The Scale Factor: A New Degree of Freedom in Phase Type Approximation

A. Bobbio [a], A. Horváth [b], M. Telek [c]

[a] *Dip. di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale*

[b] *Dip. di Informatica, Università di Torino*

[c] *Dep. of Telecommunications, Budapest University of Technology and Economics*

Abstract

This paper introduces a unified approach to phase-type approximation in which the discrete and the continuous phase-type models form a common model set. The models of this common set are assigned with a non-negative real parameter, the *scale factor*. The case when the scale factor is strictly positive results in Discrete phase-type distributions and the scale factor represents the time elapsed in one step. If the scale factor is 0, the resulting class is the class of Continuous phase-type distributions. Applying the above view, it is shown that there is no qualitative difference between the discrete and the continuous phase-type models.

Based on this unified view of phase-type models one can choose the best phase-type approximation of a stochastic model by optimizing the scale factor.

Key words: Discrete and Continuous Phase-type distributions, Phase-type expansion, approximate analysis

1 Introduction

This paper presents new comparative results on the use of Discrete Phase Type (DPH) distributions [22] and of Continuous Phase Type (CPH) distributions [23] in applied stochastic modeling.

DPH distributions of order n are defined as the time to absorption in a Discrete-State Discrete-Time Markov Chain (DTMC) with n transient states

Email addresses: `bobbio@di.unito.it` (A. Bobbio), `horvath@di.unito.it` (A. Horváth), `telek@hit.bme.hu` (M. Telek).

and one absorbing state. CPH distributions of order n are defined, similarly, as the distribution of the time to absorption in a Discrete-State Continuous-Time Markov Chain (CTMC) with n transient states and one absorbing state. The above definition implies that the properties of a DPH distribution are computed over the set of the natural numbers while the properties of a CPH distribution are defined as a function of a continuous time variable t. When DPH distributions are used to model timed activities, the set of the natural numbers must be related to a time measure. Hence, a new parameter need to be introduced that represents the time span associated to each step. This new parameter is the scale factor of the DPH distribution, and can be viewed as a new degree of freedom, since its choice largely impacts the shape and properties of a DPH distribution over the continuous time axes. When DPH distributions are used to approximate a given continuous distribution, the scale factor affects the goodness of the fit.

The paper starts discussing to what extent DPH or CPH distributions can be utilized to fit a given continuous distribution. It is shown that a DPH distribution of any order converges to a CPH distribution of the same order as the scale factor goes to zero. Even so, the DPH class contains distributions whose behavior differs substantially from the one of the corresponding distributions in the CPH class. Two main peculiar points differentiate the DPH class from the CPH class. The first point concerns the coefficient of variation: indeed, while in the continuous case the minimum coefficient of variation is a function of the order only and its lower bound is given by the well known theorem of Aldous and Shepp [1], in the discrete case the minimum coefficient of variation is proved to depend

both on the order and on the mean (and hence on the scale factor) [25]. Furthermore, it is easy to see that for any order, there exist members of the DPH class that represent a deterministic value with a coefficient of variation equal to zero. Hence, for any order (greater than 1), the coefficient of variation of the DPH class spans from zero to infinity.

The second peculiar point that differentiate the DPH class is the support of the distributions. While a CPH distribution (of any order) has always an infinite support, there exist members of the DPH class of any order that have a finite support (between a minimum non-negative value and a maximum) or have a mass equal to one concentrated in a single value (deterministic distribution).

It turns out that the possibility of

- tuning the scale factor to optimize the goodness of the fit,
- having distributions with coefficient of variation spanning from 0 to infinity,
- representing deterministic values exactly,
- coping with finite support distributions,

makes the DPH class a very interesting and challenging class of distributions to be explored in applied stochastic models. The purpose of this paper is to show how these favorable properties can be exploited in practice, and to provide guidelines to the modeler to a reasonably good choice of the distributions to be used. Indeed, since a DPH distribution tends to a CPH distribution as the scale factor approaches zero, considering the scale factor as a new decision variable in a fitting experiment, and finding the value of the optimal scale factor (with respect to some error measure) provides a valuable tool to decide whether to use a discrete or a continuous approximation to the given problem.

The fitting problem for the CPH class has been extensively studied and reported in the literature by resorting to a variety of structures and numerical techniques [2,3,9,18,24]. Conversely, the fitting problem for the DPH class has received very little attention [6].

In recent years, a considerable effort has been devoted to define models with generally distributed timings [10,7] and to merge in the same model random variables and deterministic duration [21]. Analytical solutions are possible in special cases, and the approximation of the original problems by means of CPH distributions is a rather well known technique [12,8]. This paper is aimed at emphasizing that DPH approximation may provide a more convenient alternative with respect to CPH approximation, and also to provide a way to quantitatively support this choice. Furthermore, the use of DPH approximation can be extended from stochastic models to functional analysis where time intervals with nondeterministic choice are considered [5,4]. Finally, discretization techniques for continuous problems [13,14] can be restated in terms of DPH approximations.

The rest of the paper is organized as follows. After defining the notation to be used in the paper in Section 2, Section 3 discusses the peculiar properties of the DPH class with respect to the CPH class. Some guidelines for bounding the parameters of interest and extensive numerical experiments to show how the goodness of the fit is influenced by the optimal choice of the scale factor are reported in Section 4. Section 5 discusses the quality of the approximation when passing from the analysis of a single distribution to the analysis of performance measures in complete non-Markovian stochastic models. The paper is concluded in Section 6.

2 Definition and Notation

A DPH distribution [22,23] is the distribution of the time to absorption in a DTMC with n transient states, and one absorbing state numbered $(n + 1)$. The one-step transition probability matrix of the corresponding DTMC can be partitioned as:

$$\widehat{\mathbf{B}} = \begin{bmatrix} \mathbf{B} & \mathbf{b} \\ \mathbf{0} & 1 \end{bmatrix} \tag{1}$$

where $\mathbf{B} = [b_{ij}]$ is the $(n \times n)$ matrix collecting the transition probabilities among the transient states, $\mathbf{b} = [b_{i,n+1}]^T$ is the column vector of length n grouping the probabilities from any state to the absorbing one, and $\mathbf{0} = [0]$ is the zero vector. The initial probability vector $\widehat{\alpha} = [\alpha, \alpha_{n+1}]$ is of length $(n+1)$, with $\sum_{j=1}^{n} \alpha_j = 1 - \alpha_{n+1}$. In the present paper, we consider only the class of DPH distributions for which $\alpha_{n+1} = 0$, but the extension to the case when $\alpha_{n+1} > 0$ is straightforward. The tuple $(\alpha, \mathbf{B})$ is called the representation of the DPH distribution, and n the order.

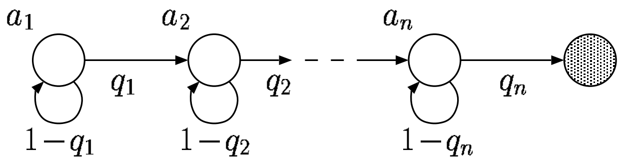
Similarly, a CPH distribution [23] is the distribution of the time to absorption in a CTMC with n transient states, and one absorbing state numbered $(n+1)$. The infinitesimal generator $\widehat{\mathbf{Q}}$ of the CTMC can be partitioned in the following way:

$$\widehat{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{q} \\ \mathbf{0} & 1 \end{bmatrix} \tag{2}$$

where, $\mathbf{Q}$ is a $(n \times n)$ matrix that describes the transient behavior of the CTMC and $\mathbf{q}$ is the column vector grouping the transition rates to the absorbing state. Let $\hat{\alpha} = [\,\alpha,\ \alpha_{n+1}\,]$ be the $(n+1)$ initial probability (row) vector with $\sum_{i=1}^{n} \alpha_i = 1 - \alpha_{n+1}$. The tuple $(\alpha,\ \mathbf{Q})$ is called the representation of the CPH distribution, and n the order.

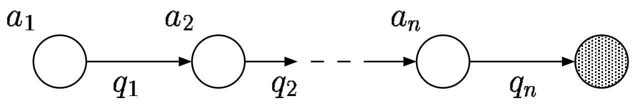
It has been shown in [6] for the discrete case and in [11] for the continuous case that the representations in (1) and (2), because of their too many free parameters, do not provide a convenient form for running a fitting algorithm. Instead, resorting to acyclic phase-type distributions, the number of free parameters is reduced significantly since both in the discrete and the continuous case a canonical form can be used. The canonical form and its constraints for the discrete case [6] is depicted in Figure 1. Figure 2 gives the canonical form and associated constraints for the continuous case. In both cases the canonical form corresponds to a mixture of Hypo-exponential distributions.

A fitting algorithm that provides acyclic CPH, acyclic DPH distributions has been provided in [2,3] and [6], respectively. Experiments suggests (an exhaustive comparison of fitting algorithms can be found in [20]) that, from the point of view of applications, it is reasonable to restrict our attention to the class of acyclic phase-type distributions.



$$\sum_{i=1}^{n} a_i = 1, \;\; 0 < q_i \le q_{i+1} \le 1, \;\; 1 \le i \le n-1$$

Fig. 1. Canonical representation of acyclic DPH distributions and its constraints



$$\sum_{i=1}^{n} a_i = 1, \;\; 0 < q_i \le q_{i+1}, \;\; 1 \le i \le n-1$$

Fig. 2. Canonical representation of acyclic CPH distributions and its constraints

3 Comparing properties of CPH and DPH distributions

CTMC are defined as a function of a continuous time variable t, while DTMC are defined over the set of the natural numbers. In order to relate the number of jumps in a DTMC with a time measure, a time span must be assigned to each step. Let δ be (in some arbitrary units) the scale factor, i.e. the time span assigned to each step. The value of δ establishes an equivalence between the sentence "probability at the k-th step" and "probability at time $k\,\delta$", and hence, defines the time scale on which the properties of the DTMC are measured. The consideration of the scale factor δ introduces a new parameter, and consequently a new degree of freedom, in the DPH class with respect to the CPH class. In the following, we discuss how this new degree of freedom impacts the properties of the DPH class and how it can be exploited in practice.

Let u be an "unscaled" DPH distributed random variable (r.v.) of order n with representation $(\alpha,\, \mathbf{B})$, defined over the set of the non-negative natural numbers. Let us consider a scale factor δ; the scaled r.v. $\tau = \delta\, u$ is defined over the discrete set of time points $(1\,\delta,\, 2\,\delta,\, 3\,\delta,\, \ldots,\, k\,\delta,\, \ldots)$, being k a non-negative natural number. For the unscaled and the scaled DPH r.v. the following equations hold.

$$\begin{aligned}
&F_u(k) = Pr\{u \le k\} = 1 - \alpha \mathbf{B}^k\, \mathbf{e} \\
&F_\tau(\delta k) = Pr\{\tau \le \delta k\} = 1 - \alpha \mathbf{B}^k\, \mathbf{e} \\
&m_u^i = E(u^i) \\
&m_\tau^i = E(\tau^i) = \delta^i\, E(u^i) \qquad i \ge 1,
\end{aligned} \tag{3}$$

where $\mathbf{e}$ is the column vector of ones, and $E(u^i)$ is the i-th moment calculated from the factorial moments of u: $E(u(u-1)\dots(u-i+1)) = i!\,\alpha(\mathbf{I}-\mathbf{B})^{-i}\mathbf{B}^{i-1}\mathbf{e}$. It is evident from (3) that the mean m_τ of the scaled r.v. τ is δ times the mean m_u of the unscaled r.v. u. While m_u is an invariant of the representation $(\alpha, \mathbf{B})$, δ is a free parameter; adjusting δ, the scaled r.v. can assume any mean value $m_\tau \geq 0$. On the other hand, one can easily infer from (3) that the coefficients of variation of τ and u are equal. A consequence of the above properties is that one can easily provide a scaled DPH of order ≥ 2 with arbitrary mean and arbitrary coefficient of variation with an appropriate scale factor. Or more formally: the unscaled DPH r.v. u of any order $n > 1$ can exhibit a coefficient of variation between $0 \leq cv_u^2 \leq \infty$. For $n = 1$ the coefficient of variation ranges between $0 \leq cv_u^2 \leq 1$.

As mentioned earlier, an important property of the DPH class with respect to the CPH class is the possibility of exactly representing a deterministic delay. A deterministic distribution with value a can be realized by means of a scaled DPH distribution with n phases with scale factor δ if $n = a/\delta$ is integer. In this case, the structure of the DPH distribution is such that phase i is connected with probability 1 only to phase $i+1$ ($i = 1, \dots, n$), and with an initial probability concentrated in state 1. If $n = a/\delta$ is not integer for the given δ, the deterministic behavior can only be approximated.

3.1 First order discrete approximation of CTMCs

Given a CTMC with infinitesimal generator $\widetilde{\mathbf{Q}}$, the transition probability matrix over an interval of length δ can be written as:

$$e^{\widetilde{\mathbf{Q}}\delta} = \sum_{i=0}^{\infty} (\widetilde{\mathbf{Q}}\delta)^i / i! = \mathbf{I} + \widetilde{\mathbf{Q}}\delta + \sigma(\delta),$$

hence the first order approximation of $e^{\widetilde{\mathbf{Q}}\delta}$ is matrix $\mathbf{\Pi}(\delta) = \mathbf{I} + \widetilde{\mathbf{Q}}\delta$. $\mathbf{\Pi}(\delta)$ is a proper stochastic matrix if $\delta < 1/q$, where $q = \max_{i,j} |\widetilde{\mathbf{Q}}_{ij}|$. $\mathbf{\Pi}(\delta)$ is the exact transition probability matrix of the CTMC assumed that at most one transition occurs in the interval of length δ.

We can approximate the behavior of the CTMC at time $(\delta, 2\delta, 3\delta, \dots, k\delta, \dots)$ using the DTMC with transition probability matrix $\mathbf{\Pi}(\delta)$. The approximate transition probability matrix at time $t = k\delta$ is:

$$\mathbf{\Pi}(\delta)^k = (\mathbf{I} + \widetilde{\mathbf{Q}}\delta)^{\frac{t}{\delta}}$$

The following theorem proves the property that the above first order approximation becomes exact as $\delta \to 0$.

Theorem 1 *As the length of the interval of the first order approximation, δ, tends to 0, such that $t = k\delta$ the approximate transition probability matrix tends to the exact one.*

Proof: The scalar version of the applied limiting behavior is well-known in the form

$$\lim_{x \to 0} (1 + ax)^{\frac{1}{x}} = e^a .$$

Since matrices $\mathbf{I}$ and $\widetilde{\mathbf{Q}}$ commute we can obtain the matrix version of the same expression as follows

$$\lim_{\delta \to 0} (\mathbf{I} + \widetilde{\mathbf{Q}}\delta)^{\frac{t}{\delta}} = \lim_{k \to \infty} \mathbf{\Pi}(t/k)^k = \lim_{k \to \infty} \left(\mathbf{I} + \frac{\widetilde{\mathbf{Q}}t}{k} \right)^k =$$

$$\lim_{k \to \infty} \sum_{j=0}^{k} \begin{pmatrix} k \\ j \end{pmatrix} \left(\frac{\widetilde{\mathbf{Q}}t}{k} \right)^j =$$

$$= \lim_{k \to \infty} \sum_{j=0}^{k} \frac{(\widetilde{\mathbf{Q}}t)^j}{j!} \frac{k!}{k^j \, (k-j)!} = \sum_{j=0}^{\infty} \frac{(\widetilde{\mathbf{Q}}t)^j}{j!} = e^{\widetilde{\mathbf{Q}}t} .$$

□

An obvious consequence of Theorem 1 for PH distributions is given in the following corollary.

Corollary 1 *Given a scaled DPH distribution of order n, representation (α, $\mathbf{I} + \mathbf{Q}\delta$) and scale factor δ, the limiting behavior as $\delta \to 0$ is the CPH distribution of order n with representation (α, $\mathbf{Q}$).*

3.2 The minimum coefficient of variation

It is known that one of the main limitation in approximating a given distribution by a PH one is the attainable minimal coefficient of variation, cv^2_{min}. In order to discuss this point, we recall two theorems that state the cv^2_{min} for the class of CPH and DPH distributions.

Theorem 2 (Aldous and Shepp [1]) *The cv^2_{min} of a* CPH *distributed r.v. of order n is $cv^2_{min} = 1/n$ and is attained by the Erlang(n) distribution independent of its mean m_c or of its parameter $\lambda = n/m_c$.*

The corresponding theorem for the unscaled DPH class has been proved in [25]. In the following, $\lfloor x \rfloor$ denotes the integer part and $\langle x \rangle$ denotes the fractional part of x.

Theorem 3 *The cv^2_{min} of an unscaled DPH r.v. of order n and mean m_u is:*

$$\begin{array}{ll} \dfrac{\langle m_u \rangle (1 - \langle m_u \rangle)}{m_u^2} & \text{if } \ m_u \le n \ , \\ \dfrac{1}{n} - \dfrac{1}{m_u} & \text{if } \ m_u > n \ , \end{array} \tag{4}$$

The unscaled DPH r.v. which exhibits this minimal coefficient of variation has the following canonical structure:

- *if $m_u \le n$: each state is connected to the next with probability 1 and the nonzero initial probabilities are $\alpha_{n-\lfloor m_u \rfloor} = \langle m_u \rangle$ and $\alpha_{n-\lfloor m_u \rfloor+1} = 1 - \langle m_u \rangle$ (Figure 3);*
- *if $m_u > n$: each state is connected to the next with probability n/m_u and the only nonzero initial probability is $\alpha_1 = 1$ (Figure 4).*

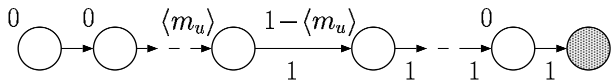


Fig. 3. $m_u \le n$

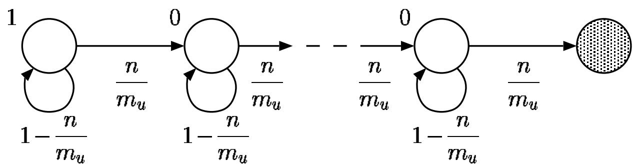


Fig. 4. $m_u > n$

Implications of the above theorems for what concerns the ability of approximating distributions with low coefficient of variation are drawn in [6].

3.3 The minimum coefficient of variation of scaled DPH distributions

For scaled DPH distribution Theorem 3 can be restated as follows.

Theorem 4 *The cv^2_{min} of a scaled DPH r.v. of order n with scale factor δ*

and mean $m_\tau = \delta\, m_u$ is:

$$
\begin{cases}
\dfrac{\left\langle \frac{m_\tau}{\delta} \right\rangle \left(1 - \left\langle \frac{m_\tau}{\delta} \right\rangle\right)}{\left(\frac{m_\tau}{\delta}\right)^2} & \text{if } m_\tau \le n\,\delta\ , \\[2ex]
\dfrac{1}{n} - \dfrac{\delta}{m_\tau} & \text{if } m_\tau > n\,\delta\ ,
\end{cases}
\tag{5}
$$

The scaled DPH r.v. which exhibits the cv^2_{min} has the same structure of Figures (3) and (4), as in the unscaled case (see Theorem 3).

Corollary 2 *For finite mean m_τ, as $\delta \to 0$ only the second part of (5) remains effective, and $cv^2_{min} \to 1/n$ as $\delta \to 0$.*

Corollary 2 proves that the cv^2_{min} of the DPH class converges to the cv^2_{min} of the CPH class of the same order as $\delta \to 0$. The following corollary presents a much stronger convergence result for the case of approximating distributions with low coefficient of variation. It is about the convergence of the distributions. In the corollary and its proof the term *best cv fitting* PH approximation will be used which refers to the PH approximation that exhibits the same mean and provides the closest approximation for the 2nd moment.

Corollary 3 *The* best *cv* fitting *scaled DPH approximation of distributions with low coefficient of variation converges, in distribution, to the* best *cv* fitting *CPH approximation of the same distribution as δ tends to* 0.

Proof: Both the CPH and the DPH classes have limits in approximating distributions with low coefficient of variation. As given in Theorem 2 and 3, the *best cv fitting* approximation of a distribution with coefficient of variation less than these limits is the Erlang(n) distribution in both the continuous and the discrete case (assuming that $m_\tau > n\delta$).

The representation (α, $\mathbf{Q}$) of the continuous Erlang(n) with mean m_τ and the representation (α, $\mathbf{B}$) of the discrete Erlang(n) with mean m_τ, scale factor δ

are:

$$\alpha = \{1, 0, \ldots, 0\},\ \mathbf{Q} = \boxed{\begin{matrix} -\frac{n}{m_\mathcal{T}} & \frac{n}{m_\mathcal{T}} & 0 & \ldots & 0 \\ 0 & -\frac{n}{m_\mathcal{T}} & \frac{n}{m_\mathcal{T}} & \ldots & \\ & & \ddots & & \\ 0 & & & \ldots & -\frac{n}{m_\mathcal{T}} \end{matrix}}$$

$$\alpha = \{1, 0, \ldots, 0\},\ \mathbf{B} = \boxed{\begin{matrix} 1-\frac{n\delta}{m_\mathcal{T}} & \frac{n\delta}{m_\mathcal{T}} & 0 & \ldots & 0 \\ 0 & 1-\frac{n\delta}{m_\mathcal{T}} & \frac{n\delta}{m_\mathcal{T}} & \ldots & \\ & & \ddots & & \\ 0 & & & \ldots & 1-\frac{n\delta}{m_\mathcal{T}} \end{matrix}}$$

Note that $\mathbf{B} = \mathbf{I} - \mathbf{Q}\delta$ and Corollary 3 follows from Corollary 1. □

In this particular case, when the structure of the *best cv fitting* scaled DPH and CPH distributions are known, we can show that the distribution of the *best cv fitting* scaled DPH distribution converges to the distribution of the *best cv fitting* CPH distribution when $\delta \to 0$. Unfortunately, the same convergence property cannot be proved in general, since the structural properties of the *best cv fitting* PH distributions are not known and they depend on the chosen (arbitrary) optimization criterion. Instead, in Section 4 we provide an extensive experimental study on the behavior of the *best cv fitting* scaled DPH and CPH distributions as a function of the scale factor δ .

3.4 DPH distributions with finite support

Another peculiar characteristic of the DPH class is to contain distributions with finite support. A DPH distribution has finite support if its structure does not contain cycles and self-loops (any cycle or self loop implies an infinite support).

Let $[a, b]$ be the finite support of a given distribution, with $a, b \geq 0$ and $a \leq b$ (when $a = b$ the finite support distribution reduces to a deterministic distribution with mass 1 at $a = b$). If a/δ and b/δ are both integers, it is possible to construct a scaled DPH of order b/δ for which the probability mass function has non-zero elements only for the values $a, a+\delta, a+2\delta, ..., b$. As an example, the discrete uniform distribution between $a = 2$ and $b = 6$ is reported in Figure 5, for scale factor $\delta = 1$.

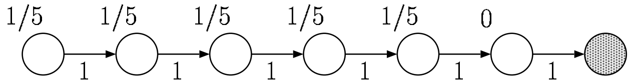


Fig. 5. DPH representation of the discrete uniform distribution $[a = 2, b = 6]$

4 The optimal δ in PH fitting

The scale factor δ provides a new degree of freedom in fitting, and, furthermore, since the limit of a DPH distribution for $\delta \to 0$ is a CPH distribution, the optimization of the scale factor in a fitting problem provides a quantitative way to decide whether a continuous or a discrete approximation performs better in the given problem. Hence, assuming δ as a decision variable, we can consider the CPH and the DPH class as a unique model set in which the choice among DPH or CPH classes is given by the optimal value of δ.

Let X be the continuous r.v. to be fit by a PH distribution, and let $F_X(x)$ be its cdf, $E(X^i)$ the i-th moment and $cv^2(X)$ the squared coefficient of variation. In order to define a fitting procedure, a distance measure between X and the approximating PH distribution needs to be defined. Then, the fitting algorithm provides the PH distribution which minimizes the chosen distance measure. The minimization of the measure is performed as follows. Starting from the initial guess, the non-linear optimization problem is solved by an iterative linearization method. In each step the partial derivatives with respect to the parameters of the PH distribution are computed numerically. Then, the simplex method is applied to determine the direction in which the distance measure decreases optimally. The constraints of the linear programming is given by probabilistic constraints, by the restriction on the structure of the PH distribution and by confining the change of parameters. (More detailed description of the fitting procedure, which is implemented in the tool PhFit [17], is given in [16].)

In order to compare, in a unified framework, the goodness of the approximation reached by CPH and DPH distributions, we need to choose a distance measure that is meaningful and applicable both in the continuous as well as in the discrete setting. The selected distance measure is the squared area difference between the original cdf $F(\cdot)$ and the approximating cdf $\hat{F}(\cdot)$:

$$D = \int_x (F(x) - \hat{F}(x))^2 dx \qquad (6)$$

The distance measure D is easily applicable for any combination of discrete and continuous distributions. All the numerical experiments reported in the sequel are based on the minimization of the area difference given in (6).

4.1 Fitting distributions with low cv^2

The following considerations provide practical upper and lower bounds to guide in the choice of a suitable scale factor δ, and are mainly based on the dependence of the minimal coefficient of variation of a scaled DPH distribution on the order n and on the mean m_τ.

Since we only consider DPH distributions with no mass at zero, the mean of any unscaled DPH distribution is greater than 1. This means that δ should be less than $E(X)$. However, a more convenient upper bound that exploits the flexibility associated with the n phases, is given by:

$$\delta \leq \frac{E(X)}{n-1} \; . \tag{7}$$

If the squared coefficient of variation of the distribution to be approximated is less than $1/n$, δ should satisfy the following relation (see Theorem 3):

$$\delta > \left(\frac{1}{n} - cv^2(X)\right) E(X) \tag{8}$$

Let X be a Lognormal r.v. with parameters $(1, 0.2)$, whose mean is $E(X) = 1$ and $cv^2(X) = 0.0408$ (this distribution is the distribution L3 taken from the benchmark examined in [9,6], hence we refer to it as L3). Table 1 reports the lower and upper bounds of δ, with $n = 2$, 4, 8, 12, computed from (8) and (7).

The cdf and pdf of the approximating CPH and DPH distributions of order $n = 10$, with different scale factors δ, are presented in Figure 6. When considering the approximate DPH distribution, the $f(x)$ values are calculated at the discrete points $(\delta, 2\delta, 3\delta, \ldots, k\delta, \ldots)$ to which the following mass is assigned:

$$f(k\delta) = 1/\delta(F(k\delta) - F((k-1)\delta)) \tag{9}$$

For the ease of visual interpretation the points are connected with a line.

When δ is less than its lower bound the required cv^2 cannot be attained; when δ becomes too large the wide separation of the discrete steps increases the approximation error; when δ is in the proper range (e.g. $n = 10$; $\delta = 0.06$) a reasonably good fit is achieved. This example also suggests that an optimal value of δ exists that minimizes the chosen distance measure D in (6).

In order to display the goodness of fit for the L3 distribution, Figure 7 shows the distance measure D as a function of δ for various values of the order n. A minimum value of D is attained in the range where the parameters fit the

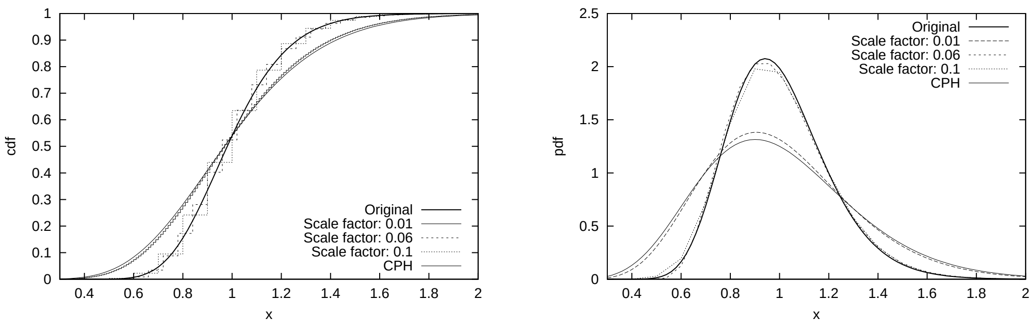


Fig. 6. Approximating the L3 distribution with 10-phase PH approximations

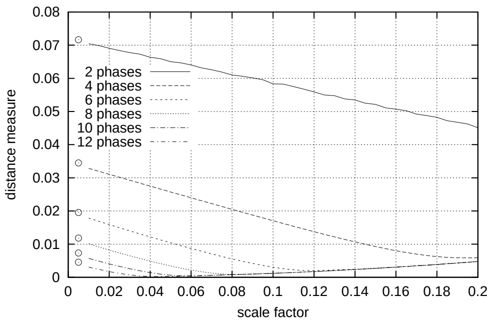


Fig. 7. Distance measure as the function of the scale factor δ for low cv^2 (L3)

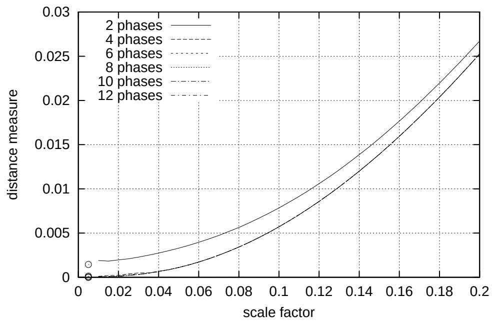


Fig. 8. Distance measure as the function of the scale factor δ for high cv^2 (L1)

bounds of Table 1. Notice also that, as δ increases, the advantage of having more phases disappears, according to Theorem 4. The circles in the left part of this figure (as well as in all the successive figures) indicate the corresponding distance measure D obtained from CPH fitting. The figure (and the subsequent ones as well) suggests that the distance measure obtained from DPH fitting converges to the distance measure obtained by the CPH approximation as δ tends to 0.

The lowest applied scale factor is 0.01. Lower values of the scale factor renders the fitting procedure numerically unstable because the diagonal elements of the transition matrix are close to 1.

n	lower bound of δ equation (8)	upper bound of δ equation (7)
4	0.2092	0.333
8	0.0842	0.1428
12	0.0425	0.0909
16	0.0217	0.0666

Table 1
Upper and lower bound of δ for fitting distribution L3

4.2 Fitting distributions with high cv^2

We have seen in the previous subsections that it is beneficial to approximate distributions with a low coefficient of variation by means of a DPH distributions. In this subsection, we investigate the optimal value of δ when fitting distributions with a high coefficient of variation.

Let X be a Lognormal r.v. with parameters $(1, 1.8)$ (this is the distribution L1 taken from the benchmark in [9,6]). For X we have $E(X) = 1$ and $cv^2(X) = 24.534$. Figure 8 shows the measure of the goodness of fit as a function of δ for various orders n (the cases when the number of phases are greater than 2 result in practically the same goodness of fit). The distance measures D decreases as $\delta \to 0$ indicating that the optimal fitting is achieved by applying CPH distribution. This example suggests that, for distributions with infinite support and $cv^2(X) > 1/n$, the optimal value of δ tends to 0, implying that the best fit is obtained by a CPH. However, this conclusion might not be true for distributions with finite support, as it is explored in the next subsection.

4.3 Fitting distributions with finite support

In this case, two features must be considered, namely the cv^2 and the maximum value of the finite support. It should be stressed that the chosen distance measure D in (6) can be considered as not completely appropriate in the case of finite support, since it does not force the approximating PH to have its mass confined in the finite support and 0 outside.

Let X be a uniform r.v. over the interval $[1,\ 2]$, with $E(X) = 1.5$ and $cv^2(X) = 0.0370$ (this is the distribution U2 taken from the benchmark in [9,6]). Figure 9 shows the distance measure as a function of δ for various orders n. It is evident that, for each n, a minimal value of δ is obtained, that provides the best approximation according to the chosen distance measure.

As a second example, let X be a uniform r.v. over the interval $[0,\ 1]$, with $E(X) = 0.5$ and $cv^2(X) = 0.333$ (this is the distribution U1 taken from the benchmark in [9,6]). Figure 10 shows the distance measure as a function of δ for various orders n. Since, in this example $cv^2(X) = 0.333$, an order $n = 3$ is large enough for a CPH to attain the coefficient of variation of the distribution. Nevertheless, the optimal δ in Figure (10), which minimizes the distance measure D for high order PH ($n > 2$), ranges between $\delta = 0.02$ and $\delta = 0.05$, thus leading to the conclusion that a DPH provides a better fit. This example evidences that the coefficient of variation is not the only factor which influences the optimal δ value. The shape of the distribution plays an essential role as well. Our experiments show that a discontinuity in the pdf (or in the

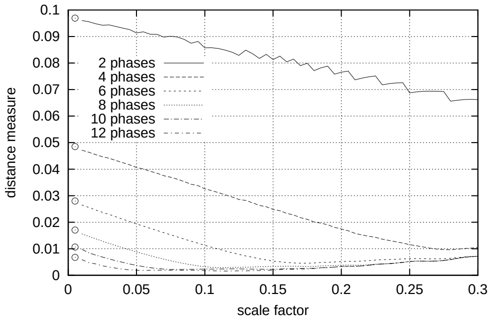


Fig. 9. Distance measure as the function of the scale factor δ for Uniform(1,2) (U2)

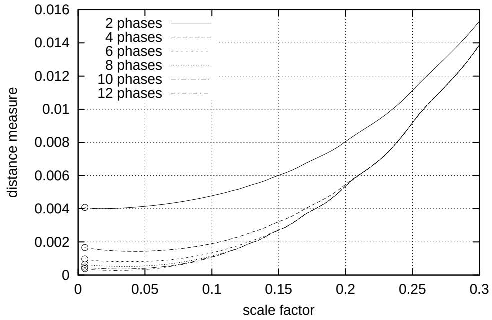


Fig. 10. Distance measure as the function of the scale factor δ for Uniform(0,1) (U1)

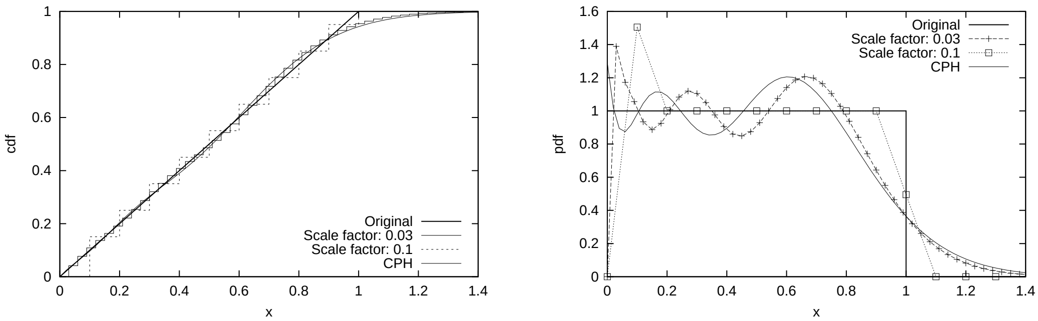


Fig. 11. Approximating the Uniform $(0, 1)$ distribution (U1)

cdf) is hard to approximate with CPH, hence in the majority of these cases DPH provides a better approximation.

Figure 11 shows the cdf and the pdf of the U1 distribution, compared with the best fit PH approximations of order $n = 10$, and various scale factors δ. In the case of DPH approximation, the $f(x)$ values are calculated as in (9). With respect to the chosen distance measure, the best approximation is obtained for $\delta = 0.03$, which corresponds to a DPH distribution with infinite support . When $\delta = 0.1$ the approximate distribution has a finite support. Hence, the value $\delta = 0.1$ (for $n = 10$) provides a DPH able to represent the logical property that the random variable is less than 1. Another fitting criterion may, of course, stress this property.

4.4 Approximating exponential distributions

When it comes to discrete approximation of continuous models, exponential durations are usually approximated by a discrete PH distribution with a single phase (see, for example, [15]). This approximation, however, is precise only if the scale factor is small enough. Figure 12 shows the goodness of fitting as function of the scale factor for the exponential distribution with mean equal

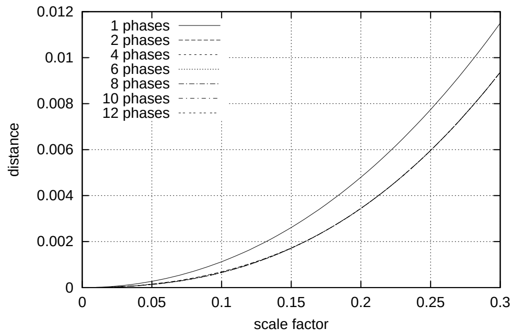


Fig. 12. Distance measure as the function of the scale factor for the Exponential(1.0) distribution

to 1.0. The figure suggests that it is not worth to use more than 2 phases to approximate exponential distributions. (Note that this statement might not be true in case of other distance measures.)

5 Approximating non-Markovian models

Section 4 has explored the problem of how to find the best fit among either a DPH or a CPH distribution by tuning the scale factor δ. When dealing with a stochastic model of a system that incorporates non exponential distributions, a well-known solution technique consists in a markovianization of the underlying non-Markovian process by substituting the non exponential distribution with a best fit PH distribution, and then expanding the state space. A natural question arises also in this case, on how to decide between discrete (using DPH) and continuous (using CPH) approximation, in order to minimize the error in the performance measures we are interested in for the overall model.

One possible way to handle this problem could consist in finding the best PH fit for any single distribution and to plug them in the model. In the present paper, we only consider the case where the PH distributions are either all discrete with the same scale factor or they are all continuous. With this limitation some of the freedom in choosing the scale factor disappears because a single scale factor is chosen for the whole model. Various embedding techniques have been explored in the literature for mixing DPH with different scale factors and CPH ([13,19]), but these techniques are out of the scope of the paper.

In order to quantitatively evaluate the influence of the scale factor on some performance measures defined at the system level, we have considered a preemptive M/G/1/2/2 queue with two classes of customers. We have chosen this example because accurate analytical solutions are available both in transient condition and in steady-state using the methods presented in e.g.[13]. The general distribution G is taken from the set of distributions (L1, L3, U1, U2)

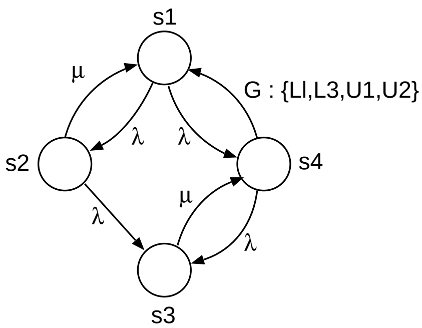


Fig. 13. The state space of the considered M/G/1/2/2 queue

already considered in the previous section.

Customers arrive at the queue with rate $\lambda = 0.5$ in both classes. The service time of a higher priority job is exponentially distributed with parameter $\mu = 1$. The service time distribution of the lower priority job is either L1, L3, U1 or U2. Arrival of a higher priority job preempts the lower priority one. The policy associated to the preemption of the lower priority job is preemptive repeat different (prd), i.e. after the departure of the higher priority customer the service of the low priority customer starts from the beginning with a new service time sample.

The system has 4 states (Figure 13): in state *s1* the server is empty, in state *s2* a higher priority customer is under service with no lower priority customer in the system, in state *s3* a higher priority customer is under service with a lower priority customer waiting, in state *s4* a lower priority job is under service (in this case there cannot be a higher priority job).

5.1 Steady state behavior

Let p_i $(i = 1, \ldots, 4)$ denote the steady state probability of the M/G/1/2/2 queue obtained from an exact analytical solution.

In order to evaluate the correctness of the PH approximation we have solved the model by substituting the original general distribution (either L1, L3, U1 or U2) with approximating DPH or CPH distributions. (The fitting was performed applying the same distance measure as before, it is given in (6)). Let $\hat{p}_i$ $(i = 1, \ldots, 4)$ denote the steady state probability of the M/PH/1/2/2 queue with the PH approximation.

The overall approximation error is measured in terms of the difference between the exact steady state probabilities p_i and the approximate steady state probabilities $\hat{p}_i$. Two error measures are defined:

$$\epsilon_{SUM} = \sum_i \frac{|p_i - \hat{p}_i|}{p_i} \quad \text{and} \quad \epsilon_{MAX} = \max_i \frac{|p_i - \hat{p}_i|}{p_i}.$$

All the experiments were performed applying both 1-phase and 2-phase approximation for the exponential durations. The evaluated numerical values for ϵ_{SUM} and ϵ_{MAX} are reported in Figures 14-15 and 16-17 for the distribution L3. Since the behavior of ϵ_{MAX} is very similar to the behavior of ϵ_{SUM} in all the cases, for the other cases we report ϵ_{SUM} only (L1: Figures 18-19, U1: Figures 20-21, U2: Figures 22-23).

Figures 24 and 25 report two cases when both the service times are non-exponential. In the first case both the service distributions are L3, while in the second case, higher (lower) priority costumers are served according to distribution L3 (L1).

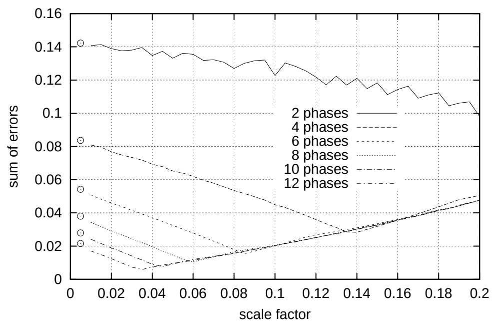


Fig. 14. ϵ_{SUM} for distribution L3 with 1-phase appr. for exponential durations

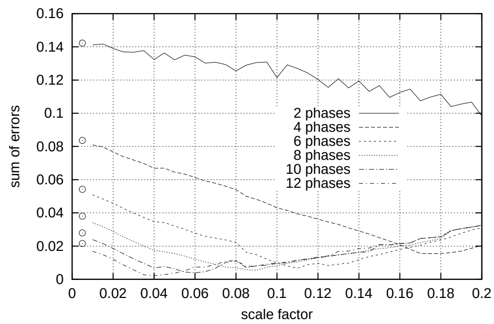


Fig. 15. ϵ_{SUM} for distribution L3 with 2-phase appr. for exponential durations

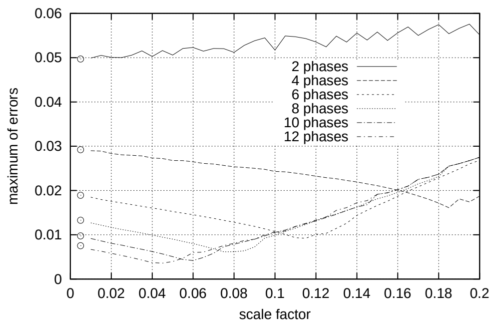


Fig. 16. ϵ_{MAX} for distribution L3 with 1-phase appr. for exponential durations

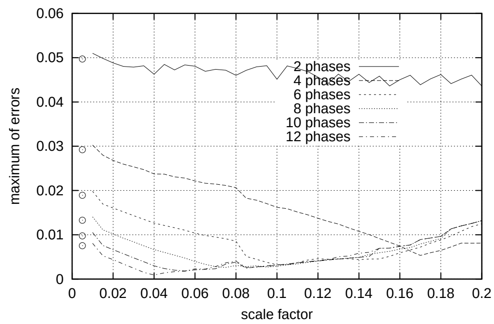


Fig. 17. ϵ_{MAX} for distribution L3 with 2-phase appr. for exponential durations

Based on the above experiments, it is rather hard to draw general guidelines for the choice of the scale factor. Some rather general observations, however,

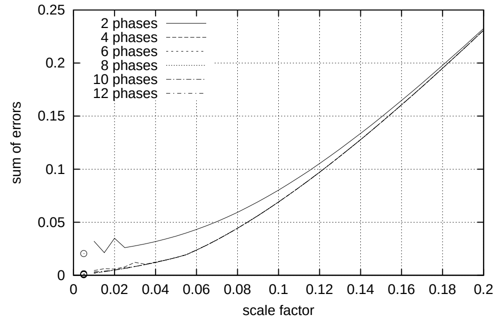


Fig. 18. ϵ_{SUM} for distribution L1 with 1-phase appr. for exponential durations

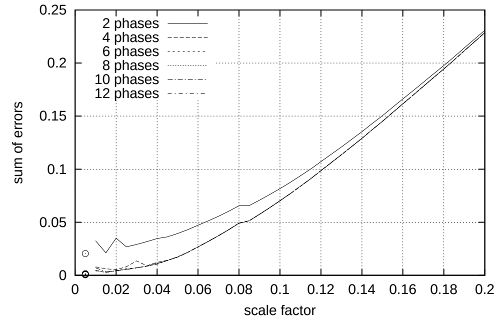


Fig. 19. ϵ_{SUM} for distribution L1 with 2-phase appr. for exponential durations

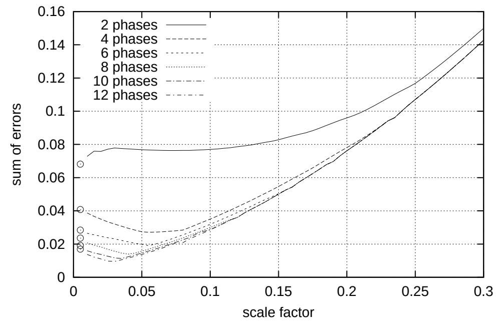


Fig. 20. ϵ_{SUM} for distribution U1 with 1-phase appr. for exponential durations

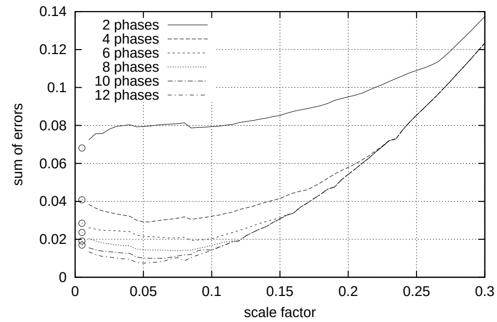


Fig. 21. ϵ_{SUM} for distribution U1 with 2-phase appr. for exponential durations

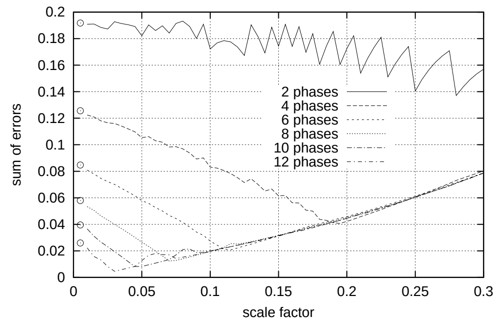


Fig. 22. ϵ_{SUM} for distribution U2 with 1-phase appr. for exponential durations

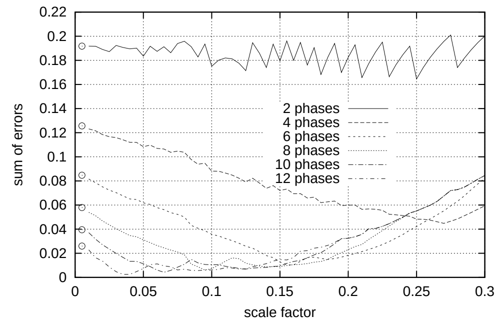


Fig. 23. ϵ_{SUM} for distribution U2 with 2-phase appr. for exponential durations

can be made:

- If all the optimal scale factors are 0, the approximation error in the evaluation of the performance indices of the global model can be minimized by resorting to continuous PH type approximation. See, for example, the case when the only non-exponential distribution in the model is L1 (Figure 18).
- When some of the individual optimal scale factors are 0, while others are not (this is the case, for example, when there are exponential durations and durations with low coefficient of variation in the same model, Figure 14), there is an optimal scale factor value (or a range) that gives optimal

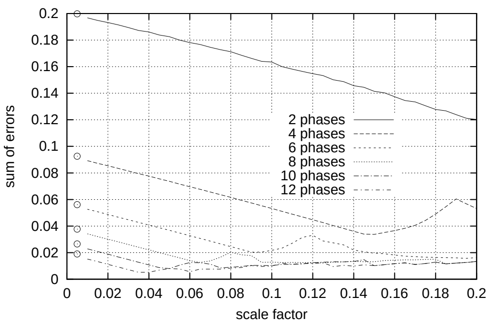


Fig. 24. ϵ_{SUM} for service times L3-L3 with 1-phase appr. for exponential durations

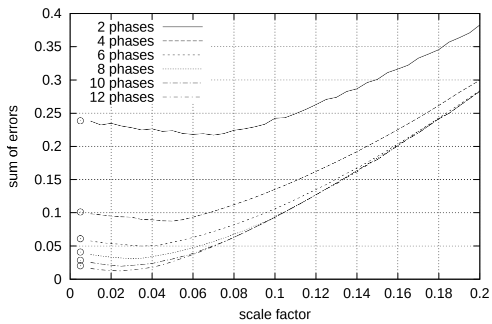


Fig. 25. ϵ_{SUM} for service times L1-L3 with 1-phase appr. for exponential durations

results for what concerns performance measures of the global model. In this case, on the one hand, decreasing the scale factor worsen the precision of the approximation because with lower scale factor durations with low coefficient of variation are approximated worse. On the other hand, increasing the scale factor worsen the approximation because the approximation of the exponential durations are getting less accurate. (See Figure 14.)

- In the cases when one of the individual optimal scale factors is rather far from 0, there is a wide range of scale factor values for which the discrete approximation yields better results than the continuous one. See, for example, Figure 22 where in case of 10 phases the discrete approximation gives better results in the range [0,0.2].
- Applying 2-phase discrete PH approximations for the exponential durations slows down the rate at which the approximation error grows with increasing scale factor. See, for example, Figure 14 compared to Figure 15.

5.2 Transient behavior

In order to investigate the approximation error in the transient behavior, we have considered distribution U2 for the service time and we have computed the transient probability of state s_1 with two different initial conditions. Figure 26 depicts the transient probability of state s_1 with initial state s_1. Figure 27 depicts the transient probability of the same state, s_1, when the service of a lower priority job starts at time 0 (the initial state is s_4). All approximations are with DPH distributions of order $n = 10$. Only the DPH approximations are depicted because the CPH approximation is very similar to the DPH one with scale factor $\delta = 0.03$. In the first case, (Figure 26), the scale factor $\delta = 0.03$, which was the optimal one from the point of view of fitting the single distribution in isolation, provides the most accurate results for the transient analysis as well. Instead, in the second case, the approximation with a scale factor $\delta = 0.2$ captures better the sharp change in the transient probability. Moreover, this value of δ is the only one among the values reported in the figure

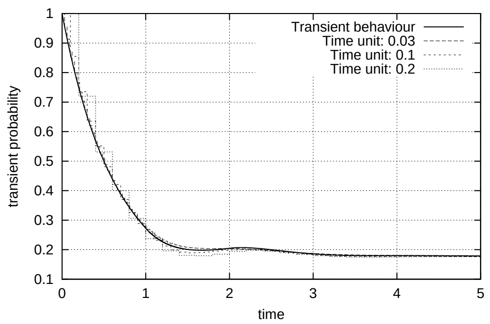


Fig. 26. Approximating transient probabilities

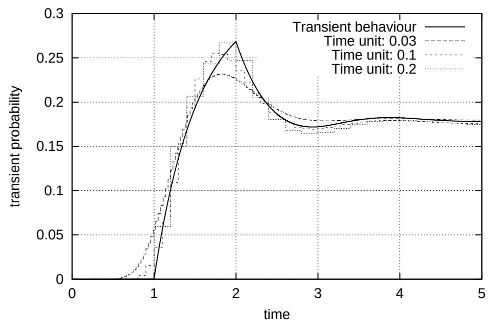


Fig. 27. Approximating transient probabilities

that results in 0 probability for time points smaller than 1. In other words, the second example depicts the advantage given by DPH distributions to model durations with finite support. This example suggests also that DPH approximation can be of importance when preserving reachability properties is crucial (like in modeling time-critical systems) and, hence, DPH approximation can be seen as a bridge between the world of stochastic modeling and the world of functional analysis and model checking [5].

6 Concluding remarks

The main result of this paper has been to show that the DPH and CPH classes of distributions of the same order can be considered a single model set as a function of a scale factor δ. The optimal value of δ, δ_{opt}, determines the best distribution in a fitting experiment. When $\delta_{opt} = 0$ the best choice is a CPH distribution, while when $\delta_{opt} > 0$ the best choice is a DPH distribution. This paper has also shown that the transition from DPH class to CPH class is continuous with respect to several properties, like the distance (denoted by D in 6) between the original and the approximate distributions. The paper presents limit theorems for special cases; however, extensive numerical experiments show that the limiting behavior is far more general than the special cases considered in the theorems.

The numerical examples have also evidenced that for very small values of δ, the diagonal elements of the transition probability matrix are very close to 1, rendering numerically unstable the DPH fitting procedure.

A deep analytical and numerical sensitivity analysis is required to draw more general conclusions for the model level “optimal δ value” and its dependence on the considered performance measure than the ones presented in this work. It is definitely a field of further research.

Finally, we summarize the advantages and the disadvantages of applying approximate DPH models (even with optimal δ value) with respect to using CPH approximations.

Advantages of using DPH:

An obvious advantage of the application of DPH distributions is that one can have a closer approximate of distributions with low coefficient of variation. An other important quantitative property of the DPH class is that it can capture distributions with finite support and deterministic values. This property allows to capture the periodic behavior of a complex stochastic model, while any CPH based approximation of the same model tends to a steady state.

Numerical experiments have also shown that DPH can better approximate distributions with some abrupt or sharp changes in the CDF or in the PDF.

Disadvantages of using DPH:

There is a definite disadvantage of discrete time approximation of continuous time models. In the case of CPH approximation, coincident events do not have to be considered (they have zero probability of occurrence). Instead, when applying DPH approximation coincident events have to be handled, and their consideration may burden significantly the complexity of the analysis.

Acknowledgments

This work has been performed under the Italian-Hungarian R&D program supported by the Italian Ministry of Foreign Affairs and the Hungarian Ministry of Education. A. Bobbio was partially supported by the MURST Under Grant ISIDE; M. Telek was partially supported by Hungarian Scientific Research Fund (OTKA) under Grant No. T-34972.

References

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

[2] S. Asmussen and O. Nerman. Fitting Phase-type distributions via the EM algorithm. In *Proceedings: "Symposium i Advent Statistik"*, pages 335–346, Copenhagen, 1991.

[3] A. Bobbio and A. Cumani. ML estimation of the parameters of a PH distribution in triangular canonical form. In G. Balbo and G. Serazzi, editors, *Computer Performance Evaluation*, pages 33–46. Elsevier Science Publishers, 1992.

[4] A. Bobbio and A. Horváth. Model checking time Petri nets using NuSMV. In *Proceedings 5-th International Workshop on Performability Modeling of Computer and Communication Systems (PMCCS5)*, pages 100–104, 2001.

[5] A. Bobbio and A. Horváth. Petri nets with discrete phase type timing: A bridge between stochastic and functional analysis. *Electronic Notes in Theoretical Computer Science*, 52(3), 2001.

[6] A. Bobbio, A. Horváth, M. Scarpa, and M. Telek. Acyclic discrete phase type distributions: Properties and a parameter estimation algorithm. *Performance Evaluation*, 2003. Accepted for publication.

[7] A. Bobbio, A. Puliafito, and M. Telek. A modeling framework to implement preemption policies in non-Markovian SPN. *IEEE Transactions Software Engineering*, 26:36–54, 2000.

[8] A. Bobbio, A. Puliafito, M. Telek, and K. Trivedi. Recent developments in non-Markovian stochastic Petri nets. *Journal of Systems Circuits and Computers*, 8(1):119–158, Feb 1998.

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

[10] G. Ciardo, R. German, and C. Lindemann. A characterization of the stochastic process underlying a stochastic Petri net. *IEEE Transactions on Software Engineering*, 20:506–515, 1994.

[11] A. Cumani. On the canonical representation of homogeneous Markov processes modelling failure-time distributions. *Microelectronics and Reliability*, 22:583–602, 1982.

[12] A. Cumani. Esp - A package for the evaluation of stochastic Petri nets with phase-type distributed transition times. In *Proceedings International Workshop Timed Petri Nets*, pages 144–151, Torino (Italy), 1985.

[13] R. German. *Performance Analysis of Communication Systems: Modeling with Non-Markovian Stochastic Petri Nets.* John Wiley and Sons, 2000.

[14] M. Gribaudo and A. Horváth. Fluid stochastic petri nets augmented with flush-out arcs: A transient analysis technique. In *Proceedings 9th International Workshop on Petri Nets and Performance Models - PNPM01*, pages 145–154. IEEE Computer Society, 2001.

[15] A. Horváth, A. Puliafito, M. Scarpa, and M. Telek. Analysis and evaluation of non-Markovian stochastic Petri nets. In *Proc. of 11th Performance TOOLS*, volume 1786 of *Lecture Notes in Computer Science*, pages 171–187, Schaumburg, IL, USA, March 2000. Springer.

[16] 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.

[17] 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.

[18] M. Johnson and M. Taaffe. Matching moments to Phase distributions: nonlinear programming approaches. *Stochastic Models*, 6:259–281, 1990.

[19] R. Jones and G. Ciardo. On phased delay stochastic Petri nets: Definition and application. In *Proceedings 9th International Workshop on Petri Nets and Performance Models - PNPM01.* IEEE Computer Society, 2001.

[20] A. Lang and J. L. Arthur. Parameter approximation for phase-type distributions. In *Matrix-analytic methods in stochastic models*, Lecture notes in pure and applied mathematics, pages 151–206. Marcel Dekker, Inc., 1996.

[21] C. Lindemann. DSPNexpress: a software package for the efficient solution of deterministic and stochastic Petri nets. *Performance Evaluation*, 22:3–21, 1995.

[22] M. Neuts. Probability distributions of phase type. In *Liber Amicorum Prof. Emeritus H. Florin*, pages 173–206. University of Louvain, 1975.

[23] M. Neuts. *Matrix Geometric Solutions in Stochastic Models.* Johns Hopkins University Press, Baltimore, 1981.

[24] L. Schmickler. MEDA: mixed Erlang distributions as Phase-type representations of empirical distribution functions. *Stochastic Models*, 8:131–156, 1992.

[25] M. Telek. Minimal coefficient of variation of discrete phase type distributions. In *3rd International Conference on Matrix-Analitic Methods in Stochastic models, MAM3*, pages 391–400, Leuven, Belgium, 2000.