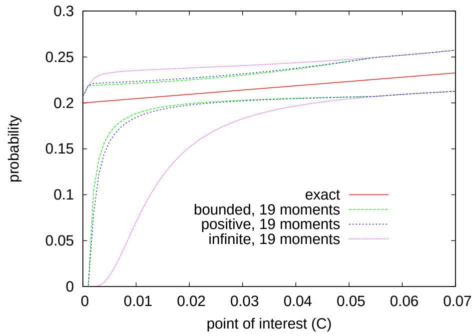

Fig. 10. Left endpoint of the bounded interval)

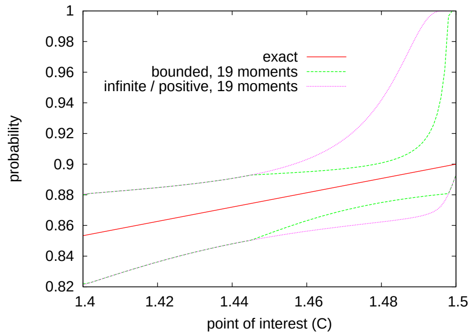

Fig. 11. Right endpoint of the bounded interval)

The gain in using even more moments is so small that it is not reasonable to do that and above this limit numerical problems likely occur.

Positive and bounded estimations have their advantage over the infinite estimation around the finite endpoint(s) of the support interval, especially when there is a nonzero probability mass there. However the infinite estimation proved to be numerically more stable. It is advisable to use it even in cases where the positive and/or bounded estimations are applicable, if the point of interest is "far" from the end point(s) of the support interval.

References

[1] W. Bryc. *Normal Distribution characterizations with applications*, volume 100 of *Lecture Notes in Statistics 1995*. 2005.

[2] V. G. Kulkarni, V. F. Nicola, and K. Trivedi. On modeling the performance and reliability of multi-mode computer systems. *The Journal of Systems and Software*, 6:175–183, 1986.

[3] 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.

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

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

[6] Á. Tari, M. Telek, and P. Buchholz. A unified approach to the moments based distribution estimation – Unbounded support. In 2^{nd} *European Performance Evaluation Workshop*, Versailles, France, 2005.

[7] Á. Tari, M. Telek, and S. Rácz. A matrix based solution of the moment problem on bounded and semi-infinite intervals. (to be published).