Approximation of Cumulative Distribution Functions by Bernstein Phase-Type Distributions

András Horváth[1], Illés Horváth[2,3], Marco Paolieri[4], Miklós Telek[2,3], and Enrico Vicario[5]

[1] Department of Computer Science, University of Turin
[2] HUN-REN-BME Information Systems Research Group
[3] Department of Networked Systems and Services, Budapest University of Technology and Economics
[4] Department of Computer Science, University of Southern California
[5] Department of Information Engineering, University of Florence

Abstract. The inclusion of generally distributed random variables in stochastic models is often tackled by choosing a parametric family of distributions and applying fitting algorithms to find appropriate parameters. A recent paper proposed the approximation of probability density functions (PDFs) by Bernstein exponentials, which are obtained from Bernstein polynomials by a change of variable and result in a particular case of acyclic phase-type distributions. In this paper, we show that this approximation can also be applied to cumulative distribution functions (CDFs), which enjoys advantageous properties; by focusing on CDFs, we propose an approach to obtain stochastically ordered approximations.

Keywords: Bernstein polynomials; phase-type distributions; Markov chains; analytic approximation.

1 Introduction

Continuous-time models of stochastic systems frequently need to include random variables with general (i.e., non-exponential) probability distributions, to represent properties enforced by design (e.g., periodic releases or deterministic timeouts) or by contract (e.g., service times guaranteed along the development process or by some agreed service level objective), or to fit observed data or learned parameters. A standard approach is to select a parametric family of probability distributions and to apply fitting algorithms to find appropriate parameter values to approximate the observed random variables. An ideal family of distributions should be sufficiently general to result in accurate approximations, but it should also support simple fitting procedures and allow efficient analysis (or simulation) of the resulting system model.

The family of phase-type (PH) distributions [11], defined as the time to absorption in Markov chains, is broadly used to approximate general random variables. By varying the number of phases in the Markov chain, this family allows a tradeoff between accuracy of the approximation and analysis cost of the resulting model. PH parameter fitting methods include maximum likelihood methods [2,4], moment matching [5,16], tail behavior matching [6][14], or both [8]. Stochastic models with PH type distributions result in underlying Markov chains with regular structures, which can be analyzed efficiently through matrix analytic methods [12,10].

In [9], an approach was proposed to approximate probability density functions (PDFs) using *Bernstein exponentials* (BE), i.e., linear combinations of Bernstein polynomials (BP) [13] [15] where the support $[0,1]$ is mapped to $[0,\infty)$ through a change of variable. This approach results in a subclass of acyclic PH distributions that preserve shape properties of approximated density functions and enjoy derivation simplicity as BE parameters can be derived in closed form, while allowing efficient model analysis.

In this paper, we study the properties of BE approximations for cumulative distribution functions (CDFs) rather than PDFs. In fact, BE approximations guarantee uniform convergence and preserve local shape properties, notably including non-negativity, but they do not preserve integral measure and thus require normalization to obtain valid PDFs with unitary measure. Conversely, we show that, when BE approximations are applied to CDFs, or to complementary CDFs (CCDFs), the resulting functions are valid cumulative distributions that belong to a subclass of acyclic PH distributions and preserve the important properties of the original CDFs including monotonicity, upper and lower bounds, and exact limit values at 0 and $+\infty$. In particular, we focus on stochastic order for models designed for the evaluation of safe guarantees of system metrics (e.g., quality of service) [7][3]. We present an approach to obtain BE approximations that guarantee smaller and greater stochastic order, and characterize the required tail conditions and minimum degree of the BE approximation.

The paper is organized as follows. In Section 2, we recall background information on Bernstein polynomials, Bernstein exponentials, and PH distributions. In Section 3, we present the properties of BE approximations of CDFs, while in Section 4 we propose an approach to obtain stochastically ordered BE approximations. In Section 5, we evaluate our approach numerically to highlight advantages and limitations. Conclusions are drawn in Section 6.

2 Background

2.1 Bernstein polynomials

For any order $n \in \mathbb{N}$, the Bernstein operator B_n maps a function $G : [0,1] \to \mathbb{R}$ onto a polynomial defined as [13]:

$$B_n(G; y) := \sum_{i=0}^{n} \binom{n}{i} G\left(\frac{i}{n}\right) y^i (1-y)^{n-i} . \quad (1)$$

The Bernstein operator is linear, i.e., $B_n(\lambda_1 G_1 + \lambda_2 G_2; y) = \lambda_1 B_n(G_1; y) + \lambda_2 B_n(G_2; y)$, and it represents first-degree polynomials exactly, i.e., $B_n(1; y) = 1$

and $B_n(y; y) = y$. $B_n(G; y)$ also preserves many properties of G, which motivated its investigation and wide application as a tool for approximation.

Boundary Conditions, Bounds, Monotonicity. $B_n(G; y)$ is exactly equal to $G(y)$ at the endpoints of the domain $[0, 1]$ and preserves upper and lower bounds, i.e., $G(0) = B_n(G; 0)$, $G(1) = B_n(G; 1)$, and $\forall y \in [0, 1], m \leq G(y) \leq M \implies \forall y \in [0, 1], m \leq B_n(G; y) \leq M$. Moreover, if G is monotonic increasing (or decreasing) over $[0, 1]$, so is $B_n(G; y)$. By combination of these properties, if $G(y)$ is a CDF (or a CCDF) with support $[0, 1]$, so is $B_n(G, y)$ for any $n \in \mathbb{N}$, i.e., the Bernstein operator B_n maps distributions to valid distributions.

Uniform Convergence. For any continuous function G, the Bernstein operator ensures asymptotic convergence to 0 of the error $|G(y) - B_n(G; y)|$ when $n \to \infty$, uniformly over the entire support $[0, 1]$:

$$\forall \epsilon > 0, \exists \bar{n} \in \mathbb{N} \text{ such that } n > \bar{n} \implies \forall y \in [0, 1], |G(y) - B_n(G; y)| < \epsilon. \quad (2)$$

For further related results and explicit bounds we refer to [15].

2.2 Bernstein exponentials

The Bernstein exponential (BE) operator extends the Bernstein operator to the class of bounded functions with infinite support $[0, \infty)$ through the change of variables $y = e^{-x}$ (i.e., $x = -\log(y)$) which maps the support $[0, 1]$ onto $[0, \infty)$.

According to this, for any order $n \in \mathbb{N}$, the BE operator maps a function $F : [0, \infty) \to \mathbb{R}$ onto an exponential mixture of the form:

$$BE_n(F; x) := \sum_{i=0}^{n} \binom{n}{i} F\left(-\log\left(\frac{i}{n}\right)\right) e^{-ix}(1 - e^{-x})^{n-i} \quad (3)$$

By design, the BE operator inherits various shape preservation properties of the Bernstein operator. Since the change of variables $y = e^{-x}$ is continuous and strictly monotonic, $BE_n(F; x)$ is exactly equal to $F(x)$ for $x = 0$ and $x \to \infty$, it preserves the bounds of F, and if $F(x)$ is monotonic, so is $BE_n(F; x)$. Moreover, $BE_n(F; x)$ converges to $F(x)$ uniformly over $[0, \infty)$ as $n \to \infty$.

2.3 Phase-type distributions

A degree n continuous-time PH distribution is given by the time to absorption in a continuous-time Markov chain (CTMC) with n transient states and one absorbing state.

$$Q = \begin{pmatrix} -5.2 & 3 & 2.2 & 0 \\ 1.2 & -2.5 & 0.5 & 0.8 \\ 4 & 2.3 & -7.55 & 1.25 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

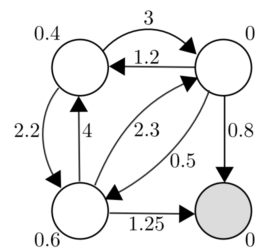

Fig. 1. A degree 3 PH distribution

A graphical example is provided in Figure 1 where the absorbing state is colored in gray. The initial probability vector and the infinitesimal generator of the CTMC are $q = \begin{pmatrix}0.4 \; 0 \; 0.6 \; 0\end{pmatrix}$ and Note that the last entry of q (in this case zero) and the last column of Q can be calculated from the rest (the row corresponding to the absorbing state is filled with zeros). Accordingly, the most widely used representation of a PH distribution includes only the parts of the initial probability vector and of the infinitesimal generator that correspond to transient states. For the above example, the representation is the vector-matrix pair

$$a = \begin{pmatrix}0.4 \; 0 \; 0.6\end{pmatrix}, \quad A = \begin{pmatrix} -5.2 & 3 & 2.2 \\ 1.2 & -2.5 & 0.5 \\ 4 & 2.3 & -7.55 \end{pmatrix}.$$

Given a vector-matrix pair (a, A), the corresponding PH distribution will be denoted by $\mathrm{PH}(a, A)$. The PDF, the CDF, and the CCDF of $\mathrm{PH}(a, A)$ will be denoted and can be calculated as

$$f_{a,A}(x) = ae^{xA}(-A\mathbf{1}), \; F_{a,A}(x) = 1 - ae^{xA}\mathbf{1}, \text{ and } \bar{F}_{a,A}(x) = ae^{xA}\mathbf{1}$$

where $\mathbf{1}$ denotes the column vector of ones.

In [9] it was shown and illustrated numerically through several examples that normalized BE approximation of a PDF results in a PH distribution. [9] provides also a more detailed description of the characteristics of BP and BE.

3 Approximation of Cumulative Density Functions by Bernstein Phase-Type Distributions

The degree n Bernstein exponential approximation of a given CDF $F(x)$ with support $[0, \infty)$ is

$$\hat{F}_n(x) = \sum_{i=0}^{n} F\left(\log \frac{n}{i}\right) \underbrace{\binom{n}{i} e^{-ix}(1 - e^{-x})^{n-i}}_{T_{n,i}(x)} \tag{4}$$

where the division by zero in case of $i = 0$ is resolved by considering the limiting value of $F(x)$ as x tends to infinity, i.e., $F\left(\log \frac{n}{0}\right) = \lim_{x \to \infty} F(x)$ which is equal to 1 if the CDF is not defective (we will denote this limit also simply by $F(\infty)$). At the other end, for $i = n$ we have $F(\log(n/n)) = F(0)$ which is 0 if there is no probability mass at 0 in the distribution.

The same Bernstein exponential can be obtained based on the CCDF.

Proposition 1. *Let $\bar{F}(x)$ be the CCDF of a given CDF $F(x)$, i.e., $\bar{F}(x) = 1 - F(x)$. The distribution obtained by the degree n Bernstein exponential approximation of $F(x)$, given by $\hat{F}_n(x)$ in* (4)*, is equal to the distribution derived from the degree n Bernstein exponential approximation of $\bar{F}(x)$, i.e., $\hat{F}_n(x) = 1 - \hat{\bar{F}}_n(x)$.*

Proof. The degree n Bernstein exponential approximation of $\bar{F}(x)$ is

$$\hat{\bar{F}}_n(x) = \sum_{i=0}^{n} \bar{F}\left(\log\frac{n}{i}\right) \cdot \binom{n}{i} e^{-ix}(1-e^{-x})^{n-i} = \tag{5}$$
$$\sum_{i=0}^{n}\left(1 - F\left(\log\frac{n}{i}\right)\right) \cdot \binom{n}{i} e^{-ix}(1-e^{-x})^{n-i} =$$
$$\sum_{i=0}^{n}\binom{n}{i} e^{-ix}(1-e^{-x})^{n-i} - \hat{F}_n(x) = \left(e^{-x} + (1-e^{-x})\right)^n - \hat{F}_n(x) = 1 - \hat{F}_n(x)$$

from which $\hat{F}_n(x) = 1 - \hat{\bar{F}}_n(x)$ directly follows. □

The following theorem shows that the approximation given in (4) corresponds to an acyclic PH distribution.

Theorem 1. *When $F(x)$ is a CDF with support $[0, \infty)$, i.e., $\lim_{x\to\infty} F(x) = 1$, then $F_{a,A}(x) = \hat{F}_n(x)$, where*

$$a = \begin{pmatrix} a_1 \;...\; a_n \end{pmatrix} \;\; \textit{with}\; a_i = F\left(\log\frac{n}{i-1}\right) - F\left(\log\frac{n}{i}\right), \tag{6}$$

and

$$A = \begin{pmatrix} -1 & 1 & 0 & & ... & & \\ 0 & -2 & 2 & 0 & & & ... \\ & & & \ddots & & & \\ & ... & & 0 & -(n-1) & n-1 \\ & & ... & & 0 & -n \end{pmatrix}. \tag{7}$$

I.e., the CDF of PH(a, A) is equal to the approximation in Eq. (4)*.*

The graphical representation of PH(a, A) is shown in Figure 2 (where the role of $F(0)$ and $F(\infty)$ can be explicitly seen).

Proof. The Laplace transform of the PDF of PH(a, A) is

$$f^*_{a,A}(s) = \sum_{i=1}^{n}\left(F\left(\log\frac{n}{i-1}\right) - F\left(\log\frac{n}{i}\right)\right)\prod_{j=i}^{n}\frac{j}{j+s} \tag{8}$$

where the product $\prod_{j=i}^{n}\frac{j}{j+s}$ corresponds to the convolution of $n-i+1$ exponential random variables with parameters $i, i+1, ..., n$.

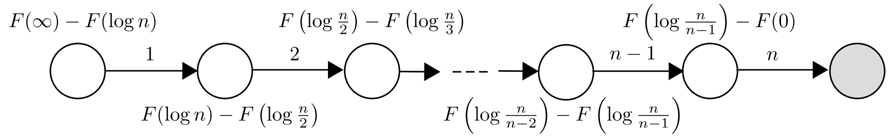

Fig. 2. Bernstein PH approximation of a cdf $F(x)$

$\hat{F}_n(x)$ is the weighted sum of terms in the form

$$T_{n,i}(x) = \binom{n}{i} e^{-ix}(1-e^{-x})^{n-i} \tag{9}$$

with derivative

$$t_{n,i}(x) = T'_{n,i}(x) = (n-i)\binom{n}{i} e^{-(i+1)x}(1-e^{-x})^{n-i-1} - i\binom{n}{i} e^{-ix}(1-e^{-x})^{n-i}$$

whose Laplace transform is

$$t^*_{n,i}(s) = \int_0^\infty e^{-sx} t_{n,i}(x)dx = \prod_{j=i+1}^{n} \frac{j}{j+s} - \prod_{j=i}^{n} \frac{j}{j+s}\,.$$

Accordingly, the Laplace transform of $\hat{f}(x) = \hat{F}'_n(x)$ is

$$\hat{f}^*(s) = \int_0^\infty e^{-sx} \hat{f}(x)dx = \sum_{i=0}^{n} F\left(\log \frac{n}{i}\right)\left(\prod_{j=i+1}^{n} \frac{j}{j+s} - \prod_{j=i}^{n} \frac{j}{j+s}\right) =$$

$$\sum_{i=1}^{n}\left(F\left(\log \frac{n}{i-1}\right) - F\left(\log \frac{n}{i}\right)\right)\prod_{j=i}^{n} \frac{j}{j+s} \tag{10}$$

from which, by comparing (8) and (10), we have $f^*_{a,A}(s) = \hat{f}^*(s)$. Additionally, using $1 - F_{a,A}(0) = \sum_{i=1}^{n} a_i = F(\infty) - F(0) = 1 - F(0)$ and $F(0) = \hat{F}_n(0)$, $F(\infty) = \hat{F}_n(\infty)$, as discussed after (3), we also have $F_{a,A}(x) = \hat{F}_n(x)$. □

The time domain equivalent of (10) is provided in the following proposition.

Proposition 2. *If* $F(0) = 0$*, then*

$$\hat{F}_n(x) = \sum_{i=1}^{n}\left(F\left(\log \frac{n}{i-1}\right) - F\left(\log \frac{n}{i}\right)\right)\sum_{j=0}^{i-1} T_{n,j}(x). \tag{11}$$

Proof. From (4) and (9), we have

$$
\begin{aligned}
\hat{F}_n(x) &= \sum_{i=0}^{n} F\left(\log \frac{n}{i}\right) \cdot T_{n,i}(x) = \sum_{i=0}^{n} F\left(\log \frac{n}{i}\right) \cdot \left(\sum_{j=0}^{i} T_{n,j}(x) - \sum_{j=0}^{i-1} T_{n,j}(x) \right) \\
&= \sum_{i=0}^{n} F\left(\log \frac{n}{i}\right) \cdot \sum_{j=0}^{i} T_{n,j}(x) - \sum_{i=1}^{n} F\left(\log \frac{n}{i}\right) \sum_{j=0}^{i-1} T_{n,j}(x) \\
&= \sum_{i=0}^{n} F\left(\log \frac{n}{i}\right) \cdot \sum_{j=0}^{i} T_{n,j}(x) - \sum_{i=0}^{n-1} F\left(\log \frac{n}{i+1}\right) \sum_{j=0}^{i} T_{n,j}(x) \\
&= \underbrace{F(0)}_{0} \cdot \underbrace{\sum_{j=0}^{n} T_{n,j}(x)}_{1} + \sum_{i=0}^{n-1} \left(F\left(\log \frac{n}{i}\right) - F\left(\log \frac{n}{i+1}\right) \right) \sum_{j=0}^{i} T_{n,j}(x) \\
&= \sum_{i=1}^{n} \left(F\left(\log \frac{n}{i-1}\right) - F\left(\log \frac{n}{i}\right) \right) \sum_{j=0}^{i-1} T_{n,j}(x)
\end{aligned}
$$

□

From Proposition 2 it also follows directly that

$$\frac{d^k}{dx^k} \sum_{j=0}^{i-1} T_{n,j}(x)|_{x=0} = 0 \qquad \text{for} \quad 0 \le k \le n-i \tag{12}$$

by considering Fig. 2 when the last $n-i$ nodes have 0 initial probability, and, as a further consequence, we also have

$$\frac{d^j}{dx^j} T_{n,i}(x)|_{x=0} = 0 \qquad \text{for} \quad 0 \le j \le n-i-1 \; . \tag{13}$$

According to Theorem 1, if $F(x)$ is non-decreasing, $F(0) = 0$ and $F(\infty) = 1$, then its $F_{a,A}(x) = \hat{F}_n(x)$ approximation based on (4) is such that $a_i > 0$ for $i = 1, ..., n$, and $\sum_{i=1}^{n} a_i = 1$.

In a BE approximation the coefficient of the term $T_{n,n}(x)$ is equal to the value of the approximation at zero. Vice versa, the coefficient of the term $T_{n,0}$ is equal to the value of the approximation as $x \to \infty$. This implies the following proposition.

Proposition 3. *Given a CDF $F(x)$ that corresponds to a distribution that is with mass at zero ($F(0) > 0$) and/or defective ($\lim_{x\to\infty} F(x) < 1$), the approximation*

$$\hat{F}_n(x) = 0 \cdot T_{n,n}(x) + \sum_{i=1}^{n-1} F\left(\log \frac{n}{i}\right) T_{n,i}(x) + 1 \cdot T_{n,0}(x) \tag{14}$$

corresponds to a non-defective distribution without mass at zero. The same can be achieved by approximating the CCDF $\bar{F}(x)$ in the form

$$\hat{\bar{F}}_n(x) = 0 \cdot T_{n,0}(x) + \sum_{i=1}^{n-1} \bar{F}\left(\log \frac{n}{i}\right) T_{n,i}(x) + 1 \cdot T_{n,n}(x). \tag{15}$$

Note that every PH distribution constructed through a Bernstein exponential approximation has the same infinitesimal generator A given in (7). For this reason, given a vector $a = (a_1 \dots a_n)$ the distribution PH(a, A) will be referred to as BPH(a). The PDF, the CDF, and the CCDF of a BPH(a) will be denoted by $f_a(x), F_a(x)$ and $\bar{F}_a(x)$, respectively.

4 Stochastically smaller and larger approximation

Here we study the possibility to create BPH(a) distributions that guarantee stochastic order with respect to the distribution we aim to approximate.

If $\bar{F}_X(x)$ and $\bar{F}_Y(x)$ are the CCDF of X and Y, then X is stochastically smaller than Y (equivalently, Y is stochastically larger than X) if and only if

$$P(X > z) = \bar{F}_X(z) \leq \bar{F}_Y(z) = P(Y > z), \forall z \geq 0. \tag{16}$$

In the sequel we will use the notation

$$F_{-\epsilon}(x) = \max(F(x) - \epsilon, 0), \quad F_{+\epsilon}(x) = \min(F(x) + \epsilon, 1), \tag{17}$$

$$\bar{F}_{+\epsilon}(x) = \min(\bar{F}(x) + \epsilon, 1), \quad \bar{F}_{-\epsilon}(x) = \max(\bar{F}(x) - \epsilon, 0), \tag{18}$$

among which $F_{-\epsilon}(x)$ and $\bar{F}_{+\epsilon}(x)$ are useful to obtain larger approximations while $F_{+\epsilon}(x)$ and $\bar{F}_{-\epsilon}(x)$ to obtain smaller ones. In case of $\epsilon > 0$, the distributions corresponding to the CDFs in (17) and to the CCDFs in (18) are either with mass at zero or are defective. As shown by Proposition 3 we can still easily obtain approximations of them that correspond to non-defective distributions without mass at zero using (14) or (15).

Let $\hat{x}$ and $\check{x}$ be such that $F(\hat{x}) = 1 - \bar{F}(\hat{x}) = 1 - \epsilon$ and $F(\check{x}) = 1 - \bar{F}(\check{x}) = \epsilon$, then (17) and (18) can be written as

$$F_{+\epsilon}(x) = \begin{cases} F(x) + \epsilon & \text{if } x \leq \hat{x}, \\ 1 & \text{if } x > \hat{x}, \end{cases} \text{ and } F_{-\epsilon}(x) = \begin{cases} 0 & \text{if } x \leq \check{x}, \\ F(x) - \epsilon & \text{if } x > \check{x}, \end{cases} \tag{19}$$

$$\bar{F}_{+\epsilon}(x) = \begin{cases} 1 & \text{if } x \leq \check{x}, \\ \bar{F}(x) + \epsilon & \text{if } x > \check{x}, \end{cases} \text{ and } \bar{F}_{-\epsilon}(x) = \begin{cases} \bar{F}(x) - \epsilon & \text{if } x \leq \hat{x}, \\ 0 & \text{if } x > \hat{x}. \end{cases} \tag{20}$$

Furthermore, let $\hat{n}$ be such that $\log \frac{n}{\hat{n}} < \hat{x}$, but $\log \frac{n}{\hat{n}-1} > \hat{x}$, and similarly, let $\check{n}$ be such that $\log \frac{n}{\check{n}} < \check{x}$, but $\log \frac{n}{\check{n}-1} > \check{x}$. That is,

$$\hat{n} = \left\lceil \frac{n}{e^{\hat{x}}} \right\rceil, \text{ and } \check{n} = \left\lceil \frac{n}{e^{\check{x}}} \right\rceil. \tag{21}$$

We assume that ϵ is a small error term such that $0 < \epsilon \ll 1/2$. In this case $\check{x} \leq \hat{x}$ and $\check{n} \geq \hat{n}$.

A consequence of (21) is that $\lim_{n\to\infty} \frac{\hat{n}}{n} = e^{-\hat{x}}$ and $\lim_{n\to\infty} \frac{\check{n}}{n} = e^{-\check{x}}$. That is, both $\hat{n}$ and $\check{n}$ increase to infinity with n.

Theorem 2. *(a) Let $\bar{F}(x)$ be a continuous CCDF with the following property:*
- *$\bar{F}(x) \le ce^{-ax}$ for some $0 < c < \infty$ and $1 < a < \infty$;*
- *there exists a finite n_0 for which $\frac{d^{n_0}}{dx^{n_0}}\bar{F}(x)|_{x=0} \neq 0$.*

Then for any $\epsilon > 0$ small enough there exists an n such that the function

$$\hat{\hat{\bar{F}}}_{+\epsilon,n}(x) = \sum_{i=1}^{n} \bar{F}_{+\epsilon}\left(\log\frac{n}{i}\right) \cdot T_{n,i}(x) \tag{22}$$

is the CCDF of a BPH distribution that stochastically dominates $\bar{F}(x)$.

(b) Let $\bar{F}(x)$ be a continuous CCDF with the following properties:
- *$\bar{F}(x) > ce^{-ax}$ for some $0 < c < \infty$ and $a < \infty$;*
- *$\frac{d}{dx}\bar{F}(x)|_{x=0}$ is finite.*

Then for any $\epsilon > 0$ small enough there exists an n such that the function

$$\hat{\hat{\bar{F}}}_{-\epsilon,n}(x) = \sum_{i=1}^{n-1} \bar{F}_{-\epsilon}\left(\log\frac{n}{i}\right) \cdot T_{n,i}(x) + T_{n,n}(x) \tag{23}$$

is the CCDF of a stochastically smaller BPH distribution.

Proof. (a) According to Prop. 3, $\hat{\hat{\bar{F}}}_{+\epsilon,n}$ is a proper CCDF. Eq. (22) can be further written as

$$\hat{\hat{\bar{F}}}_{+\epsilon,n}(x) = \sum_{i=1}^{\check{n}-1} \left(\bar{F}\left(\log\frac{n}{i}\right) + \epsilon\right) \cdot T_{n,i}(x) + \sum_{i=\check{n}}^{n} T_{n,i}(x)\,. \tag{24}$$

We investigate three separate domains: $x \to 0$, $x \to \infty$ and the main body of $\bar{F}(x)$ (on an interval separated away from both 0 and ∞).

For the behavior around 0, the larger index terms are responsible due to Eqs. (12) and (13); around $x = 0$, $\bar{F}(x)$ and $\sum_{i=\check{n}}^{n} T_{n,i}(x)$ both start from 1 and decrease; for $\bar{F}(x)$, the first $n_0 - 1$ derivatives are 0 due to the assumption, while for $\sum_{i=\check{n}}^{n} T_{n,i}(x)$, the first $n - \check{n}$ derivatives are 0, so as long as $n - \check{n} > n_0$, $\sum_{i=\check{n}}^{n} T_{n,i}(x)$ dominates $\bar{F}(x)$ on some interval $[0, x_1]$.

At $x \to \infty$, the dominant e^{-x} term is obtained for $i = 1$, from which

$$\hat{\hat{\bar{F}}}_{+\epsilon,n}(x) \sim \left(\bar{F}(\log n) + \epsilon\right) \cdot e^{-x}, \tag{25}$$

which dominates $\bar{F}(x)$ due to the assumption $\bar{F}(x) \le ce^{-ax}$ for $1 < a < \infty$ on some interval $[x_2, \infty)$.

For the main body of $\bar{F}(x)$ on the interval $[x_1, x_2]$, we utilize the fact that $BE_n(\bar{F}_{+\epsilon}; x)$ approximates $\bar{F}(x) + \varepsilon$ uniformly. $BE_n(\bar{F}_{+\epsilon}; x)$ only differs from $\bar{F}_\epsilon$ in the coefficient of the $T_{n,0}(x)$ term, which converges pointwise to 0 as n increases, so this term will vanish over $[x_1, x_2]$. This ensures

$$\bar{F}(x) \le \hat{\hat{\bar{F}}}_{+\epsilon,n}(x)$$

on $[x_1, x_2]$ via triangle inequality.

(b) According to Prop. 3, $\hat{\bar{F}}_{-\epsilon,n}$ is a proper CCDF.Eq. (23) can be rewritten as

$$\hat{\bar{F}}_{-\epsilon,n}(x) = \sum_{i=\hat{n}}^{n-1} \left(\bar{F}\left(\log \frac{n}{i}\right) - \epsilon \right) \cdot T_{n,i}(x) + T_{n,n}(x). \tag{26}$$

For the behavior around 0, once again the larger index terms are responsible:

$$\left(\bar{F}\left(\log \frac{n}{n-1}\right) - \epsilon \right) \cdot T_{n,n-1}(x) + T_{n,n}(x) \sim 1 + (1-\epsilon)e^{-(n-1)x}(1-e^{-x}),$$

whose derivative at 0 is $-(1-\epsilon)(n-1)$. The assumption that $F(x)$ has a finite derivative at 0 ensures that there exists a finite n such that $\hat{\bar{F}}_{-\epsilon,n}(x) < \bar{F}(x)$ on $[0, x_1]$.

At $x \to \infty$, the dominant term is obtained for $i = \hat{n}$, so

$$\hat{\bar{F}}_{-\epsilon,n}(x) \sim \left(\bar{F}\left(\log \frac{n}{\hat{n}+1}\right) - \epsilon \right) \cdot e^{-(\hat{n}+1)x}. \tag{27}$$

As long as the assumption $\bar{F}(x) > ce^{-ax}$ for some $a < \infty$ holds, n can be chosen large enough so that $\hat{n} > a$, and $\hat{\bar{F}}_{-\epsilon,n}(x) < \bar{F}(x)$ on $[x_2, \infty)$. For the main body of $\bar{F}(x)$ on the interval $[x_1, x_2]$, pointwise convergence can be applied similarly to part (a). ☐

The assumptions on the derivatives in Theorem 2 are necessary; for any $\bar{F}$ BPH CCDF, there exists a finite n_0 for which $\frac{d^{n_0}}{dx^{n_0}}\bar{F}(x)|_{x=0} \neq 0$, so no BPH CCDF can dominate a function whose every higher order derivative at $x = 0$ is 0, such as $1 - e^{-\frac{1}{x^2}}$.

Similarly, every F BPH CCDF has a finite derivative at 0, so no BPH CCDF can be stochastically smaller than a function like $\hat{F}(x) = 1 - \sqrt{x}$ that has an infinite derivative at 0.

Bounded support random variables cannot be dominated either; we state this as a separate corollary.

Corollary 1. *Consider a CCDF $\bar{F}(x)$. Then*

$$\exists x > 0,\ \bar{F}(x) = 1 \implies \nexists\ \textit{BPH(a)},\ \forall x \geq 0,\ \bar{F}(x) \leq \bar{F}_a(x)$$

and

$$\exists x > 0,\ \bar{F}(x) = 0 \implies \nexists\ \textit{BPH(a)},\ \forall x \geq 0,\ \bar{F}(x) \geq \bar{F}_a(x)$$

Practically relevant examples for bounded support random variables include shifted and truncated distributions:

$$\bar{F}_{se}(x) = \begin{cases} 1 & x \leq 1 \\ e^{-(x-1)} & x > 1 \end{cases}, \quad \bar{F}_{te}(x) = \begin{cases} \dfrac{e^{1-x}-1}{e-1} & x \leq 1 \\ 0 & x > 1 \end{cases} \tag{28}$$

Figures 3 and 4 illustrate the issues when trying to find stochastically larger or smaller distributions for $\bar{F}_{se}$ and $\bar{F}_{te}$.

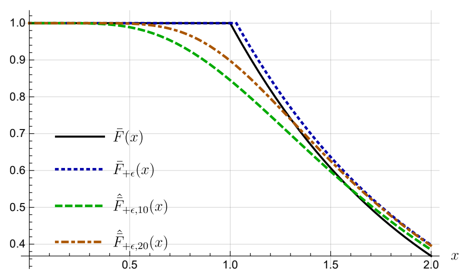

Fig. 3. BPH approximations of a shifted exponential distribution: there does not exist larger BPH distribution.

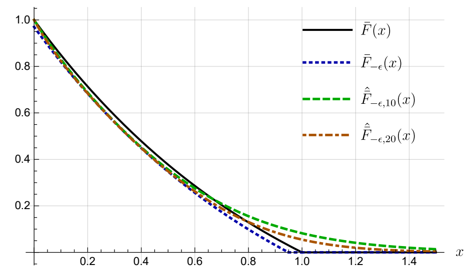

Fig. 4. BPH approximations of a truncated exponential distribution: there does not exist smaller BPH distribution.

5 Numerical investigations

In this section we use the order k Erlang distribution with mean equal to one, with CCDF

$$\bar{F}(x) = \sum_{i=0}^{k-1} e^{-kx}(kx)^i/i! \ , \tag{29}$$

to numerically investigate BPH approximations based on the CCDF (which, as shown in Theorem 1, is equivalent to using the CDF). We approximate $\bar{F}(x)$ itself and its increased and decreased variants $\bar{F}_{+\epsilon}(x)$ and $\bar{F}_{-\epsilon}(x)$ as well, in order to obtain stochastically larger and smaller BPH distributions.

While an Erlang distribution is itself a PH distribution, it provides a straightforward way to analyze several crucial characteristics of BPH approximations for the following reasons. The order k can be used to control both the behavior at zero and as x tends to infinity. The larger k, the longer the CCDF remains close to one (the first $k-1$ derivatives of the CCDF are zero at $x=0$), and the more phases we need to construct a stochastically larger BPH. At $x \to \infty$, the larger k, the faster the CCDF decays (at rate $e^{-kx}x^{k-1}$) and the more phases we need to construct a stochastically smaller BPH. Moreover, the Erlang distribution is known to have the smallest possible squared coefficient of variation (SCV) among PH distributions of a given order k independent of the mean [1], namely $1/k$. This allows us to study easily also the impact of the SCV on the goodness of the approximation.

In case of the Erlang CCDF, when approximating $\bar{F}_{+\epsilon}(x)$ and $\bar{F}_{-\epsilon}(x)$, Theorem 2 guarantees that there exists an n such that $\hat{\bar{F}}_{+\epsilon,n}(x)$ ($\hat{\bar{F}}_{-\epsilon,n}(x)$) is stochastically larger (smaller) than $\hat{F}(x)$. For a given n and ϵ, checking whether $\hat{\bar{F}}_{+\epsilon,n}(x) \geq \bar{F}(x)$ for every $x \geq 0$ ($\hat{\bar{F}}_{-\epsilon,n}(x) \leq \bar{F}(x)$ for every $x \geq 0$) is not straightforward.

We checked the involved functions at $x=0$ and $x \to \infty$ analytically, and analyzed the difference between $\bar{F}(x)$ and the approximations numerically over the

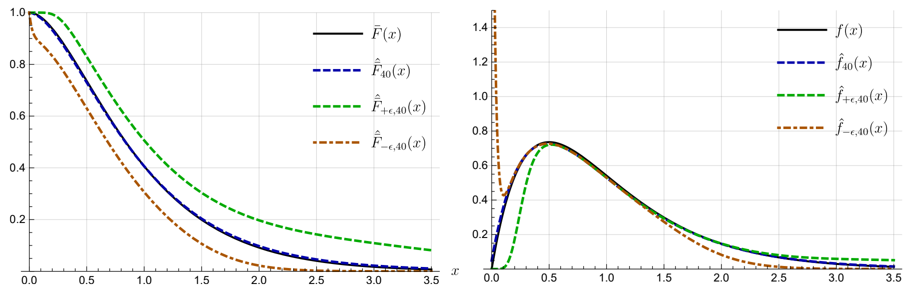

Fig. 5. Approximating the Erlang CCDF with $k = 2, n = 40, \epsilon = 0.1$: on the left the resulting CCDFs, on the right the corresponding PDFs.

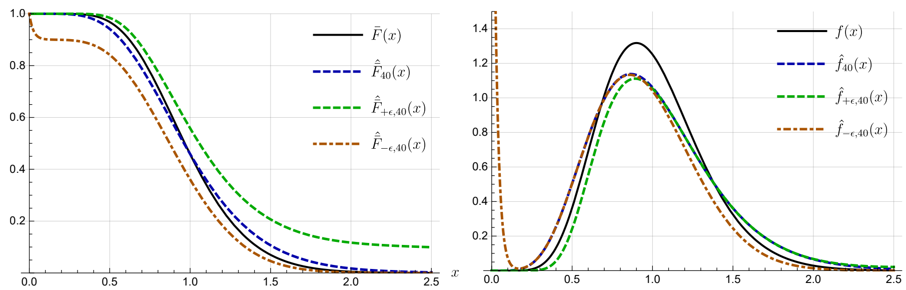

Fig. 6. Approximating the Erlang CCDF with $k = 10, n = 40, \epsilon = 0.1$: on the left the resulting CCDFs, on the right the corresponding PDFs.

main body of the functions. If $\min_x \hat{\hat{\bar{F}}}_{+\epsilon,n}(x) - \bar{F}(x) = 0$ then $\hat{\hat{\bar{F}}}_{+\epsilon,n}(x)$ is stochastically larger than $\bar{F}(x)$; vice versa, if $\max_x \hat{\hat{\bar{F}}}_{-\epsilon,n}(x) - \bar{F}(x) = 0$ then $\hat{\hat{\bar{F}}}_{-\epsilon,n}(x)$ is stochastically smaller than $\bar{F}(x)$. (Note that $\bar{F}(0) = \hat{\hat{\bar{F}}}_{+\epsilon,n}(0) = \hat{\hat{\bar{F}}}_{-\epsilon,n}(0) = 1$ is guaranteed by the approximation, which implies that $\min_x \hat{\hat{\bar{F}}}_{+\epsilon,n}(x) - \bar{F}(x) \leq 0$ and $\max_x \hat{\hat{\bar{F}}}_{-\epsilon,n}(x) - \bar{F}(x) \geq 0$.)

Figure 5 shows the CCDFs and PDFs resulting from the approximation of the Erlang CCDF with $k = 2, n = 40$ and $\epsilon = 0.1$. Approximating $\bar{F}(x)$ itself via (3) gives a good approximation but does not guarantee stochastic order. Approximating $\bar{F}_{+\epsilon}(x)$ and $\bar{F}_{-\epsilon}(x)$ provides a larger and a smaller distribution, respectively, but the resulting CCDFs are far from the original due to the relatively large values of ϵ. A particular consequence of using $\bar{F}_{-\epsilon}(x)$ can be observed for the PDF $\hat{f}_{-\epsilon,40}(x)$ at zero where we have $\hat{f}_{-\epsilon,40}(0) = 4.05$. This is due to the fact that $\hat{\hat{\bar{F}}}_{-\epsilon,40}(x)$ has no derivatives equal to zero at zero. The larger ϵ, the larger $\hat{f}_{-\epsilon,40}(0)$. The SCV in this case is relatively large, 1/2, and hence, approximating $\hat{F}(x)$ provides a CCDF and a corresponding PDF that follows closely the original CCDF and PDF.

Figure 6 shows analogous experiments for $k = 10$. Similar to the $k = 2$ case, approximating $\bar{F}_{+\epsilon}(x)$ and $\bar{F}_{-\epsilon}(x)$ provides a larger and a smaller distribution, respectively, and we have a peak at zero in the PDF of the smaller distribution.

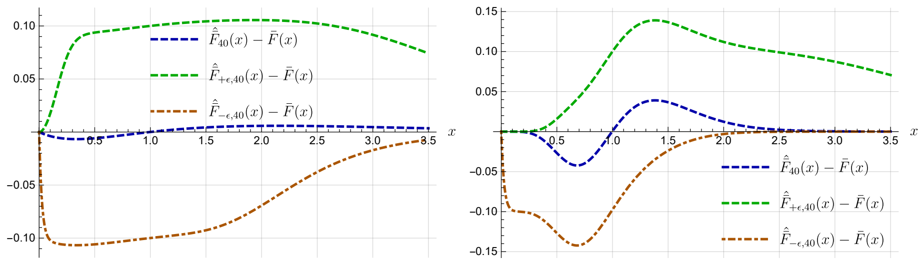

Fig. 7. Difference in the CCDFs approximating the Erlang CCDF with $n = 40, \epsilon = 0.1$; with $k = 2$ on the left and with $k = 10$ on the right.

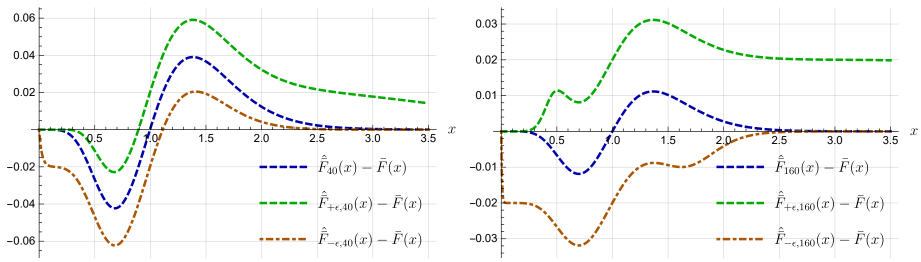

Fig. 8. Difference in the CCDFs approximating the Erlang CCDF with $k = 10, \epsilon = 0.02$; with $n = 40$ on the left and with $n = 160$ on the right.

The main difference with respect to $k = 2$ is the much lower SCV (1/10 for $k = 10$). Accordingly, $\hat{\hat{f}}(x)$ is unable to capture the "narrow" shape of $f(x)$. The rigid structure of the BPH distribution (fixed intensities and distributed initial probabilities, see Figure 2) is not ideal to obtain low SCV.

In Figure 7 we show the difference between the original CCDF and the approximating CCDFs for the two experiments considered so far, i.e., $n = 40, \epsilon = 0.1$ with $k = 2$ and $k = 10$. For the approximating $\bar{F}_n(x)$, close to zero the resulting CCDF is smaller than the original, and later it becomes larger, which is the typical behaviour when the derivative of the CCDF is to zero at zero.

As expected, for smaller values of ϵ, it can be necessary to increase the number of phases in order to obtain a stochastically larger (or smaller) distribution. In Figure 8, using $k = 10$ and $\epsilon = 0.02$, we show the difference between the approximating CCDFs and the original one with $n = 40$ and $n = 160$. With 40 phases, there is no stochastic order between the Erlang CCDF and the approximations. Indeed, both $\hat{\hat{F}}_{+\epsilon,40}(x) - \bar{F}(x)$ and $\hat{\hat{F}}_{-\epsilon,40}(x) - \bar{F}(x)$ cross the x axis. For $n = 160$, stochastic order is guaranteed.

We report also some results for several different values of n, namely, $n = 5, 10, 20, 40, 80$ and 160. We investigate the minimum and the maximum of $\hat{\hat{F}}_n(x) - \bar{F}(x)$, $\hat{\hat{F}}_{+\epsilon,n}(x) - \bar{F}(x)$ and $\hat{\hat{F}}_{-\epsilon,n}(x) - \bar{F}(x)$. The results are shown in Figure 9 using $k = 2$ and $k = 10$ with $\epsilon = 0.05$. The minimum and maximum of $\hat{\hat{F}}_n(x) - \bar{F}(x)$ shows a symmetric behavior with respect to the x axis. For the stochastically larger and smaller approximations, $\min_x \hat{\hat{F}}_{+\epsilon,n}(x) - \bar{F}(x)$

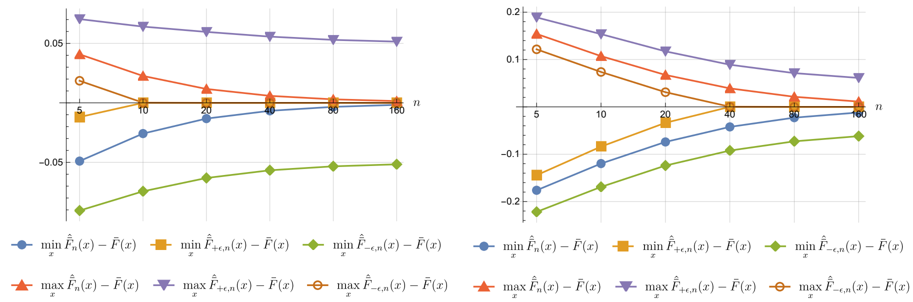

Fig. 9. Minimum and the maximum of $\hat{\bar{F}}_n(x) - \bar{F}(x)$, $\hat{\bar{F}}_{+\epsilon,n}(x) - \bar{F}(x)$ and $\hat{\bar{F}}_{-\epsilon,n}(x) - \bar{F}(x)$ for various values of n with $\epsilon = 0.05$ and $k = 2$ (left) and $k = 10$ (right).

is symmetric to $\max_x \hat{\bar{F}}_{-\epsilon,n}(x) - \bar{F}(x)$ and $\min_x \hat{\bar{F}}_{-\epsilon,n}(x) - \bar{F}(x)$ is symmetric to $\max_x \hat{\bar{F}}_{+\epsilon,n}(x) - \bar{F}(x)$. A larger (smaller) distribution is guaranteed once n is increased large enough so that $\min_x \hat{\bar{F}}_{+\epsilon,n}(x) - \bar{F}(x)$ ($\max_x \hat{\bar{F}}_{-\epsilon,n}(x) - \bar{F}(x)$) reaches the x axis.

Next, we numerically investigate the minimal n as a function of ϵ, denoted by $n_+(\epsilon)$ and $n_-(\epsilon)$, in order to obtain stochastically larger or smaller approximations, respectively. The test functions considered are the following.

- The Weibull(2,2) CCDF $\bar{F}(x) = e^{-x^2/4}$ satisfies condition (a) of Theorem 2 but not (b) because $\bar{F}(x)$ decays at $x \to \infty$ faster than exponential.
- The Weibull(1/2,1/2) CCDF $\bar{F}(x) = e^{-\sqrt{2x}}$ does not satisfy condition (a) of Theorem 2 since $\bar{F}(x)$ decays at $x \to \infty$ slower than exponential. Neither condition (b) is satisfied since the derivative of the CCDF at $x = 0$ is infinite.
- The Erlang CCDF (29) with any order k satisfies both conditions (a) and (b) of Theorem 2. We use $k = 2$ and $k = 10$ and refer to the corresponding cases as Erlang(2) and Erlang(10), respectively.
- The (shifted) Pareto(1) CCDF $\bar{F}(x) = 1/(x + 1)$ satisfies condition (b) but not (a) because the decay of $\bar{F}(x)$ is slower than exponential at $x \to \infty$.
- Similarly, the (shifted) Pareto(5) CCDF $\bar{F}(x) = 1/(x+1)^5$ satisfies condition (b) but not (a).

Table 1 shows $n_+(\epsilon)$ and $n_-(\epsilon)$ for each choice of $\bar{F}$. When such $n_+(\epsilon)$ and/or $n_-(\epsilon)$ do not exist the table indicates $\nexists$ (this happens always in accordance with Theorem 2). It turns out that for all of the test functions, if $n_+(\epsilon)$ and/or $n_-(\epsilon)$ exist then larger and smaller CCDFs are obtained for any $n \geq n_+(\epsilon)$ and $n \geq n_-(\epsilon)$, respectively, i.e.,

$$\hat{\bar{F}}_{+\epsilon,n}(x) \geq \bar{F}(x) \quad x \in [0, \infty),\ \forall n \geq n_+(\epsilon) \quad \text{and}$$
$$\hat{\bar{F}}_{-\epsilon,n}(x) \leq \bar{F}(x) \quad x \in [0, \infty),\ \forall n \geq n_-(\epsilon).$$

Table 1 shows that apart from the exception (and possible corner case) Pareto(1), $n_+(\epsilon)$ and $n_-(\epsilon)$ typically increase linearly in $1/\epsilon$, with a constant

	Weib.(2,2)		Weib.(1/2,1/2)		Erlang(2)		Erlang(10)		Pareto(1)		Pareto(5)	
ϵ	$n_+(\epsilon)$	$n_-(\epsilon)$	$n_+(\epsilon)$	$n_-(\epsilon)$	$n_+(\epsilon)$	$n_-(\epsilon)$	$n_+(\epsilon)$	$n_-(\epsilon)$	$n_+(\epsilon)$	$n_-(\epsilon)$	$n_+(\epsilon)$	$n_-(\epsilon)$
0.1	7	$\nexists$	$\nexists$	$\nexists$	5	7	24	29	$\nexists$	1	$\nexists$	11
0.01	68	$\nexists$	$\nexists$	$\nexists$	27	28	192	180	$\nexists$	1	$\nexists$	74
0.001	687	$\nexists$	$\nexists$	$\nexists$	271	237	1976	1869	$\nexists$	1	$\nexists$	746

Table 1. Minimal order, $n_+(\epsilon)$ and $n_-(\epsilon)$, to obtain stochastically larger and smaller approximations, respectively, as function of ϵ.

factor depending on $\bar{F}$. For Pareto(1), $\hat{\bar{F}}_{-\epsilon,n}(x) \leq \bar{F}(x)$ holds already for $n = 1$, for any choice of ϵ examined.

Finally, we apply some of the already studied approximations to M/G/1 queues. The service time distribution is Erlang (see (29)) with either $k = 2$ or $k = 10$. We use the same values of ϵ for approximations as in Table 1, namely 0.1, 0.01 and 0.001, and the minimal n that allows us to obtain stochastically larger and smaller approximate BPH service time distributions (this is also indicated in Table 1 as $n_+(\epsilon)$ and $n_-(\epsilon)$, respectively). The utilization of the queue is set to 0.7. The queue length distribution of the resulting M/BPH/1 queue can be calculated by the procedure provided in [9] that has linear complexity in the order n and hence allows us to use large values of n in the computation.

Since the BPH approximations guarantee stochastic order with respect to the original service time distribution, the CCDF of the queue length distribution and its upper and lower bounds for $k = 2$ are illustrated in Figure 10 (left), while the difference between the bounds are plotted on the right. As expected, the bounds become tighter as $\epsilon \searrow 0$. The largest difference between the upper bound and the lower bound is 0.5202, 0.0885 and 0.01357, respectively, for $\epsilon = 0.1$, $\epsilon = 0.01$ and $\epsilon = 0.001$. For $k = 10$ the results are shown in Figure 11. In this case the largest difference between the upper bound and the lower bound is 0.6769, 0.09476 and 0.01388, respectively, for the same values of ϵ.

The figures indicate that for $k = 2$ and $k = 10$, the differences between the upper bound and the lower bound are rather similar in spite of the essentially different service time distribution (the SCV of the service time is 1/2 for $k = 2$ and 1/10 for $k = 10$). We note that the figures hide an important aspect of the approximation which is highlighted in Table 1. Namely, in case of $k = 10$ the minimal order guaranteeing stochastic ordering is about seven times larger than in case of $k = 2$, however this increase does not lead to unfeasible computations due to the simplicity of the construction and the application of BPH approximations.

6 Conclusions

We applied Bernstein exponentials to the approximation of CDFs and showed that the resulting CDFs are valid and describe random variables that belong to a subclass of acyclic PH distributions, allowing efficient approximations of

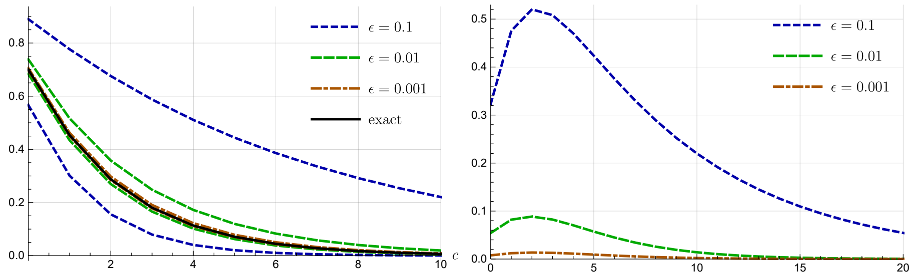

Fig. 10. Bounds on the probability of more than c jobs present in the queue in case of Erlang service time with $k = 2$ (left) and the difference between the upper bound and the lower bound (right); obtained by larger and smaller BPH approximations with minimal order guaranteeing stochastic order for various values of ϵ. The plots are valid at integer values of c, and the discrete points are connected for better visibility.

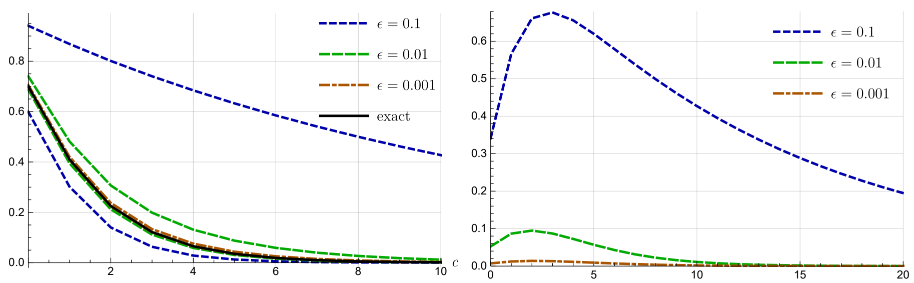

Fig. 11. The same results as in Figure 10 with $k = 10$.

non-Markovian models. We also provided an approach to obtain stochastically ordered approximations, which open the way to the application in problems where a safe approximation of performance metrics is required.

In future work, we plan to analyze the approximation of scaled functions, i.e., $F(cx)$, where c gives some freedom in choosing the points where F is sampled and also to relax some conditions in Theorem 2. Another problem to face is the approximation error highlighted in Section 5 for stochastically lower approximations near zero (e.g., $\hat{f}_{-\epsilon,40}(0) \gg 0$ when $f(0) = 0$) through alternative approaches to obtain stochastically larger or smaller approximations.

Acknowledgments

M. Telek was supported by the OTKA K-138208 project of the Hungarian Scientific Research Fund. E. Vicario was supported bt the RESTART project, as a part of the Italian PNNR programme.

References

1. D Aldous and L Shepp. The least variable phase type distribution is erlang. *Stochastic Models*, 3 (3), 1987.
2. S. Asmussen and O. Nerman. Fitting Phase-type distributions via the EM algorithm. In *Proceedings of Symposium i Advent Statistik*, pages 335–346, Copenhagen, 1991.
3. François Baccelli and Armand M Makowski. Multidimensional stochastic ordering and associated random variables. *Operations Research*, 37(3):478–487, 1989.
4. A. Bobbio and A. Cumani. ML estimation of the parameters of a PH distribution in triangular canonical form. In *Computer Performance Evaluation*, pages 33–46. Elsevier, 1992.
5. A. Bobbio, A. Horváth, and M. Telek. Matching three moments with minimal acyclic phase type distributions. *Stochastic Models*, 21:303–326, 2005.
6. A. Feldman and W. Whitt. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*, 31:245–279, 1998.
7. Jean-Michel Fourneau and Nihal Pekergin. A numerical analysis of dynamic fault trees based on stochastic bounds. In *International Conference on Quantitative Evaluation of Systems*, pages 176–191. Springer, 2015.
8. 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.
9. András Horváth and Enrico Vicario. Construction of Phase Type Distributions by Bernstein Exponentials. In *Proceedings of ASMTA 2023*, volume 14231 of *Lecture Notes in Computer Science*, pages 201–215. Springer, 2023.
10. G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*. SIAM, 1999.
11. M. Neuts. Probability distributions of phase type. In *Liber Amicorum Prof. Emeritus H. Florin*, pages 173–206. University of Louvain, 1975.
12. M.F. Neuts. *Matrix Geometric Solutions in Stochastic Models*. Johns Hopkins University Press, Baltimore, 1981.
13. George M. Phillips. *Interpolation and Approximation by Polynomials*. Springer New York, 2003.
14. Alma Riska, Vesselin Diev, and Evgenia Smirni. An em-based technique for approximating long-tailed data sets with ph distributions. *Performance Evaluation*, 55(1-2):147–164, 2004.
15. Theodore J. Rivlin. *An Introduction to the Approximation of Functions*. Courier Corporation, 1981.
16. M. Telek and G. Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Performance Evaluation*, 64(9-12):1153–1168, Aug. 2007.