Approximating Distributions and Transient Probabilities by Matrix Exponential Distributions and Functions

(Article begins on next page)

25 April 2024

CHAPTER 1

APPROXIMATING DISTRIBUTIONS AND TRANSIENT PROBABILITIES BY MATRIX EXPONENTIAL DISTRIBUTIONS AND FUNCTIONS

András Horváth[1], Marco Paolieri[2], Enrico Vicario[2]

[1] Department of Computer Science - University of Turin - Italy
  horvath@di.unito.it
[2] Department of Information Engineering - University of Florence - Italy
  marco.paolieri@unifi.it, enrico.vicario@unifi.it

Abstract. A typical problem in modeling distributed systems is to fit measured data by an analytically tractable distribution. To this end phase type and matrix exponential distributions are often used not only because of their analytical tractability but also because they are easy to use as building blocks of more complex models. In this chapter we first give a brief introduction to these distributions and then provide a simple procedure to construct matrix exponential distributions based on the theory of Bernstein polynomials (BP). Namely, by a change of variable we derive from BP a family of functions, called in the sequel Bernstein expolynomials (BEs), which corresponds to a subset of the family of matrix exponential distributions. Furthermore, we show that BEs can be used to capture the transient probabilities of continuous time Markov chains (CTMC). They can provide closed-form approximations which are useful in the analysis of models where the process subordinated to a possibly non-Markovian period is described by a CTMC. The application of BEs for approximating both distributions and transient probabilities will be illustrated through several numerical examples.

1.1 Introduction

In stochastic modeling of distributed systems it is often necessary to approximate a random duration (which can be described by measured data) by selecting appropriately a member of a family of distributions. The choice of the family of the distribution usually must fulfill the following two conditions. First, the distributions must be composable, i.e., it must be possible to use the distributions as building blocks of a more complex model. Second, the distributions and the models composed of them must be tractable, i.e., they must lead to models whose analysis is feasible.

A well-known and widely-used family of distributions that provides the above two characteristics is the family of phase type (PH) distributions [28] that are defined as the distribution of time to absorption in a CTMC. Another family enjoying the two characteristics is that of the matrix exponential (ME) distributions. The ME family [26] is a proper superset of the class of PH distributions, it has the same algebraic form as PH distributions but does not have a simple probabilistic interpretation. It has been shown recently in [5] and [12] that many methods developed for models with PH distributions can be adopted to models with ME distributions.

Having chosen the family of distributions an appropriate member of the family must be chosen by applying a fitting procedure. Most of the existing techniques fall into two categories: those that are based on the maximum likelihood (ML) principle and those that aim to match moments. A third and much less explored possibility is to use a family of functions to fit the probability density function of the distribution we aim to approximate. This approach was used in [16] applying Jacobi polynomials and the main idea was to map the interval [-1,1], on which the Jacobi polynomials form an orthogonal class, to the interval $[0, \infty]$ by an appropriate change of variable. The resulting distributions are in the class of ME distributions.

In this work we show that also Bernstein polynomials (BP) can be used to construct ME distributions. These polynomials have a number of favorable properties but they are not feasible for distribution fitting in their standard form. We will show that by a change of variable it is possible to derive from them a family of functions which maintains the favorable features of BPs and provides the missing ones. We will refer to this family of functions as Bernstein expolynomials (BEs).

We can use BEs also to approximate transient probabilities of CTMCs which provides a useful ingredient for a recently proposed technique for the analysis of non-Markovian processes. This technique, called the method of stochastic state classes, is based on grouping into classes those states of the stochastic process that share a common future [13]. An extension of this approach has been proposed in [23] which takes advantage of subordinated Markovian periods and requires a closed-form description of the transient probabilities of the subordinated CTMCs.

The chapter is organised as follows. Section 1.2 provides a brief introduction to PH and ME distributions. In Section 1.3 we give a description of BPs and introduce

BEs. In Section 1.4 and Section 1.5 we illustrate the application of BEs to distribution fitting and to approximation of transient probabilities of CTMCs, respectively. Conclusions are drawn in Section 1.6.

1.2 Phase type and matrix exponential distributions

A continuous time random variable is said to have an order n Phase type distribution if it corresponds to the time to absorption of a CTMC with n transient states and a single absorbing state. Accordingly, a PH distribution is determined by the initial probabilities and the transition intensities of its absorbing CTMC. Figure 1.1 gives an example for the graphical representation of an order 3 PH distribution where the initial probabilities are written in the states and the transition intensities on the arcs. The initial probabilities and the infinitesimal generator of the CTMC depicted in Figure 1.1 are

$$\alpha = |0.5\ 0.2\ 0.3\ 0|, \ Q = \begin{vmatrix} -0.7 & 0.2 & 0.5 & 0 \\ 1.2 & -2.95 & 1.5 & 0.25 \\ 2 & 1 & -3.5 & 0.5 \\ 0 & 0 & 0 & 0 \end{vmatrix}$$

Since the entries of α have to sum to one and the sum of the entries of each row in Q must be equal to 0, the above representation is redundant. A non-redundant representation can be achieved by considering only the transient states. This representation for the above example is

$$\beta = |0.5\ 0.2\ 0.3|, \ T = \begin{vmatrix} -0.7 & 0.2 & 0.5 \\ 1.2 & -2.95 & 1.5 \\ 2 & 1 & -3.5 \end{vmatrix} \tag{1.1}$$

The cumulative distribution function of a PH random variable, X, represented by a vector-matrix pair (β, T) can be written as

$$F(x) = P\{X \le x\} = 1 - \beta e^{Tx} \mathbb{1} \tag{1.2}$$
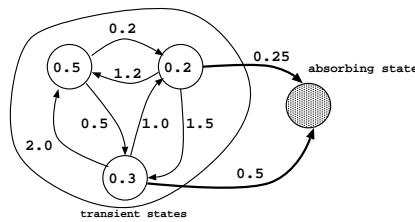
Figure 1.1   Graphical representation of an order 3 PH distributions

where $\mathbb{1}$ is a vector of ones. The rationale behind the above formula is that $\beta e^{Tx}\mathbb{1}$ is the probability that the Markov chain, started according to β and evolving according to T, is in one of the transient states at time x. The probability density function (pdf) is achieved by taking the derivative of $F(x)$ which leads to

$$f(x) = \beta e^{Tx}(-T)\mathbb{1} \tag{1.3}$$

By eigenvalue decomposition of the infinitesimal generator, the pdf and the cdf of a PH distribution can be written without taking the exponential of a matrix. The resulting form in general is

$$f(x) = \sum_{i=1}^{r} p_i(x)e^{\gamma_i} \tag{1.4}$$

where r is the number of the distinct eigenvalues of T, γ_i is the ith eigenvalue and $p_i(x)$ is a polynomial in x whose degree equals to the multiplicity of the ith eigenvalue minus one. There can be conjugate pairs of complex eigenvalues and their effect in the pdf is the presence of a term of type

$$p_i(x)e^{Re(\gamma_i)}\cos(\omega x + \phi)$$

For example, the pdf of the PH distribution represented graphically in Figure 1.1 is

$$0.0189382\,e^{-4.5455x} + 0.0803571\,e^{-2.5x} + 0.100705\,e^{-0.104499x}$$
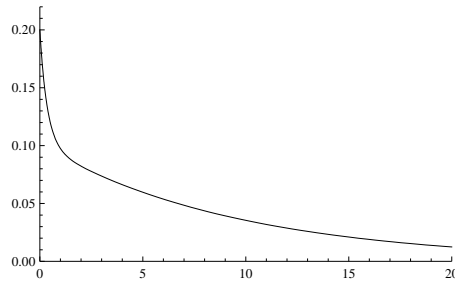
and we depicted it in Figure 1.2.

The Laplace transform of the pdf of a PH distribution is

$$f^*(s) = E\left[e^{-sX}\right] = \beta(sI - T)^{-1}(-T)\mathbb{1} \tag{1.5}$$

where I is the identity matrix. From the above Laplace transform it is straightforward to obtain the moments and we have

$$m_i = E\left[X^i\right] = i!\beta(-T)^{-i}\mathbb{1} \tag{1.6}$$

Figure 1.2   Pdf of the PH distribution represented in Figure 1.1

Some well-known and widespread used distributions are in the PH class. These include the exponential distribution with representation

$$\beta = |1|, \quad T = |-\lambda|$$

the hyper-exponential distributions whose representation is

$$\beta = |p_1, p_2, ..., p_n|, \quad T = \begin{vmatrix} -\lambda_1 & & & \\ & -\lambda_2 & & \\ & & \ddots & \\ & & & -\lambda_n \end{vmatrix}$$

and the order n Erlang distribution with representation

$$\beta = |1, 0, ..., 0|, \quad T = \begin{vmatrix} -\lambda & \lambda & & \\ & -\lambda & \lambda & \\ & & \ddots & \\ & & & -\lambda \end{vmatrix}$$

Matrix exponential (ME) distributions, such as PH distributions, are represented by a vector-matrix pair and have the same algebraic properties as PH distributions, i.e., equations (1.2-1.6) hold for ME distributions as well. The difference is that in case of ME distributions, the vector-matrix pair does not necessarily have the simple stochastic interpretation owned by PH distributions. Accordingly, the ME family is a proper superset of the PH family.

For both PH and ME distributions, the vector-matrix representation is not unique. Different representations can be obtained by similarity transforms. I.e., given a representation (β, T) and a matrix S such that $S\mathbb{1} = \mathbb{1}$, the vector-matrix pair $(\beta S^{-1}, STS^{-1})$ provides the same distribution as (β, T). For example, applying the similarity transform given by

$$S = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0.2 & 0.8 \end{vmatrix}$$

to the PH distribution given in (1.1), we obtain

$$\beta S^{-1} = |0.5 \ 0.125 \ 0.375|, \quad STS^{-1} = \begin{vmatrix} -0.7 & 0.075 & 0.625 \\ 1.2 & -3.325 & 1.875 \\ 1.84 & 0.835 & -3.125 \end{vmatrix}$$

which is still a Markovian representation, i.e., the above vector-matrix pair is clearly in the PH family. By applying

$$S = \begin{vmatrix} 0.5 & 0.5 & 0 \\ 0 & 1 & 0 \\ 0 & 0.5 & 0.5 \end{vmatrix}$$

in the similarity transform, we obtain a third representation

$$\beta S^{-1} = |1 \ -0.6 \ 0.6|, \ STS^{-1} = \begin{vmatrix} 0.5 & -2.625 & 2. \\ 2.4 & -5.65 & 3. \\ 3.2 & -1.575 & -2. \end{vmatrix} \tag{1.7}$$

which is not a Markovian representation anymore (it has a negative value in the vector and negative off-diagonal entries in the matrix). Thus one might think that the vector-matrix pair given in (1.7) does not represent a PH distribution. This is not true however because there exists a Markovian vector-matrix pair that produces the very same distribution. The problem of searching for an equivalent Markovian pair given a non-Markovian one was studied in [34].

For almost all order n ME distribution there exists an equivalent order m PH distribution with $n \le m < \infty$ [4]. Exceptions are those ME distributions whose pdf is zero infinitely many times. An example for this case is given by the following vector-matrix pair [11]

$$\beta = |1.04865, -0.0340166, -0.0146293|, \ T = \begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & -20 \\ 0 & 20 & -1 \end{vmatrix}$$

whose corresponding pdf is

$$1.04865 \, e^{-x} \cos(20x + 2) + 1.04865 \, e^{-x}$$

and it is depicted in Figure 1.3. Such pdf can be realized with a PH distribution only with an infinite number of phases. Another feature in which ME distributions outperform PH distributions is the capacity of having low variability. The lowest coefficient of variation of an order n PH distribution is $1/n$ exhibited by the order n Erlang distribution [3]. No such simple result is known for the ME class but it is known that the minimal coefficient of an order n $(n > 2)$ ME is lower than that of an order n PH distribution.

The class of PH distributions and, consequently, the ME class are dense in the field of positive valued distributions. This implies that any positive valued distribution can be approximated by them with any accuracy. This fact does not provide however directly a practical method to fit distributions by PH or ME distributions. Several authors proposed fitting methods and most of these fall into two categories:
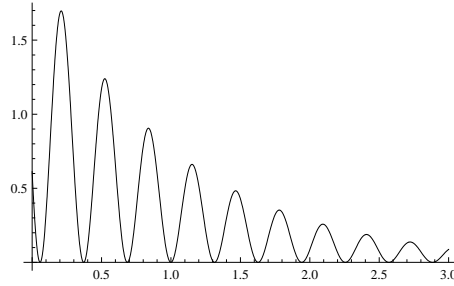
Figure 1.3   ME pdf touching the x-axis infinitely many times

maximum likelihood (ML) based estimation of the parameters and moment matching techniques.

One of the first works on ML estimation considered acyclic PH distributions is [7] (i.e., PH distributions whose underlying Markov chain is acyclic) while an approach for the whole family, based on the expectation-maximization method, is proposed in [30]. Since these early papers, many new methods and improvements have been suggested for the whole PH family and for its sub-classes (see, e.g., [35, 29]). Much less research tackled ML based fitting of ME distributions because of the lack of a practical stochastic interpretation. One such method, based on semi-infinite programming, is described however in Chapter 9 of [18] where the computational complexity of the problem is discussed and an algorithm is devised.

For what concerns moment matching methods the following results are available. For low order (≤ 3) PH and ME distributions moment bounds and moment matching formulas are either known in an explicit manner or there exist iterative numerical methods to check if given moments are possible to capture [33, 24, 20]. For higher order there exist matching algorithms, but these often result in improper density functions and the validity check is a non-trivial problem [18, 25]. In [8] a simple method is provided that constructs a minimal order acyclic PH distribution given three moments. Characterization of moments of PH and ME distributions is discussed in [10]. Moreover, tool support is available for the construction of PH and ME distributions. Specifically, ML based fitting is implemented in PhFit [22] and a set of moment matching functions is provided in BuTools [1].

The main reason to apply PH and ME distributions in stochastic modeling is that these distributions can be easily used as building blocks of more complex models. This fact is well-known for what concerns PH distributions because it is straightforward that if we are given a system in which all sojourn times are according to PH distributions and the next state distribution is Markovian then the overall system behavior can be described by a Markov chain. Starting from [15], the construction of the Markov chain, often referred to as the expanded Markov chain, was proposed in the literature for several modeling formalisms, such as stochastic Petri nets (SPN), stochastic process algebras or stochastic automata networks [2, 17, 31]. The down-

side of dealing with the expanded chain is that if the model has many states and/or it describes many activities performed in parallel, then the expanded chain can have a huge number of states. To alleviate this problem several authors proposed techniques for the compact representation of the expanded chain. Such techniques are either based on Kronecker-algebra (among the first works see, e.g., [32]) or, more recently, on decision diagrams techniques (e.g., [14]).

The necessary theoretical background to use ME distributions as building blocks was developed instead only recently. In this case the overall model is not a Markov chain, but the transient system behavior can still be described by a set of ordinary differential equations. In the context of SPN this was shown for a subclass of ME distributions in [12] while in [11] the result was extended to the whole family. In the context of quasi-birth-and-death processes the possibility of using ME distributions was investigated in [6].

1.3 Bernstein polynomials and expolynomials

The degree-n Bernstein polynomial [27] approximating a function f on $[a, b]$ is defined as

$$B_n(x) = \sum_{i=0}^{n} f\left(a + \frac{i}{n}(b-a)\right) \binom{n}{i} \frac{(x-a)^i(b-x)^{n-i}}{(b-a)^n} \qquad (1.8)$$

Bernstein polynomials exhibit a series of favorable properties which motivates its use in distribution approximation:

- *Globality:* the approximant is global in the sense that a single polynomial is used to approximate the whole interval $[a, b]$.

- *Positivity:* since all the polynomials $(x-a)^i(b-x)^{n-i}, i = 0, 1, \ldots, n$, are positive in the interval $[a, b]$ and we assume that also the function f is positive (in our case f is a pdf), the polynomial in (1.8) is positive on $[a, b]$.

- *Simplicity of derivation:* a BP is determined in a straightforward manner by the samples of the approximated function without any optimisation process.

- *Convergence of the approximation:* if the approximated function is continuous then the degree-n BP converges uniformly to it as n is increased. Moreover, the approximation error can be bounded by a Lipschitz inequality [27].

Conversely, the following requirements are not guaranteed by BPs but can be accommodated by a suitable adaptation.

- *Divergence of the approximant:* a BP is either a constant function or it diverges to plus or minus infinity. For this reason it cannot be used to approximate infinite support distributions and it is not feasible for the approximation of transient probabilities of CTMCs if the required interval overlaps with the steady state regime.

- *Unit-measure:* BP approximation does not guarantee that the approximant preserves the integral of the approximated function. This is clearly a problem when fitting distributions because the approximant is expected to have unit-measure.

The last problem can be managed simply by normalising the approximant. The problem of divergence can be solved instead by applying the change of variable $x \rightarrow e^{-x}$ to (1.8) resulting in the following sum of exponentials with negative exponents

$$
E_n(x) = \sum_{i=0}^{n} f\left(-\log\left(e^{-a} - \frac{i}{n}\left(e^{-a} - e^{-b}\right)\right)\right) \times
$$
$$
\binom{n}{i} \frac{(e^{-a} - e^{-x})^i (e^{-x} - e^{-b})^{n-i}}{(e^{-a} - e^{-b})^n} \tag{1.9}
$$

which converges as x tends to infinity. The functions defined in (1.9), which we call Bernstein expolynomials, maintain the favourable properties of BPs. It is easy to see that a function in the form of (1.9) belongs to the ME family whose general form is given in (1.4).

Both BPs and BEs catch exactly the approximated function at the limits of the considered interval, $[a, b]$. The important difference between BPs and BEs is that the latter can be used to approximate a function on the interval $[0, \infty]$. Applying (1.9) with $b = \infty$ the limiting value of the BE will catch exactly the limiting value of the approximated function.

Another difference between the two approximations is the way they sample the approximated function. BPs result in an equidistant sampling of the considered interval, i.e., the value of the approximated function is taken at the values $a + \frac{i}{n}(b-a), i = 0, 1, \ldots, n$. In order to consider the sampling applied by the expolynomial approximation, let us assume $a = 0, b = \infty$. In this case the approximated function is sampled at the values $-\log(i/n), i = n, n - 1, \ldots, 0$. This implies that more samples are taken from the beginning of the considered interval. This effect can be mitigated by approximating a scaled version of the function, namely $f(x/c)$, and bringing back then the approximant to the original scale. Without going into details, we illustrate this effect in two figures. In Fig. 1.4 we depict for different values of c, assuming that $n = 10$ and the interval to be approximated is [0,10], the positions where the value of the approximated function, $f(x)$ itself, is taken. The larger c, the closer the sampling points are to a (for except the last sampling point which is always at b). Instead, as c tends to 0 the sampling tends to be equidistant. In Fig. 1.5 the effect of c for $n = 10, a = 0, b = \infty$ is shown.

When the functions in (1.9) are applied to fit the probability density function (pdf) of a distribution, the interval of the approximation must be set to [0,∞]. This guarantees that the limiting value of the approximant is 0 which is a necessary condition for a pdf. It is easy to check that the approximating pdf is in the class of ME distributions.
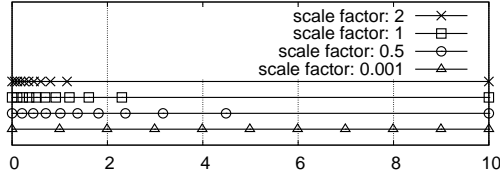
Figure 1.4 Sampling points for different scale factors if $n = 10$ and the interval to fit is [0,10].
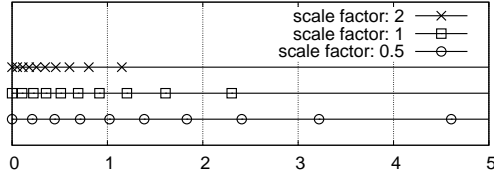
Figure 1.5 Sampling points for different scale factors if $n = 10$ and the interval to fit is [0,∞] (also the value of the function at infinity is used).

1.4 Application of BEs to distribution fitting

In this section we illustrate the application of BE to distribution fitting through a benchmark defined in [9] for the assessment of acyclic PH distributions. The benchmark is composed of 9 functions that assume relevance in reliability analysis and quantitative evaluation problems. The 9 distributions are the following (in parenthesis we give the abbreviation with which we refer to them in the rest of the paper).

- Weibull: $f(x) = \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}} \begin{cases} \eta = 1, \beta = 1.5 & \text{(W1)} \\ \eta = 1, \beta = 0.5 & \text{(W2)} \end{cases}$

- Lognormal: $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \begin{cases} \mu = 1, \sigma = 1.8 & \text{(L1)} \\ \mu = 1, \sigma = 0.8 & \text{(L2)} \\ \mu = 1, \sigma = 0.2 & \text{(L3)} \end{cases}$

- Uniform on $[0, 1]$, (U1)

- Uniform on $[1, 2]$, (U2)

- Shifted Exponential: $f(x) = \frac{1}{2}e^{-x} + \frac{1}{2}e^{-(x-1)} \cdot I(x \geq 1)$, (SE)

- Matrix Exponential: $f(x) = \left(1 + \frac{1}{(2\pi)^2}\right)(1 - cos(2\pi x))e^{-x}$, (ME)

We will consider the following measures, used also in [9], to evaluate the goodness of fit achieved by the BE approximations.

	W1	W2	L1	L2	L3	U1	U2	SE	ME
$c_1(F)$	0.903	2	13.7	3.74	2.77	0.5	1.5	1.5	1.045
$c_2(F)$	0.376	20	4629	12.6	0.313	0.08\dot{3}	0.08\dot{3}	1.25	0.953
$c_3(F)$	0.246	592	4.29×10^7	164	0.108	0	0	2	1.99

Table 1.1   Mean, variance and third central moment of the distributions of the benchmark.

- visual appearance of density plots,

- relative error of the mean: $\hat{e}_1 = |c_1(\hat{F}) - c_1(F)|/c_1(F)$,

- relative error of the variance: $\hat{e}_2 = |c_2(\hat{F}) - c_2(F)|/c_2(F)$,

- relative error of the third central moment: $\hat{e}_3 = |c_3(\hat{F}) - c_3(F)|/c_3(F)$,

where $c_1(F)$ and $c_1(\hat{F})$ denote the mean of the original distribution and the mean of the approximant, respectively, and $c_i(F)$ and $c_i(\hat{F})$ are the ith central moment of the original distribution and that of the approximant. Table 1.1 provides $c_1(F), c_2(F)$ and $c_3(F)$ for all the distributions of the benchmark.

Figures 1.6-1.10 depict the pdf of the distributions and the pdf of the approximations for various degrees of the BE. For all the experiments we have chosen such scaling that the number of samples that fall before the median of the distribution is the same as the number of samples that fall after the median. Moreover, as the mean is of crucial importance in almost all modeling situations, we modified the approximants in such a way that its mean is equal to the mean of the original distribution. This can be done easily by scaling the approximant by an appropriate constant. Consequently, for all approximants $\hat{e}_1 = 0$. By visual inspection the following considerations can be made.

The approximant is not good for those distributions whose tail is heavier then exponential (distribution W2 and L1). These two distributions are hard to approximate by general purpose PH fitting methods and can be handled only by methods that pay special attention to the tail [19, 21]. In case of distribution L1 we obtain an approximant that is particularly far from the original function. This is due to the adjustment of the mean. Without the adjustment, the mean of the BE approximant is far from the mean of L1 but visually it looks better than the one depicted in Figure 1.7. Adjusting the mean makes the visual impression worse.

The distributions with low coefficient of variation (L1, U1 and U2) requires a high degree approximation to capture their low variability.

The distributions with sharp changes are in general hard to approximate by PH distribution and, consequently, low degree BE approximations do not provide good approximants.

In most of the cases, as it is foreseen by the theory of BP, increasing the degree of the approximant results in better fit.

Table 1.2 reports error measures obtained applying BE approximation of degree 10, 20 and 100 to each function of the benchmark. For distributions U1 and U3, the
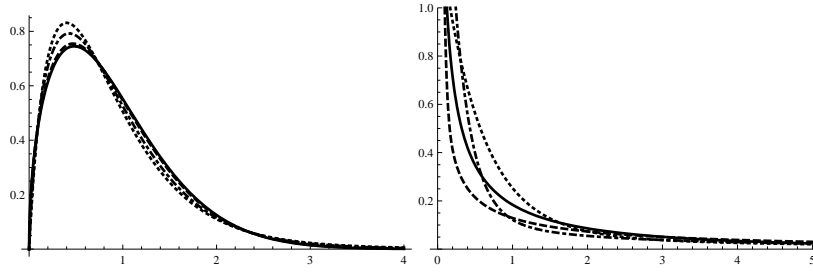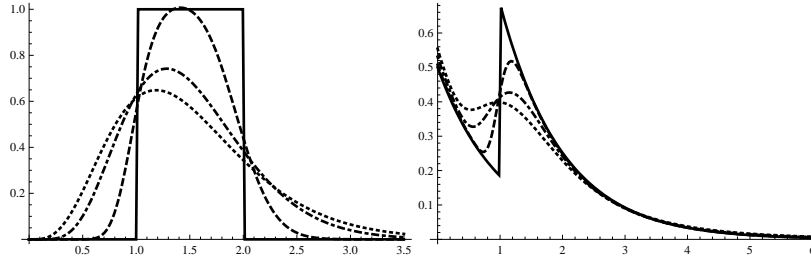
Figure 1.6 BE approximation of distributions W1 (left) and W2 (right) with degree 10 (dotted), 20 (dot-dashed) and 100 (dashed).
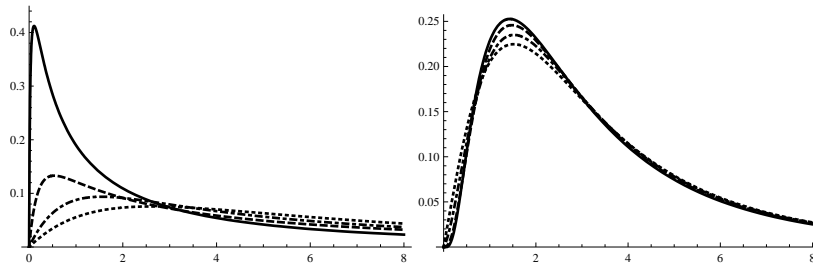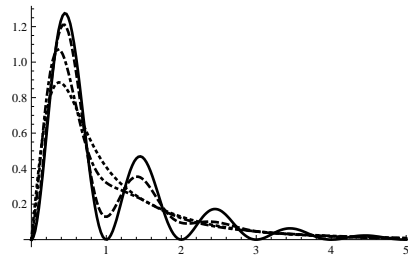

Figure 1.7 BE approximation of distributions L1 (left) and L2 (right) with degree 10 (dotted), 20 (dot-dashed) and 100 (dashed).
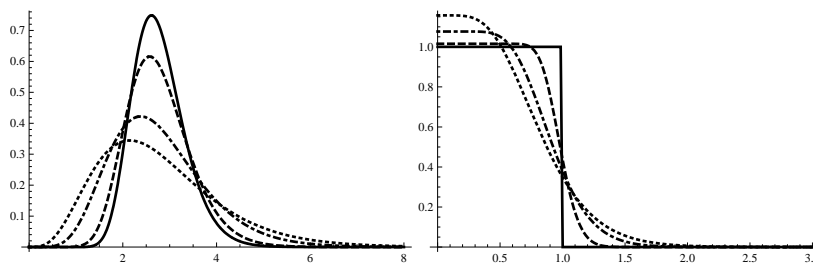
Figure 1.8 BE approximation of distributions L3 (left) and U1 (right) with degree 10 (dotted), 20 (dot-dashed) and 100 (dashed).

Figure 1.9 BE approximation of distributions U2 (left) and SE (right) with degree 10 (dotted), 20 (dot-dashed) and 100 (dashed).

Figure 1.10 BE approximation of distributions ME with degree 10 (dotted), 20 (dot-dashed) and 100 (dashed).

third central moment, $c_3(F)$, is 0 and for this reason the tables report the third central moment of the approximants, $c_3(\hat{F})$, instead of the relative error.

As already observed, BE approximation suffers heavy tails: in case of distributions W2 and L1 there are large relative errors of the variance and of the third central moment and even increasing the degree of the approximation does not decrease the error.

In case of low coefficient of variation (distributions L3, U1 and U2), low degree approximations are not able to capture the variance of the original distribution but increasing the degree improves the fitting. Note that in case of distributions U1 and U2, the error of the third central moment is low even if the relative error of the variance is large.

It is interesting to note that visual inspection and numerical error measures can contradict each other as it happens in case of distribution ME for which the degree 100 approximation seems much better than the degree 10 approximation but the relative errors of the moments tell the opposite.

Now we give a brief comparison between the Jacobi polynomial based method presented in [16] and the one proposed in this paper. As already mentioned, the method of [16] requires the setting of four parameters for which no guidelines are presented. For distribution U2 the author of [16] provides a good setting of the four parameters and for this reason we use U2 in the comparison. In Figure 1.11 we depicted the approximations obtained by the two methods presented in [16]. Method

	$\hat{e}_2$	$\hat{e}_3$		$\hat{e}_2$	$\hat{e}_3$		$\hat{e}_2$	$\hat{e}_3$
$W1_{10}$	0.3500	1.9312	$L2_{10}$	0.1522	0.8746	$U2_{10}$	4.5881	0.1510
$W1_{20}$	0.2080	1.0049	$L2_{20}$	0.1593	0.8740	$U2_{20}$	2.8260	0.0602
$W1_{100}$	0.0458	0.0936	$L2_{100}$	0.1423	0.8646	$U2_{100}$	0.5303	0.0062
$W2_{10}$	0.2877	0.9346	$L3_{10}$	4.4687	4.8666	SE_{10}	0.3343	0.4270
$W2_{20}$	0.2606	0.9299	$L3_{20}$	2.3452	1.1749	SE_{20}	0.2164	0.2205
$W2_{100}$	0.4333	0.9423	$L3_{100}$	0.4878	0.5199	SE_{100}	0.0703	0.0950
$L1_{10}$	0.9554	0.9999	$U1_{10}$	0.5569	0.0576	ME_{10}	0.0270	0.1490
$L1_{20}$	0.9476	0.9999	$U1_{20}$	0.2906	0.0299	ME_{20}	0.1926	0.2016
$L1_{100}$	0.9382	0.9999	$U1_{100}$	0.0615	0.0055	ME_{100}	0.0357	0.1788

Table 1.2 Relative errors of the variance and of the third central moment of BE approximations of degree 10, 20 and 100.

Jacobi method A   Jacobi method B
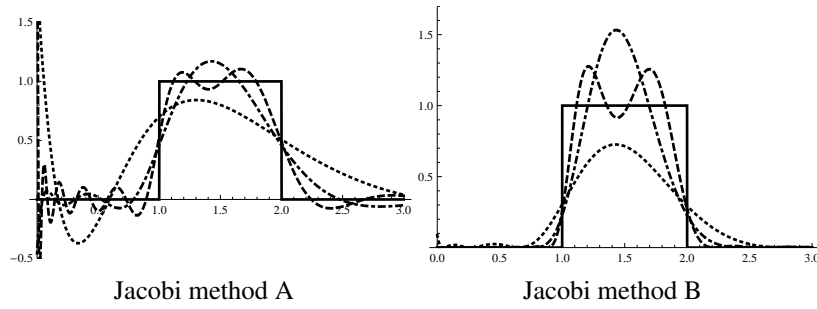
Figure 1.11 Approximation of U2 by the methods proposed in [16] with degrees 5 (dotted), 10 (dot-dashed) and 20 (dashed) on the left and with degrees 10 (dotted), 20 (dot-dashed) and 40 (dashed) on the right (the squaring applied by method B doubles the degree of the approximation).

A does not guarantee positivity and the pdfs given by this method have negative values. Moreover, an oscillating behaviour can be observed around 0. Method B provides proper pdfs and gives better approximation than the BE approximant of the same degree. The drawback however of the Jacobi polynomial based method is that it requires heavy numerical computations and involves numerical problems. Indeed, we were not able to compute the Jacobi polynomial based approximant for degrees higher than 20. Moreover, the computation of the approximant of degree 20 took about 2 minutes while the construction of the BE approximant is immediate.

1.5 Application of BEs to transient probabilities

In this section we briefly experiment the approximation of transient probabilities of CTMCs by BEs. This can be useful when a distributed computing system is

composed of many loosely interacting sub-modules and the BE based representation of the transient probabilities allows to use a compact representation of the behavior of a building block and thus reduce the state space of the overall model. It is also useful when the theory of stochastic state classes [13, 23] is used where it is necessary to have a closed-form approximation of the transient behaviour of the subordinated CTMCs that are present in the model

We define the CTMC under study by a Petri net which is depicted in Fig. 1.12. In the net we have three non-exponential transitions, $t0$, $t1$ and $t2$, while the other transitions are with exponential firing time with parameter $1/2$. The exponential transitions give rise to a CTMC which is subordinated to the activity period of the non-exponential transitions.

This CTMC describes the interaction of senders and receivers exchanging messages through a buffer that may fail and be repaired. After each successful receive action (transition te), a choice is made between transition tf which restarts a new receive cycle and transition $tout$ by which the receiver leaves. We start the model with 4 tokens in place pa, 4 tokens in place $free$ and 4 tokens in place pd. With this initial marking the subordinated CTMC has 1218 states.

In Fig. 1.13-1.18 we have depicted some transient probabilities of the subordinated CTMC and the corresponding approximations for various degrees. We denote by $[\exp(tQ)]_{i,j}$ the probability of being in marking j at time t assuming that the chain started in marking i. The considered markings are: marking 1 which is the already described initial marking, marking 2 where the token distribution (not listing empty places) is $\#pa = 3, \#pb = 1, \#free = 3, \#pd = 4$, marking 10 with $\#pb = 4, \#pd = 4$ and marking 1000 with $\#pa = 3, \#pc = 1, \#pe = 2, \#pf = 2, \#failed = 1, \#busy = 1$. The corresponding transient probabilities cover a wide range of cases. To illustrate the effect of changing the sampling pattern by scaling, in Fig. 1.14, 1.16 and 1.17 we give approximations for two different scalings. In particular, $[\exp(tQ)]_{1,1}$ is approximated well by a degree-1 BE (one exponential term plus a constant) if the scaling is chosen well. It is worth to observe that BEs are able to catch the probabilities in the steady state regime as well.

1.6 Conclusions

In this chapter we have provided an introduction of the family of PH and ME distributions. Furthermore, we have proposed the family of Bernstein expolynomials for the construction of ME distributions. We have tested the approach fitting a benchmark of distributions. In a similar manner, Bernstein expolynomials can be used to approximate the transient probabilities of CTMCs and they can provide compact, closed-form expressions. The computational cost of the approximation is extremely low.

The following aspects have not been covered and will be included in a subsequent research paper. The goodness of the approximations has to be verified both through statistical measures (like moments) and by plugging them into applications. Bernstein polynomial and Bernstein expolynomial can be applied to multivariate func-
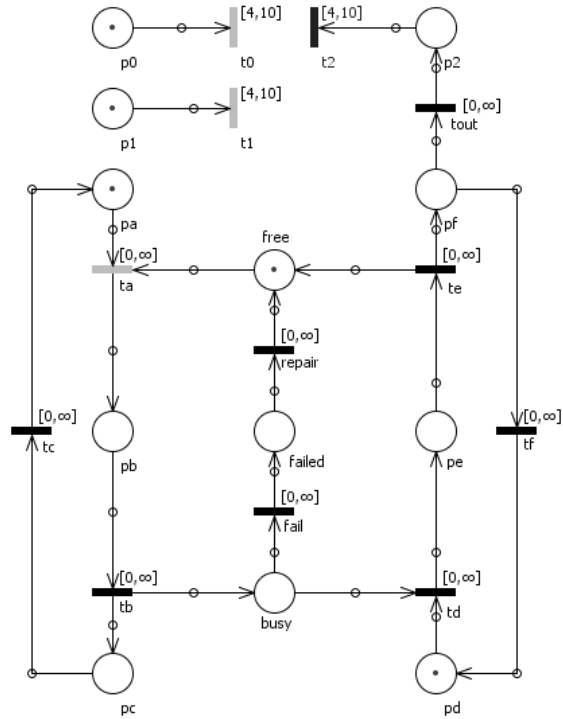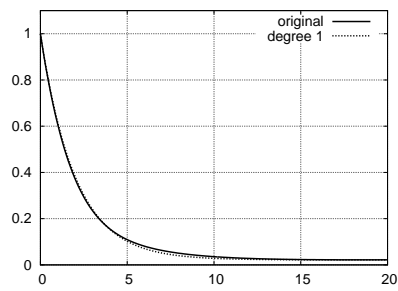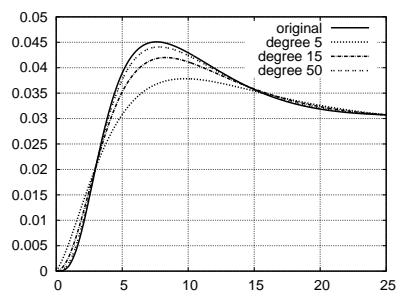
Figure 1.12   Petri net whose subordinated CTMC is considered for experimenting BEs
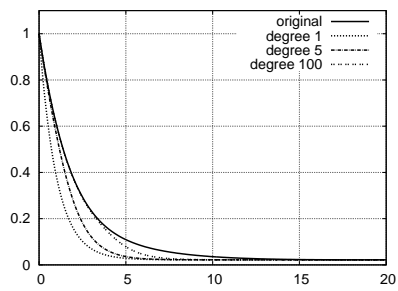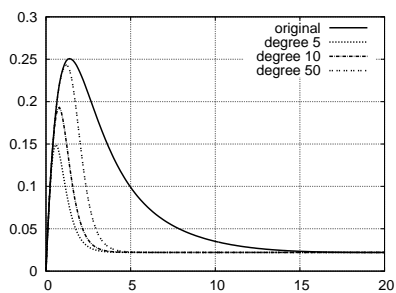

Figure 1.13   Fitting $[\exp(tQ)]_{1,1}$ without scaling
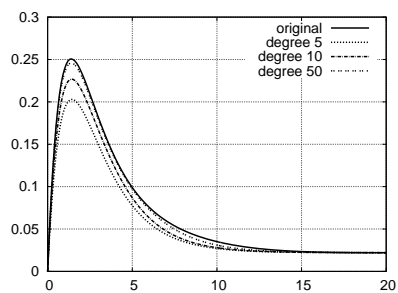
Figure 1.14   Fitting $[\exp(tQ)]_{1,1}$ with convenient scaling

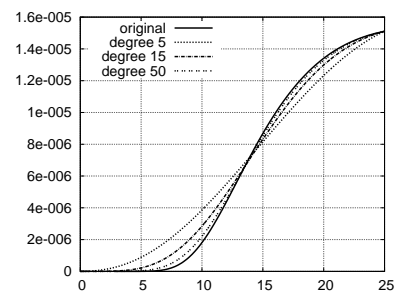**Figure 1.15** Fitting $[\exp(tQ)]_{1,10}$ with scalings that are convenient for the whole interval

Figure 1.16 Fitting $[\exp(tQ)]_{1,2}$ with a scaling that concentrates sampling points close to 0

Figure 1.17 Fitting $[\exp(tQ)]_{1,2}$ with a scaling that is convenient for the whole interval

Figure 1.18 Fitting $[\exp(tQ)]_{1,1000}$ with scalings that are convenient for the whole interval

tions as well and this can be necessary when the approximations are applied in the theory of stochastic state classes. For distributions, Bernstein expolynomial approximations result in ME distributions with a fixed pole-structure. This restriction can be mitigated by extending the class of Bernstein expolynomials but the approximations then might require numerical optimisation.

REFERENCES

1. *BuTools*. http://webspn.hit.bme.hu/~telek/tools/butools.

2. M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. *Modelling with Generalized Stochastic Petri Nets*. John Wiley & Sons, 1995.

3. D. Aldous and L. Shepp. The least variable phase type distribution is Erlang. *Stochastic Models*, 3(3):467–473, 1987.

4. S. Asmussen and C. A. O'Cinneide. Matrix-exponential distributions - distributions with a rational Laplace transform. In S. Kotz and C. Read, editors, *Encyclopedia of Statistical Sciences*, pages 435–440. John Wiley and Sons, 1997.

5. N.G. Bean and B.F. Nielsen. Quasi-birth-and-death processes with rational arrival process components. Technical Report IMM-Technical Report-2007-20, Technical University of Denmark, 2007.

6. N.G. Bean and B.F. Nielsen. Quasi-birth-and-death processes with rational arrival process components. *Stochastic Models*, 26(3):309–334, 2010.

7. A. Bobbio and A. Cumani. ML estimation of the parameters of a ph distri- butions in triangular canonical form. *Computer Performance Evaluation*, pages 33–46, 1992.

8. A. Bobbio, A. Horváth, and M. Telek. Matching three moments with minimal acyclic phase type distributions. *Stochastic Models*, 21:303–326, 2005.

9. A. Bobbio and M. Telek. A benchmark for PH estimation algorithms: results for Acyclic-PH. *Stochastic Models*, 10:661–677, 1994.

10. L. Bodrog, A. Horváth, and M. Telek. Moment characterization of matrix exponential and Markovian arrival processes. *Annals of Operations Research*, 160:51–68, 2008.

11. P. Buchholz, A. Horváth, and M. Telek. Stochastic Petri nets with low variation matrix exponentially distributed firing times. *Int. Journal of Performability Engineering*, 7(5):441–454, 2011.

12. P. Buchholz and M. Telek. Stochastic Petri nets with matrix exponentially distributed firing times. *Performance Evaluation*, 67:1373–1385, 2010.

13. L. Carnevali, L. Grassi, and E. Vicario. State-density functions over DBM domains in the analysis of non-Markovian models. *IEEE Trans. on Soft. Eng.*, 35:178 – 194, March-April 2009.

14. G. Ciardo and A. S. Miner. A data structure for the efficient kronecker solution of GSPNs. In *Proc of the 8th International Workshop on Petri Nets Performance Models (PNPM 99)*, 1999.

15. A. Cumani. ESP - a package for the evaluation of stochastic Petri nets with Phase-type distributed transition times. In *Proc. of the International Workshop on Timed Patri Nets*, pages 144–151, Torino, Italy, July 1985.

16. D. Dufresne. Fitting combinations of exponentials to probability distributions. *Applied Stochastic Models in Business and Industry*, 23(1):23–48, 2007.

17. A. El-Rayes, M. Kwiatkowska, and G. Norman. Solving infinite stochastic process algebra models through matrix-geometric methods. In J. Hillston and M. Silva, editors, *Proc. 7th Process Algebras and Performance Modelling Workshop (PAPM'99)*, pages 41–62. University of Zaragoza, 1999.

18. M. Fackrell. *Characterization of Matrix-exponential Distributions*. PhD thesis, Faculty of Engineering, Computer and Mathematical Sciences, University of Adelaide, 2003.

19. A. Feldman and W. Whitt. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*, 31:245–279, 1998.

20. A. Horváth, S. Rácz, and M. Telek. Moments characterization of order 3 matrix exponential distributions. In *Proc. of the 16th International Conference on Analytical and Stochastic Modeling Techniques and Applications (ASMTA) 2009*, volume 5513 of *Lecture Notes in Control and Information Sciences*, pages 174–188, Madrid, Spain, June 2009.

21. A. Horváth and M. Telek. Approximating heavy tailed behavior with Phase-type distributions. In *Proc. of 3rd International Conference on Matrix-Analytic Methods in Stochastic models*, Leuven, Belgium, June 2000.

22. A. Horváth and M. Telek. Phfit: A general phase-type fitting tool. In *Proc. of 12th Performance TOOLS*, volume 2324 of *Lecture Notes in Computer Science*, Imperial College, London, April 2002.

23. A. Horváth and E. Vicario. Aggregated stochastic state classes in quantitative evaluation of non-markovian stochastic Petri nets. In *Proc. of 6th International Conference on the Quantitative Evaluation of Systems (QEST)*, Budapest, Hungary, Sept 2009.

24. G. Horváth and M. Telek. A canonical representation of order 3 phase type distributions. *Lecture Notes in Theoretical Computer Science*, 4748:48–62, 2007.

25. G. Horváth and M. Telek. A minimal representation of Markov arrival processes and a moments matching method. *Performance Evaluation*, 64:1153–1168, 2007.

26. L. Lipsky. *Queueing Theory: A Linear Algebraic Approach*. Springer, 2008.

27. G. Lorentz. *Bernstein Polynomials*. University of Toronto Press, 1953.

28. M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approachg*. Dover, 1981.

29. A. Panchenko and A. Thümmler. Efficient phase-type fitting with aggregated traffic traces. *Performance Evaluation*, 64(7-8):629–645, 2007.

30. O. Nerman S. Asmussen and M. Olsson. Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23(4):419–441, 1996.

31. I. Sbeity, L. Brenner, B. Plateau, and W.J. Stewart. Phase-type distributions in stochastic automata networks. *European Journal of Operational Research*, 186(3):1008 – 1028, 2008.

32. M. Scarpa and A. Bobbio. Kronecker representation of stochastic Petri nets with discrete PH distributions. In *Proc. of International Computer Performance and Dependability Symposium - IPDS98*, page 52V61, 1998.

33. M. Telek and A. Heindl. Matching moments for acyclic discrete and continuous phase-type distributions of second order. *International Journal of Simulation Systems, Science & Technology*, 3(3-4), 2002. Special Issue on: Analytical & Stochastic Modelling Techniques.

34. M. Telek and G. Horváth. A minimal representation of markov arrival processes and a moments matching method. *Performance Evaluation*, 64(9–12):1153–1168, 2007.

35. A. Thümmler, P. Buchholz, and M. Telek. A novel approach for fitting probability distributions to trace data with the EM algorithm. In *Proc. of the International Conference on Dependable Systems and Networks (DSN)*, pages 712–721, June 2005.