Acyclic Discrete Phase Type Distributions: Properties and a Parameter Estimation Algorithm ⋆

A. Bobbio [a] A. Horváth [b,1] M. Scarpa [c] M. Telek [b,1]

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

[b] *Department of Telecommunications, Technical University of Budapest*

[c] *Istituto di Informatica e Telecomunicazioni, Università di Catania*

Abstract

This paper provides a detailed study on Discrete Phase Type (*DPH)* distributions and its acyclic subclass referred to as Acyclic *DPH* (*ADPH*). Previously not considered similarities and differences between *DPH* and Continuous Phase Type (*CPH)* distributions are investigated and minimal representations, called canonical forms, for the subclass of *ADPH* distributions are provided. We investigate the consequences of the recent result about the minimal coefficient of variation of the *DPH* class [18] and show that below a given order (that is a function of the expected value) the minimal coefficient of variation of the *DPH* class is always less than the minimal coefficient of variation of the *CPH* class. Since all the previously introduced Phase Type fitting methods were designed for fitting over the CPH class we provide a DPH fitting method for the first time. The implementation of the DPH fitting algorithm is found to be simple and stable. The algorithm is tested over a benchmark consisting of 10 different continuous distributions. The error resulted when a continuous distribution sampled in discrete points is fitted by a *DPH* is also considered.

Key words: Continuous and Discrete Phase Type Distributions, Acyclic Discrete Phase Type Distributions, Canonical Forms.

⋆ This work has been performed under a cooperative visiting exchange program supported by the Italian Ministry of Foreign Affairs and the Hungarian Ministry of Education (TÉT).

Email addresses: bobbio@di.unito.it (A. Bobbio), horvath@webspn.hit.bme.hu (A. Horváth), mscarpa@sun195.iit.unict.it (M. Scarpa), telek@hit.bme.hu (M. Telek).

[1] This work was partially supported by Hungarian Scientific Research Fund (OTKA) under Grant No. T-34972.

1 Introduction

Discrete Phase Type *(DPH)* distributions have been introduced and formalized in [10], but they have received little attention in applied stochastic modeling since then, because the main research activity and application oriented work was addressed towards Continuous Phase Type *(CPH)* distributions [11].

However, in recent years a new attention has been devoted to discrete models since it has been observed that they can be utilized in the numerical solution of non-Markovian processes, or they are more closely related to physical observations [15,16]. Moreover, new emphasis has been put on discrete stochastic Petri Nets [5,6,17]. Finally, *DPH*s may have a wide range of applicability in stochastic models in which random times must be combined with constant durations. In fact, one of the most interesting property of the *DPH* distributions is that they can represent in an exact way a number of distributions with finite support, like the deterministic and the (discrete) uniform, and hence one can mix inside the same formalism distributions with finite and infinite support.

In particular, while it is known that the minimal coefficient of variation for the *CPH* family depends only on the order n and is attained by the Erlang distribution [1] ($cv = 1/n$), it is trivial to show, for the *DPH* family, that for any order n the deterministic distribution with $cv = 0$ is a member of the family. Since, the range of applicability of the *PH* distributions may depend on the range of variability of the coefficient of variation given the order n, it is interesting to investigate, for the *DPH* family, how the coefficient of variation depends on the model parameters.

The convenience of using the *DPH* family in applied stochastic modeling has motivated the present paper whose aim is to investigate more closely the properties of the *DPH* family and to provide results that can be profitably exploited for the implementation of an algorithm to estimate the model parameters given an assigned distribution or a set of experimental points [4].

The *DPH* representation of a given distribution function is, in general, non unique [13] and non minimal. Hence, we first explore a subclass of the *DPH* class for which the representation is an acyclic graph (*Acyclic DPH-ADPH*) and we show that, similarly to the continuous case [8], the *ADPH* class admits a unique minimal representation, called canonical form.

We recall the theorem about the minimal coefficient of variation of the *DPH* class as a function of the order and of the mean [18]. This theorem shows that below a given order (that is a function of the mean) the minimal coefficient of variation of the *DPH* class is always less than the minimal coefficient of variation of the *CPH* class. This result, combined with the well known result of [1] (the minimal cv for an n-phase *CPH* distribution is $1/n$ independent of its mean), offers the possibility of comparing the applicability of the *CPH* and *DPH* families to fit distributions with low cv.

An algorithm is presented for the estimation of the *ADPH* model parameters to fit distributions or a set of experimental data. The algorithm is based on the maximum likelihood (ML) principle. A z-transform version of the algorithm is derived from the continuous case [3], while a novel time domain version is provided. It is shown that the time domain algorithm is easier to implement and more stable. The algorithm is then tested on a benchmark of 10 different continuous distributions that have been already utilized for a similar study in the continuous case [4]. However, since a continuous distribution needs to be discretized in order to feed the fitting algorithm, the role of the discretization interval on the performance of the algorithm and on the goodness of the fit is extensively discussed.

The structure of the paper is as follows. Section 2 introduces the basic definitions and notation, and provides a simple example to emphasize some differences between the *CPH* and *DPH* class, differences that are not evident from a comparative analysis reported for instance in [9]. Section 3 derives the canonical form (and their main properties) for the class of *Acyclic-DPH (ADPH)*. Section 4 gives the theorem to describe the minimal coefficient of variation for the *DPH* class as a function of the order and of the mean and shows the shape of the structures that realize minimal coefficient of variation. Section 5 presents the ML estimation algorithm, both in z-transform domain and in time domain. Section 6 discusses the role of the discretization interval on the accuracy of the obtainable approximation, while Section 7 is devoted to present the results of the benchmark analysis. Finally, Section 8 concludes the paper.

2 Definition and Notation

A *DPH* distribution [10,11] is the distribution of the time until absorption in a Discrete-State Discrete-Time Markov Chain *(DTMC)* with n transient states, and one absorbing state. (The case when $n = \infty$ is not considered in this paper.) If the transient states are numbered $1, 2, \ldots, n$ and the absorbing state is numbered $(n+1)$, the one-step transition probability matrix of the corresponding *DTMC* can be partitioned as:

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

where $\mathbf{B} = [b_{ij}]$ is the $(n \times n)$ matrix grouping the transition probabilities among the transient states, $\mathbf{b} = [b_{i,n+1}]^T$ is the $n-$dimensional column vector grouping the probabilities from any state to the absorbing one, and $\mathbf{0} = [0]$ is the zero vector. Since $\widehat{\mathbf{B}}$ is the transition probability matrix of a *DTMC*, the following relation holds: $\sum_{j=1}^{n} b_{ij} = 1 - b_{i,n+1}$.

The initial probability vector of the *DTMC* is an $(n+1)$ dimensional vector $\widehat{\alpha} = [\alpha, \alpha_{n+1}]$, with $\sum_{j=1}^{n} \alpha_j = 1 - \alpha_{n+1}$. In the present paper, we only consider the class of *DPH* distribu-

tions for which $\alpha_{n+1} = 0$, but the extension to the case when $\alpha_{n+1} > 0$ is straightforward.

Let τ be the time till absorption into state $(n+1)$ in the *DTMC*. We say that τ is a *DPH* r.v. of order n and representation $(\alpha, \mathbf{B})$ [11]. Let $f(k)$, $F(k)$ and $\mathcal{F}(z)$ be the probability mass, cumulative probability and probability generating function of τ, respectively. It follows:

$$f(k) = Pr\{\tau = k\} \;=\; \alpha\, \mathbf{B}^{k-1}\, \mathbf{b} \qquad\qquad \text{for } k > 0 \tag{1}$$

$$F(k) = Pr\{\tau \leq k\} \;=\; \alpha \sum_{i=0}^{k-1} \mathbf{B}^i\, \mathbf{b} \;=\; 1 - \alpha \mathbf{B}^k\, \mathbf{e} \tag{2}$$

$$\mathcal{F}(z) = E\{z^\tau\} \;=\; z\,\alpha\,(\mathbf{I} - z\mathbf{B})^{-1}\,\mathbf{b} \;=\; \frac{U(z)}{V(z)} = \frac{z^n + u_{n-1}z^{n-1} + \ldots + u_1 z + u_0}{v_n z^n + v_{n-1}z^{n-1} + \ldots + v_1 z + v_0} \tag{3}$$

where $\mathbf{e}$ is an n-dimensional column vector with all the entries equal to 1 and $\mathbf{I}$ is the $(n \times n)$ identity matrix. A *DPH* distribution is a non-negative, discrete distribution over $\{1, 2, \ldots\}$.

Since the degree of the denominator $V(z)$ in (3) equals the order n, and the degree of the numerator $U(z)$ is at most n. Due to the constraints $u_0 = 0$ (because z is a factor of the numerator) and $\sum_i u_i / \sum_i v_i = 1$ (because $\mathcal{F}(1) = 1$), it turns out that a *DPH* has at most $N_G = 2n - 1$ degrees of freedom. However, its representation has $N_F = n^2 - n + (n-1) = n^2 - 1$ free parameters ($n^2 - n$ in matrix $\mathbf{B}$ and $n - 1$ in vector α). Therefore, the matrix representation is very redundant with respect to the degrees of freedom, and it is reasonable to look for minimal representations.

2.1 Properties of DPHs different from CPHs

A number of properties of the *DPH* family have been derived in [10]. Moreover, the *DPH* family inherits many properties from the *CPH* family [9], for which a more extensive literature exists [2,14,12,7].

However, when *DPH*s and *CPH*s are used to approximate general distributions to transform a non-Markovian process into a *DTMC* or a *CTMC* over an expanded state space, a very crucial feature is to keep the order of the *DPH* or *CPH* distributions as low as possible, since the order affects multiplicatively the size of the expanded state space. Hence, it is very important to establish to what extent the properties of the family are dependent upon the order.

A well known and general result for *CPH* distributions [1] is that the squared coefficient of variation (cv^2) of a *CPH* of order n is not less than $1/n$, and this limit is reached by the Continuous Erlang, CE(λ, n), (or Gamma(λ, n)) distribution of order n (independently of

the parameter λ, and hence independently of the mean of the distribution).

The simple relation, established in [9], to compare the *CPH* and *DPH* classes, preserves the mean but not the coefficient of variation. Hence, in the case of the *DPH* family the consideration about the minimal coefficient of variation requires a more extensive analysis, and it is deferred to Section 4.

However, it is trivial to show that the mentioned property for *CPH*s does not hold for *DPH*'s. In fact, it is clear that for any order n, the *DPH* with representation $(\alpha, \mathbf{J})$ given by:

$$\alpha = [1, 0, \ldots, 0] \qquad \mathbf{J} = \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 \end{bmatrix} \tag{4}$$

represents a deterministic time to absorption $\tau = n$, with $cv = 0$. Hence for any order n there exists at least one *DPH* with $cv = 0$. In order to emphasize other differences, we carry on a simple comparative example on a 2-*CPH* versus a 2-*DPH*.

Example 1 Let τ_C and τ_D be the *CPH* and *DPH* r.v. shown in Figure 1, with representations $(\gamma, \mathbf{\Lambda})$ and $(\alpha, \mathbf{B})$, respectively:

$$\gamma = [1, 0] \,, \quad \mathbf{\Lambda} = \begin{bmatrix} -\lambda_1 & \lambda_1 \\ 0 & -\lambda_2 \end{bmatrix} \,; \qquad \alpha = [1, 0] \,, \quad \mathbf{B} = \begin{bmatrix} 1-\beta_1 & \beta_1 \\ 0 & 1-\beta_2 \end{bmatrix} \,.$$

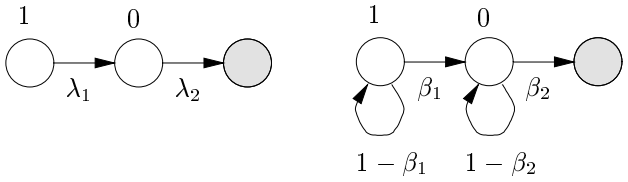


Fig. 1. Two-state *CPH* and *DPH* structures

The mean m, the variance σ^2 and the squared coefficient of variation cv^2 of τ_C and τ_D are given

below.

$$
\begin{aligned}
m_C &= \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \qquad ; \qquad m_D = \frac{1}{\beta_1} + \frac{1}{\beta_2} \\
\sigma_C^2 &= \frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} \qquad ; \qquad \sigma_D^2 = \frac{1}{\beta_1^2} - \frac{1}{\beta_1} + \frac{1}{\beta_2^2} - \frac{1}{\beta_2} \\
cv_C^2 &= \frac{\sigma_C^2}{m_C^2} = \frac{\lambda_1^2 + \lambda_2^2}{(\lambda_1 + \lambda_2)^2} \qquad ; \qquad cv_D^2 = \frac{\sigma_D^2}{m_D^2} = \frac{\beta_1^2 - \beta_1^2\beta_2 + \beta_2^2 - \beta_1\beta_2^2}{(\beta_1 + \beta_2)^2}
\end{aligned}
\tag{5}
$$

Both distributions are characterized by two free parameters: the *CPH* by (λ_1, λ_2), the *DPH* by (β_1, β_2). First, we suppose to fix the value of λ_1 and β_1, and to find the value λ_2^{min} and β_2^{min} that minimize the squared coefficient of variation in (5). The values λ_2^{min} and β_2^{min} are obtained by equating to 0 the derivative of cv^2 with respect to λ_2 and β_2, and are given by:

$$
\lambda_2^{min} = \lambda_1 \quad ; \quad \beta_2^{min} = \frac{\beta_1(2 + \beta_1)}{2 - \beta_1} \ ,
$$

where $0 \leq \beta_2, \beta_1 \leq 1$. The minimal coefficient of variation of the *CPH* structure is obtained when the parameters λ_1 and λ_2 are equal, while the *DPH* structure exhibits the minimal coefficient of variation when, in general, β_1 differs from β_2.

In order to investigate the dependence of the minimal coefficient of variation with respect to the mean, let us assume that the two free parameters of the considered structures are (λ_2, m_C) and (β_2, m_D). Rearranging Equations (5), we have:

$$
\begin{aligned}
\lambda_1 &= \frac{\lambda_2}{m_C\lambda_2 - 1} \qquad ; \qquad \beta_1 = \frac{\beta_2}{m_D\beta_2 - 1} \\
cv_C^2 &= \frac{2 - 2m_C\lambda_2 + m_C^2\lambda_2^2}{m_C^2\lambda_2^2} \qquad ; \qquad cv_D^2 = \frac{2 - 2m_D\beta_2 - m_D\beta_2^2 + m_D^2\beta_2^2}{m_D^2\beta_2^2}
\end{aligned}
\tag{6}
$$

For a given mean (m_C and m_D), the minimal coefficient of variation is obtained by equating to 0 the derivative of cv^2 with respect to λ_2 and β_2, respectively. It is obtained:

$$
\lambda_2^{min} = \frac{2}{m_C} \quad ; \quad \beta_2^{min} = \frac{2}{m_D},
$$

where as a result of the given initial probability vector $[1, 0]$ the mean $m_D \geq 2$ and $\beta_2^{min} \leq 1$. Substituting this value into (6), we obtain:

$$
\lambda_1 = \frac{2}{m_C} \ , \quad cv_C^2 = \frac{1}{2} \quad ; \quad \beta_1 = \frac{2}{m_D} \ , \quad cv_D^2 = \frac{1}{2} - \frac{1}{m_D}.
$$

In the *CPH* case, the minimal coefficient of variation is obtained (as in the previous case) when $\lambda_1 = \lambda_2$ and it is independent of the mean m_C. In the *DPH* case, differently from the previous case, the minimum is attained when $\beta_1 = \beta_2$ (discrete Erlang distribution DE($\frac{2}{m}$, 2)), but the value of the minimum depends on the mean m_D.

3 Acyclic DPHs

Definition 1 *A* DPH *is called Acyclic* DPH *(*ADPH*) if its states can be ordered in such a way that matrix* $\mathbf{B}$ *is an upper triangular matrix.*

By Definition 1, a generic *ADPH* of order n is characterized by $N_F = (n^2 + 3n - 2)/2$ free parameters ($n \cdot (n+1)/2$ in the upper triangular matrix $\mathbf{B}$ and $n-1$ in the initial probability vector α).

Definition 1 implies that a state i can be directly connected to a state j only if $j \geq i$. In an *ADPH*, each state is visited only once before absorption. We define an *absorbing path*, or simply a *path*, the sequence of states visited from an initial state to the absorbing one. In an *ADPH* of order n, the number of paths is finite and is at most $2^n - 1$. The length of a path is the (integer) number of states visited via the path before absorption.

The k-th path, r_k, of length $\ell \leq n$, is characterized by a set of indices, representing the states visited before absorption:

$$r_k = (x_1, x_2, \ldots, x_\ell) \qquad \text{such that} \qquad \begin{cases} 1 \leq x_j \leq n & \forall j : 1 \leq j \leq \ell \\ x_j < x_{j+1} & \forall j : 1 \leq j < \ell \\ b_{x_j, x_{j+1}} > 0 & \forall j : 1 \leq j < \ell \\ b_{x_\ell, n+1} > 0 \end{cases} \tag{7}$$

where the last two conditions mean that in a path any two subsequent indices represent states that are connected by a direct arc (non-zero entry in the $\mathbf{B}$ matrix), and the last index represents a state that is connected by a direct arc to the absorbing state $(n+1)$. Note that the path description, r_k, defines explicitly the initial state of the underlying *DTMC*. The below defined quantities ($P(r_k)$, $\mathcal{F}(z, r_k)$) are all conditional to the initial state (x_1).

Assuming that the underlying *DTMC* starts in state with index x_1, the path r_k in (7), occurs with probability:

$$P(r_k) = \prod_{j=1}^{\ell} \frac{b_{x_j, x_{j+1}}}{1 - b_{x_j, x_j}} \tag{8}$$

and the generating function of the time to arrive to the absorbing state through path r_k is

$$\mathcal{F}(z, r_k) = \prod_{j=1}^{\ell} \frac{(1 - b_{x_j,x_j})z}{1 - b_{x_j,x_j}\, z} \ . \tag{9}$$

$P(r_k)$ is the product of the probabilities of choosing the consecutive states of the path and $\mathcal{F}(z, r_k)$ is the product of the generating functions of the sojourn times spent in the consecutive states of the path.

Let L_i be the set of all the paths starting from state i (i.e., for which $x_1 = i$). The generating function of the time to absorption assuming the *DTMC* starts from state i is:

$$\mathcal{F}_i(z) = \sum_{r_k \in L_i} P(r_k)\ \mathcal{F}(z, r_k)$$

where it is clear from (8) that $\sum_{r_k \in L_i} P(r_k) = 1$.

Corollary 1 *The generating function of an* ADPH *is the mixture of the generating functions of its paths (see also [8]):*

$$\mathcal{F}_{ADPH}(z) = \sum_{i=1}^{n} \alpha_i \sum_{r_k \in L_i} P(r_k)\ \mathcal{F}(z, r_k) \tag{10}$$

Example 2 Let us consider the *ADPH* in Figure 2, with representation:

$$\alpha = [\alpha_1 \ \ \alpha_2]\,, \mathbf{B} = \begin{bmatrix} 0.3 & 0.5 \\ 0 & 0.6 \end{bmatrix} \tag{11}$$

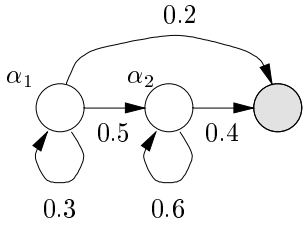


Fig. 2. An example of *ADPH*.

Three different paths can be identified to reach the absorbing state. The paths are depicted in

Figure 3 and have the following structure:

- r_1 is a path of length 1 starting from state 1. Equations (8) and (9) applied to r_1 provide:

$$P(r_1) = \frac{b_{13}}{1 - b_{11}} = \frac{2}{7} \quad ; \quad \mathcal{F}(z, r_1) = \frac{(1 - b_{11})z}{1 - b_{11}z} = \frac{0.7z}{1 - 0.3z}$$

- r_2 is a path of length 2 starting from state 1. Equations (8) and (9) provide:

$$P(r_2) = \frac{b_{12}}{1 - b_{11}} \frac{b_{23}}{1 - b_{22}} = \frac{5}{7} \; ; \; \mathcal{F}(z, r_2) = \frac{(1 - b_{11})z}{1 - b_{11}z} \frac{(1 - b_{22})z}{1 - b_{22}z} = \frac{0.7z}{1 - 0.3z} \frac{0.4z}{1 - 0.6z}$$

- r_3 is a path of length 1 starting from state 2. Equations (8) and (9) provide:

$$P(r_3) = \frac{b_{23}}{1 - b_{22}} = 1 \quad ; \quad \mathcal{F}(z, r_3) = \frac{(1 - b_{22})z}{1 - b_{22}z} = \frac{0.4z}{1 - 0.6z}$$

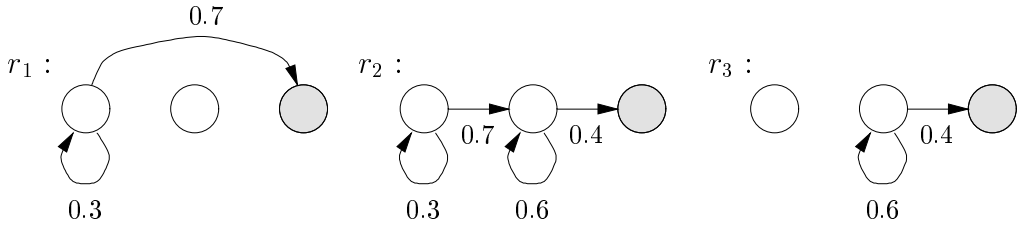


Fig. 3. Possible paths of the *ADPH* of Figure 2.

From (9) and from Corollary 1, it follows that the generating function of the time to absorption does not depend on the particular order of the geometrically distributed sojourn times. Hence, we can reorder the eigenvalues (diagonal elements) of the matrix $\mathbf{B}$ into a decreasing sequence $q_1 \geq q_2 \geq \ldots \geq q_n$. For the sake of convenience, we define the symbols $p_i = (1 - q_i)$ which represent the exit rate from state i. Since the sequence of the q_i's is in a decreasing order, the sequence of the p_i's is in an increasing order: $p_1 \leq p_2 \leq \ldots \leq p_n$.

Any path r_k can be described as a binary vector $\mathbf{u}_k = [u_i]$ of length n defined over the ordered sequence of the q_i's. Each entry of the vector is equal to 1 if the corresponding eigenvalue q_i is present in the path, otherwise the entry is equal to 0. Hence, any path r_k of length ℓ has ℓ ones in the vector $\mathbf{u}_k$. With this representation any path is characterized by a unique binary number $1 \leq \#\mathbf{u}_k \leq 2^n - 1$.

Definition 2 *A path r_k of length ℓ of an* ADPH *is called a* **basic path** *(basic series [8]) if it contains the ℓ fastest phases $q_{n-\ell+1}, \ldots, q_{n-1}, q_n$. The binary vector associated to a basic path is called a basic vector and it contains $(n - \ell)$ initial 0's and ℓ terminal 1's, so that the unique binary number of a basic vector is $\#\mathbf{u}_k = 2^\ell - 1$.*

Theorem 1 *The generating function of a path of an* ADPH *is a mixture of the generating functions of its basic paths.*

Proof: The following lemma gives the basic relationship to prove the theorem.

Lemma 1 *The generating function of a phase with parameter q_i can be represented as the mixture of the generating functions of a phase with parameter q_{i+1} and a sequence of two phases with parameter q_i and q_{i+1}.*

The above lemma is a consequence of the relationship:

$$\frac{(1-q_i)z}{1-q_iz} = w_i \, \frac{(1-q_{i+1})z}{1-q_{i+1}z} + (1-w_i) \, \frac{(1-q_i)z}{1-q_iz} \, \frac{(1-q_{i+1})z}{1-q_{i+1}z} \tag{12}$$

where $w_i = (1-q_i)/(1-q_{i+1})$, hence $0 \leq w_i \leq 1$.

A path r_k is composed by geometric phases according to its associated binary vector $\mathbf{u}_k$. Starting from the rightmost component of $\mathbf{u}_k$ which is not ordered as a basic path (Definition 2), the application of (12) splits the path into two paths which are enriched in components with higher indices. Repeated application of (12) can transform any path in a mixture of basic paths. Cumani [8] has provided an algorithm which performs the transformation of any path into a mixture of basic paths in a finite number of steps. □

Example 3 Let $n = 5$ and let r_k be a path of length $\ell = 2$ characterized by the binary vector $\mathbf{u}_k = [0, 1, 0, 1, 0]$ (corresponding to the phases with parameters q_2 and q_4). By applying Lemma 1 to the rightmost phase of r_k (phase 4), the associated binary vector $\mathbf{u}_k$ can be decomposed in a mixture of two binary vectors one containing phase 5 and the second one containing the sequence of phases $(4, 5)$. Thus the original path is split into the mixture of the following two paths:

$$\mathbf{u}_k = [0, 1, 0, 1, 0] \Longrightarrow \begin{cases} [0, 1, 0, 0, 1] \\ [0, 1, 0, 1, 1] \end{cases}$$

Now for each obtained binary vector, we take the rightmost phase which is not already ordered in a basic path, and we decompose the corresponding path into two paths according to equation (12). The complete decomposition tree is shown next, where all the final binary vectors are basic

vectors according to Definition 2.

$$\mathbf{u}_k = [0,1,0,1,0] \Longrightarrow \left\{ \begin{array}{l} [0,1,0,0,1] \left\{ \begin{array}{l} [0,0,1,0,1] \left\{ \begin{array}{l} [0,0,0,1,1] \\ [0,0,1,1,1] \end{array} \right. \\ [0,1,1,0,1] \left\{ \begin{array}{l} [0,1,0,1,1] \left\{ \begin{array}{l} [0,0,1,1,1] \\ [0,1,1,1,1] \end{array} \right. \\ [0,1,1,1,1] \end{array} \right. \end{array} \right. \\ [0,1,0,1,1] \left\{ \begin{array}{l} [0,0,1,1,1] \\ [0,1,1,1,1] \end{array} \right. \end{array} \right.$$

Corollary 2 *Canonical Form CF1.*
Any ADPH *has a unique representation as a mixture of basic paths called* Canonical Form 1 (CF1). *The* DTMC *associated to the* CF1 *is given in Figure 4, and its matrix representation* $(\mathbf{a}, \mathbf{P})$ *takes the form:*

$$\mathbf{a} = [a_1, a_2, \ldots, a_n] \qquad , \qquad \mathbf{P} = \begin{bmatrix} q_1 & p_1 & 0 & 0 & \ldots & 0 \\ 0 & q_2 & p_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \ldots & q_n \end{bmatrix} \tag{13}$$

with: $\sum_{1}^{n} a_i = 1$ and: $p_1 \leq p_2 \leq \ldots \leq p_n$

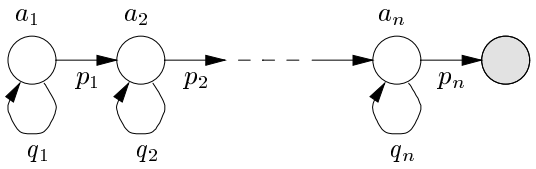


Fig. 4. Canonical Form *CF1*.

Proof - The Corollary is a direct consequence of Theorem 1.

Due to the particular structure of the matrix in (13), the relevant elements can be stored into an n-dimensional vector $\mathbf{p}$ containing the p_i's, so that we will use the notation $(\mathbf{a}, \mathbf{p})$ as the representation of a *CF1*, where $\mathbf{a}$ and $\mathbf{p}$ are n-dimensional vectors (where $0 \leq a_i \leq 1, 0 < p_i \leq 1$).

Example 4 The transformation of the *ADPH* of Figure 2 into the canonical form *CF1* proceeds along the following steps. We first order the eigenvalues of the matrix $\mathbf{B}$ into a decreasing

sequence to obtain: $q_1 = b_{22} = 0.6$ and $q_2 = b_{11} = 0.3$ with $q_1 \geq q_2$. Then, any path is assigned its characteristic binary vector. If the binary vector is not in basic form, each path is transformed into a mixture of basic paths by repeated application of (12), along the line shown in Example 3. Since the *ADPH* of Figure 2 is of order $n = 2$, we have two basic paths $\mathbf{b}_1 = [0, 1]$ and $\mathbf{b}_2 = [1, 1]$.

Path r_1 - The associated binary vector is $\mathbf{u}_1 = [0, 1]$ and is coincident with the basic path $\mathbf{b}_1$. Hence:

$$\mathcal{F}(z, r_1) = \mathcal{F}(z, \mathbf{b}_1)$$

Path r_2 - The associated binary vector is $\mathbf{u}_2 = [1, 1]$ and is coincident with the basic path $\mathbf{b}_2$. Hence:

$$\mathcal{F}(z, r_2) = \mathcal{F}(z, \mathbf{b}_2)$$

Path r_3 - The associated binary vector is $\mathbf{u}_3 = [1, 0]$ and is not a basic path. Hence, r_3 must be transformed in a mixture of basic paths as shown in Figure 5. Application of (12) provides:

$$\mathcal{F}(z, r_3) = w_1 \mathcal{F}(z, \mathbf{b}_1) + (1 - w_1)\mathcal{F}(z, \mathbf{b}_2)$$

with $w_1 = \dfrac{p_1}{p_2} = \dfrac{4}{7}$.

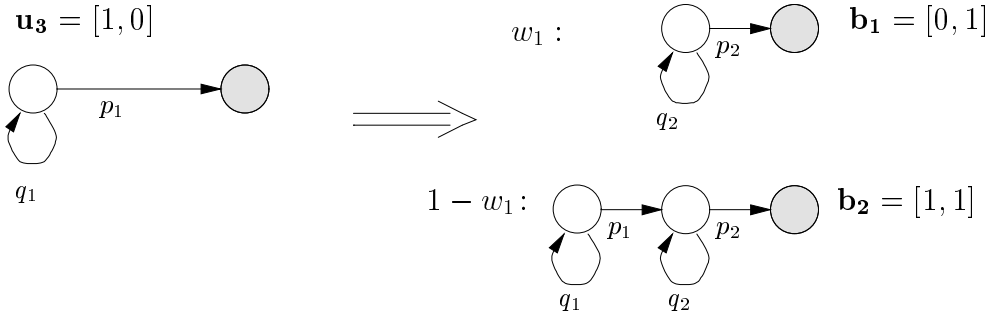


Fig. 5. Transformation of the path r_3.

The generating function of the *ADPH* can be finally written as:

$$\begin{aligned}\mathcal{F}(z) &= \alpha_1 \left[P(r_1)\mathcal{F}(z, r_1) + P(r_2)\mathcal{F}(z, r_2)\right] + \alpha_2 P(r_3)\mathcal{F}(z, r_3) \\ &= \alpha_1 P(r_1)\mathcal{F}(z, \mathbf{b}_1) + \alpha_1 P(r_2)\mathcal{F}(z, \mathbf{b}_2) + \\ &\quad \alpha_2 P(r_3) w_1 \mathcal{F}(z, \mathbf{b}_1) + \alpha_2 P(r_3)(1 - w_1)\mathcal{F}(z, \mathbf{b}_2)\end{aligned} \tag{14}$$

Equation (14) can be rearranged in the following *CF1* form, with $a_1 = \left(\frac{2}{7}\alpha_1 + \frac{4}{7}\alpha_2\right)$, and $a_2 = \left(\frac{5}{7}\alpha_1 + \frac{3}{7}\alpha_2\right)$:

$$\mathcal{F}(z) = a_1 \mathcal{F}(z, \mathbf{b}_1) + a_2 \mathcal{F}(z, \mathbf{b}_2) \tag{15}$$

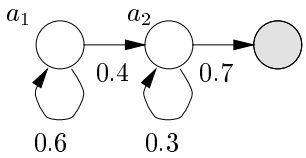


Fig. 6. Canonical form of the ADPH of Figure 2.

The *DTMC* associated to the obtained *CF1* is depicted in Figure 6, and its representation is:

$$\mathbf{a} = [a_1 \;\; a_2] \, , \mathbf{p} = [0.4 \;\; 0.7] \tag{16}$$

3.1 Properties of canonical forms

Property 1 *The* CF1 *is a minimal representation of an* ADPH.

In fact, the number of free parameters of a *CF1-ADPH* of order n is $N_F = 2n - 1$ and is equal to the number of degrees of freedom N_G computed from (3).

Given a canonical form *CF1* of order n and representation $(\mathbf{a}, \mathbf{p})$ (Figure 4), the mean, the second moment and the probability generating function are expressed as:

$$m = \sum_{i=1}^{n} a_i \sum_{j=i}^{n} \frac{1}{p_j} \tag{17}$$

$$d = \sum_{i=1}^{n} a_i \left[\sum_{j=i}^{n} \left(\frac{1}{p_j^2} - \frac{1}{p_j} \right) + \left(\sum_{j=i}^{n} \frac{1}{p_j} \right)^2 \right] \tag{18}$$

$$\mathcal{F}(z) = \sum_{i=1}^{n} a_i \prod_{j=i}^{n} \frac{p_j z}{1 - (1 - p_j) z} \tag{19}$$

Even if the canonical form *CF1* is the simplest minimal form to use in computations, sometimes it can be more convenient to have a minimal representation in which the initial probability is concentrated in the first state. Borrowing terminology from the continuous case [8], we define:

Definition 3 *Canonical Form CF2.*
An ADPH is in canonical form CF2 (Figure 7) if transitions are possible from phase 1 *to all the other phases (including the absorbing one), and, from phase* i $(2 \leq i \leq n)$ *to phase* i *itself and* $i + 1$. *The initial probability is* 1 *for phase* $i = 1$ *and* 0 *for any phase* $i \neq 1$.

The matrix representation of the canonical form *CF2* is:

$$\alpha = [1\ 0 \cdots 0], \mathbf{B} = \begin{bmatrix} q_n & c_1 & c_2 & c_3 & \cdots & c_{n-1} \\ 0 & q_1 & p_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q_{n-1} \end{bmatrix} \tag{20}$$

It is trivial to verify that *CF2* is a minimal form and that the equivalence of distributions of the time to absorption between *CF1* and *CF2* is established by the following relationship:

$$c_k = a_k p_n \tag{21}$$

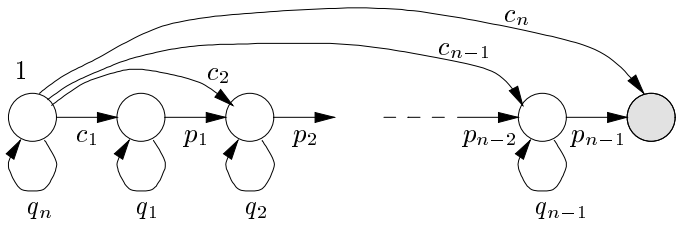


Fig. 7. Canonical Form *CF2*

Definition 4 *Canonical Form CF3.*
An ADPH is in canonical form CF3 (Figure 8) if from any phase i ($1 \leq i \leq n$) transitions are possible to phase i itself, $i+1$ and $n+1$. The initial probability is 1 for phase $i = 1$ and 0 for any phase $i \neq 1$.

The matrix representation of *CF3* is:

$$\alpha = [1\ 0 \cdots 0], \mathbf{B} = \begin{bmatrix} q_n & e'_n & 0 & 0 \cdots & 0 & 0 \\ 0 & q_{n-1} & e'_{n-1} & 0 \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 \cdots & q_2 & e'_2 \\ 0 & 0 & 0 & 0 \cdots & 0 & q_1 \end{bmatrix} \tag{22}$$

It is also easy to verify that *CF3* is a minimal form and that the equivalence of the distribution of the time to absorption between *CF1* and *CF3* is established by

$$s_i = \sum_{j=1}^{i} a_j, \qquad e'_i = \frac{a_i}{s_i} p_i, \qquad e_i = \frac{s_{i-1}}{s_i} p_i.$$

In *CF3* the phases are ordered according to decreasing sojourn time. A path through the last i phases of a *CF1* is represented by a path through the first i phases and a jump to the absorbing state in *CF3*.

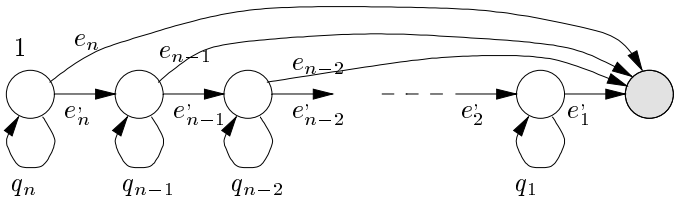


Fig. 8. Canonical Form *CF3*

4 Comparing the minimal coefficient of variation for CPH and DPH

It has been shown in Section 2.1, that a deterministic distribution with $cv = 0$ is a member of the *DPH* as well as the *ADPH* class (4), and moreover that the minimal cv depends on the mean. Since the flexibility in approximating a given distribution function may depend on the range of variability of the coefficient of variation, in this Section we compare the *CPH* and *DPH* families from the point of minimal coefficient of variation. For this purpose we recall the theorem that describes the minimal coefficient of variation for the *DPH* class [18].

To state the theorem, the following notation is introduced. $\tau_n(m)$ is a *DPH* of order n with mean m. Given a real number x, define $I(x) = \lfloor x \rfloor$ the integer part of x and $R(x)$ the fractional part of x, respectively, i.e., $x = I(x) + R(x)$, $0 \leq R(x) < 1$.

Theorem 2 *The minimal squared coefficient of variation, cv^2_{min}, of a* DPH *r.v. $\tau_n(m)$ of order n with mean m is:*

$$cv^2_{min}(\tau_n(m)) = \begin{cases} \dfrac{R(m)(1-R(m))}{m^2} & \text{if } m \leq n \ , \\ \\ \dfrac{1}{n} - \dfrac{1}{m} & \text{if } m > n \ , \end{cases} \tag{23}$$

The DPH *which exhibits this minimal cv^2 is referred to as* MDPH*, and has the following structure (in* CF1 *form):*

- *if $m \leq n$:*
  $p_i = 1, \forall i$ and the nonzero initial probabilities are $a_{n-I(m)} = R(m)$ and $a_{n-I(m)+1} = 1 - R(m)$ (Figure 9);
- *if $m > n$:*
  $p_i = n/m, \forall i$ and the only nonzero initial probability is $a_1 = 1$ (discrete Erlang distribution $DE(n/m, n)$) (Figure 10).

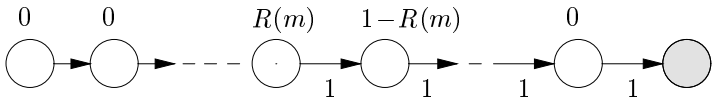


Fig. 9. MDPH with $m \leq n$

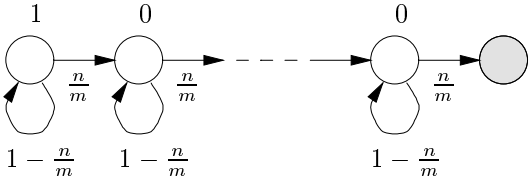


Fig. 10. MDPH with $m > n$

Comment - The *MDPH* structure is uniquely specified given the order n and the mean m. The *MDPH* structure with $m \leq n$ is the mixture of two deterministic *CF1-ADPH*s with length $I(m)+1$ and initial probability $R(m)$ and with length $I(m)$ and initial probability $1-R(m)$. This structure derives from the following identity: if x is real, $x = R(x)\,(I(x)+1) + (1-R(x))\,I(x)$. Hence, for $m \leq n$, the corresponding *MDPH* structure has an effective order $I(m)+1$, being the initial probabilities from state 1 to $n - I(m)$ equal to 0. Hence, in contrast with the continuous case, increasing the order beyond $n > m$ does not have any effect on the minimal cv. The case $m > n$ is more similar to the *CPH* case, and tends to be equal to the *CPH* case as $m \to \infty$.

Theorem 2, combined with the well known result of [1] (the minimal cv^2 for a *CPH* of order n is equal to $1/n$ independent of its mean), offers the possibility of comparing the variability of the *CPH* and *DPH* families.

For fixed m, as the order n increases beyond $n > m$ the minimal cv^2 of the *DPH* remains unchanged, while the minimal cv^2 of the *CPH* decreases as $1/n$. Hence, given m a value $n = n_C$ can be found, such that the minimal cv^2 of the *CPH* of order n_C is less or equal then the minimal cv^2 of the *DPH* of the same order. Recalling Equation (23), the value of n_C is the smallest positive integer which satisfies:

$$\frac{1}{n_C} < \frac{R(m)(1-R(m))}{m^2} \ . \tag{24}$$

It is clear from (24) that if m is integer, $R(m) = 0$ and $n_C \to \infty$. Using the relation $m = I(m) + R(m)$, in Equation (24), we can find the value of $R(m)$ that minimizes (24), for any positive integer $I(m)$, and the corresponding minimal value of n_C.

Setting the derivative of n_C with respect to $R(m)$ to zero with $I(m) = const$, we get:

$$\frac{d\,n_C}{d\,R(m)} = 0 \quad \text{iff} \quad R(m) = \frac{I(m)}{1+2I(m)}$$

From which the minimal value of n_C, corresponding to any mean with integer part equal

$I(m)$	n_{Cmin}
1	8
2	24
3	48
4	80
5	120

Table 1
Values of n_{Cmin} as a function of the integer part of the mean $I(m)$

to $I(m)$, is given by:

$$n_{Cmin} = 4\,I(m)\,(1 + I(m)) \tag{25}$$

From Equation (25) we can get Table 1 which gives us the minimal order n_{Cmin} as a function of the integer part of the mean $I(m)$ for which the *CPH* class provides a minimal cv^2 less than the *DPH* of the same order.

Example 5 Figure 11 shows the minimal cv^2 as a function of the number of phases for the *DPH* family versus the *CPH* family, when the mean is $m = 4.5$. (Note that in Figure 11, $m < n$ when $n \geq 5$.) According to Theorem 2 the minimal cv^2 for the *DPH* class remains unchanged ($cv^2_{min} = 1/81$) for $n \geq 5$, while the the minimal cv^2 for the *CPH* class ($cv^2_{min} = 1/n$) decreases to 0 as $n \to \infty$.

Application of Equation (25) tells us that if $I(m) = 4$ (i.e., the mean is any value $4 \leq m < 5$), the minimal number of phases for which the *CPH* has a cv^2 less than the *DPH* is $n_{Cmin} = 80$, corresponding to a mean $m = 4.444\cdots$.

Let us now consider the dual case, in which we fix the order n. We already know from Table 1 that if $n < n_{Cmin}$ no *CPH* can have a minimal cv^2 less than the *DPH*. However, if m increases with fixed n, we arrive in a situation in which $m > n$, and applying the second part of (23) we see that $cv^2_{min} \to 1/n$ as $m \to \infty$. Hence, as m increases, the behavior of the *DPH* class tends to be similar to the one of the *CPH* class.

Example 6 Figure 12 shows the minimal cv^2 as a function of the mean for a *DPH* of order $n = 5$. Note that for $m \leq n(= 5)$, cv^2_{min} equals zero for any m integer, and cv^2_{min} tends to the value of the *CPH* class $(1/n)$ as $m \to \infty$.

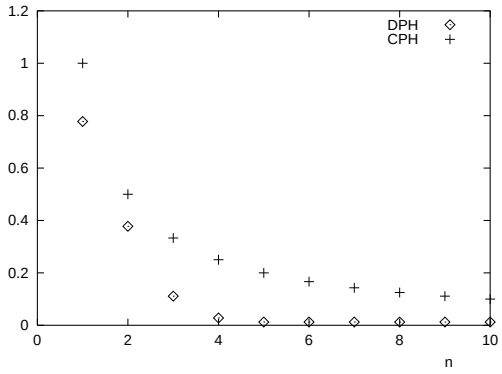


Fig. 11. Minimal squared coefficient of variation for $m = 4.5$

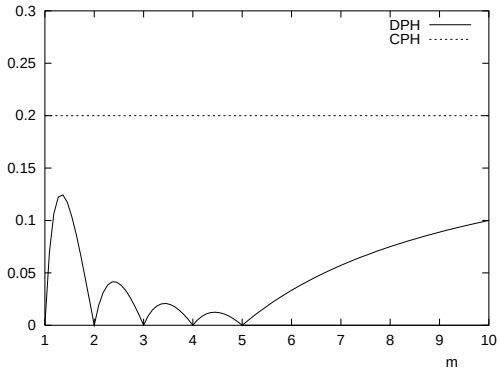


Fig. 12. Minimal squared coefficient of variation for $n = 5$

5 A fitting algorithm for parameter estimation

We describe a fitting algorithm for estimating the parameters of an *ADPH* in *CF1* form, based on the Maximum Likelihood (ML) principle [3,4]. We first derive the closed form expression for the *pmf* both in the z-transform domain and in the time domain, and for its derivatives with respect to the model parameters, then the implemented *ML* estimation algorithm is briefly sketched. The range of applicability of both techniques is finally discussed.

5.1 The probability mass function

The generating function of a *CF1-ADPH* with n phases and representation $(\mathbf{a}, \mathbf{p})$ is provided in (19) and may be written as

$$\mathcal{F}(z) = \sum_{i=1}^{n} a_i \mathcal{F}^{(i)}(z). \tag{26}$$

where $\mathcal{F}^{(i)}$ is the generating function of a path of length $(n-i+1)$ from phase i to $(n+1)$, and is given by:

$$\mathcal{F}^{(i)}(z) = \prod_{k=i}^{n} \frac{1-q_k}{z^{-1}-q_k}. \tag{27}$$

Let σ_i $(\sigma_i \leq n-i+1)$ denote the number of distinct eigenvalues out of the set $\{q_i, q_{i+1}, ..., q_n\}$ and let us further denote by $(\hat{q}_1^{(i)}, \hat{q}_2^{(i)}, ..., \hat{q}_{\sigma_i}^{(i)})$ the σ_i-dimensional vector of the distinct eigenvalues and by $(\hat{m}_1^{(i)}, \hat{m}_2^{(i)}, ..., \hat{m}_{\sigma_i}^{(i)})$ the vector of their multiplicities. With this nota-

tions, $\hat{m}_j^{(i)}$ is the multiplicity of $\hat{q}_j^{(i)}$ and $\sum_{j=1}^{\sigma_i} \hat{m}_j^{(i)} = n - 1 + i$. Equation (27) can be rewritten as:

$$\mathcal{F}^{(i)}(z) = \prod_{j=1}^{\sigma_i} \frac{(1 - \hat{q}_j^{(i)})^{\hat{m}_j^{(i)}}}{(z^{-1} - \hat{q}_j^{(i)})^{\hat{m}_j^{(i)}}},$$

After a partial fraction decomposition, we have:

$$\mathcal{F}^{(i)}(z) = \sum_{j=1}^{\sigma_i} \sum_{l=1}^{\hat{m}_j^{(i)}} \frac{c_{jl}^{(i)}}{(z^{-1} - \hat{q}_j^{(i)})^l} = \sum_{j=1}^{\sigma_i} \sum_{l=1}^{\hat{m}_j^{(i)}} \frac{c_{jl}^{(i)}\ z^l}{(1 - \hat{q}_j^{(i)} z)^l}, \tag{28}$$

where $c_{jl}^{(i)}$ ($j \in \{1, 2, \ldots, \sigma_i\}$ and $l \in \{1, 2, \ldots, \hat{m}_j^{(i)}\}$) are coefficients determined by the partial fraction decomposition. In [3], a recursive algorithm is proposed for the computation of the coefficients $c_{jl}^{(i)}$.

Using the fact that the z-transform of a series like:

$$h(k) = \begin{cases} 0, & \text{if } 0 \le k < l; \\ c\ \dfrac{(k-1)(k-2)...(k-(l-1))}{(l-1)!} q^{k-l}, & \text{if } k \ge l \ge 2; \end{cases}$$

is:

$$\mathcal{H}(z) = \frac{c\ z^l}{(1 - qz)^l},$$

($h(k)$ is the geometric series for $l = 1$), the inverse of (28) is

$$f^{(i)}(k) = \sum_{j=1}^{\sigma_i} \left[c_{j1}^{(i)} (\hat{q}_j^{(i)})^{k-1} + \sum_{l=2}^{\hat{m}_j^{(i)}} c_{jl}^{(i)} \frac{(k-1)(k-2)...(k-(l-1))}{(l-1)!} (\hat{q}_j^{(i)})^{k-l} \right], \quad k \ge 1, \tag{29}$$

which is the conditional *pmf* of absorption in k steps in state $(n+1)$ assuming that the chain started from phase i. Applying (26), the unconditional *pmf* of absorption in k steps becomes:

$$f(k) = \sum_{i=1}^{n} a_i\, f^{(i)}(k) \tag{30}$$

In the time domain, the *pmf* of the time to absorption is obtained from (1):

$$f(k) = \mathbf{a}\ \mathbf{P}^{k-1}\ \mathbf{p}_n\ , \tag{31}$$

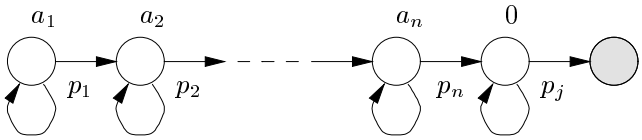


Fig. 13. Structure used to determine the derivative with respect to p_j

where $\mathbf{a}$ and $\mathbf{P}$ are given in (13), and $\mathbf{p}_n$ is a n-dimensional column vector whose first $n-1$ elements are equal to 0, and the n-th element is equal to p_n.

5.2 The derivatives of the probability mass function

In order to solve the non-linear constrained optimization problem that arises from the application of the ML principle (see Section 5.3), the derivatives of the *pmf* with respect to the parameters $(\mathbf{a}, \mathbf{p})$ are needed. Because the *pmf* depends linearly on the entries of $\mathbf{a}$, the derivatives with respect to these parameters are immediate. In order to express the derivative of the *pmf* with respect to p_j, we rearrange (19):

$$\mathcal{F}(z) = \frac{p_j}{z^{-1} - 1 + p_j} \sum_{i=1}^{j} a_i \prod_{k=i,k\neq j}^{n} \frac{p_k}{z^{-1} - q_k} + \sum_{i=j+1}^{n} a_i \prod_{k=i}^{n} \frac{p_k}{z^{-1} - q_k}, \tag{32}$$

where the second term of the r.h.s. does not depend on p_j. The derivative of (32) with respect to p_j is:

$$\frac{\partial \mathcal{F}(z)}{\partial p_j} = \left[\frac{1}{z^{-1} - q_j} - \frac{p_j}{(z^{-1} - q_j)^2} \right] \sum_{i=1}^{j} a_i \prod_{k=i,k\neq j}^{n} \frac{p_k}{z^{-1} - q_k} =$$

$$\frac{1}{p_j} \left[\sum_{i=1}^{j} a_i \mathcal{F}^{(i)}(z) - \frac{p_j}{z^{-1} - q_j} \sum_{i=1}^{j} a_i \mathcal{F}^{(i)}(z) \right], \tag{33}$$

where $\mathcal{F}^{(i)}(z)$ is given in (27). The second term in the r.h.s. may be interpreted as the generating function of a *CF1*-like model that is obtained by adding one further phase, with exit probability p_j, and null initial probability (Figure 13) to the original *CF1* model. Hence, any algorithm that can be used to evaluate the *CF1* structure of Figure 4, can be utilized to evaluate the derivatives with respect to the p factors as in Figure 13.

Using the partial fraction decomposition method, the coefficients of the augmented model of Figure 13, may be calculated by iterating the same recursive algorithm described earlier [3], just one step more.

Using the matrix formulation (31), the time domain equivalent of (33) becomes:

$$\frac{\partial f(k)}{\partial p_j} = \frac{1}{p_j}\left[\hat{\mathbf{a}}\ \mathbf{P}^{k-1}\ \mathbf{p}_n\ -\ \hat{\mathbf{a}}^*\ (\mathbf{P}_j^*)^{k-1}\ \mathbf{p}_{n+1}^*\right] \tag{34}$$

where $\hat{\mathbf{a}}$ ($\hat{\mathbf{a}}^*$) is a row vector of length n $(n+1)$ with elements $\hat{a}_i = \hat{a}_i^* = a_i$ if $1 \le i \le j$, and $\hat{a}_i = \hat{a}_i^* = 0$ otherwise, $\mathbf{p}_{n+1}^*$ is a column vector of length $n+1$ with elements $p_i^* = 0$ if $1 \le i \le n$, $p_{n+1}^* = p_j$, and

$$\mathbf{P}_j^* = \begin{bmatrix} 1-p_1 & p_1 & & \cdots & 0 \\ 0 & 1-p_2 & p_2 & \cdots & 0 \\ & \cdots & & \cdots & 0 \\ 0 & \cdots & \cdots & 1-p_n & p_n \\ 0 & \cdots & \cdots & 0 & 1-p_j \end{bmatrix}$$

is the transition probability matrix that is obtained by adding one more transient phase to the original *CF1* structure as in Figure 13.

Since $\mathbf{P}$ and $\mathbf{P}_j^*$ $(j = 1, ..., n)$ are upper triangular matrices whose only non-zero elements are located in the main diagonal or in the first (upper) subdiagonal the numerical solution of Equations (31) and (34) does not require the complexity of a general vector-matrix multiplication algorithm.

5.3 ML estimation

Let $\phi = \phi_1, ..., \phi_\nu$ be a set of ν integer data samples. Values in ϕ may derive from experimental observations or from the discretization of a continuous *cdf*. Let us denote $\bar{\mathbf{a}}$ and $\bar{\mathbf{p}}$ the maximum likelihood estimators of $\mathbf{a}$ and $\mathbf{p}$, respectively. The likelihood function has the form

$$\mathcal{L}(\phi, \mathbf{a}, \mathbf{p}) = \prod_{i=1}^{\nu} f(\phi_i, \mathbf{a}, \mathbf{p}).$$

The estimation problem consists of finding the parameters $(\bar{\mathbf{a}}, \bar{\mathbf{p}})$ such that the likelihood function $\mathcal{L}(\phi, \bar{\mathbf{a}}, \bar{\mathbf{p}})$ is maximal, under the constraints of the *CF1* form:

- $0 \le p_1 \le p_2 \le \cdots \le p_n \le 1$,
- $a_i \ge 0, \quad \sum_{i=1}^{n} a_i = 1.$

The estimation problem is then formulated in terms of a non-linear constrained optimization problem, that is solved by resorting to an iterative application of a linear programming algorithm. The logarithm of the likelihood function is linearized around the current point by means of a first order series expansion:

$$\begin{aligned} &\log \mathcal{L}(\phi, \mathbf{a} + \Delta, \mathbf{p} + \Delta) = \\ &\quad \log \mathcal{L}(\phi, \mathbf{a}, \mathbf{p}) + \frac{\partial \log \mathcal{L}(\phi, \mathbf{a}, \mathbf{p})}{\partial \mathbf{a}} \; \Delta \; \mathbf{a}^T + \frac{\partial \log \mathcal{L}(\phi, \mathbf{a}, \mathbf{p})}{\partial \mathbf{p}} \; \Delta \; \mathbf{p}^T \end{aligned} \tag{35}$$

Given an initial guess $\mathbf{a}_0, \mathbf{p}_0$, $\mathcal{L}(\phi, \mathbf{a}, \mathbf{p})$ is linearized according to (35) and linear programing is used to find the maximum of the linearized function inside a small box around $\mathbf{a}_0, \mathbf{p}_0$ according to the constraints. The solution of this step is used as the initial guess in the subsequent step of the iterative procedure, and the procedure is iterated until a preassigned tolerance level is reached or the maximum number of iterations is exceeded.

5.4 Comparison of the algorithms

The z-transform algorithm is based on a partial fraction decomposition method applied to Equations (27) and (28) for the computation of the *pmf*, and to Equation (33) for the computation of the derivatives. The most time consuming and unstable part of the algorithm is the evaluation of the coefficients $c_{jl}^{(i)}$ in (28). The instability comes from the fact that when two eigenvalues tend to be equal, the associated coefficients grow unboundedly, and when the eigenvalues are coincident the expression of the partial fraction expansion changes.

During the iterative estimation procedure the p parameters may become closer, and a criterion should be set to decide whether two close eigenvalues are "coincident" and to modify the partial fraction expansion accordingly. Practically, a small quantity ϵ is assigned, and when the difference between two eigenvalues becomes less than ϵ, they are considered to be coincident. However, this procedure introduces numerical instabilities and inaccuracies.

Once the coefficients $c_{jl}^{(i)}$ are determined, the evaluation of the *pmf* and of its derivatives even for a large k can be done recursively at a very low computational cost.

On the other hand, using the time domain analysis, the *pmf* is evaluated through Equation (31) while the derivatives are evaluated through Equation (34). The solution of both equations requires a vector matrix multiplication that must be replicated k times. Due, however, to the very special and sparse structure of the involved matrices and vectors a very specialized algorithm can be used. Moreover, since all the entries in the vectors and matrices are non-negative numbers less than 1, the vector matrix multiplication remains very stable for any value of k. Of course, in this case, the complexity of the algorithm

increases with k. Hence, the time domain algorithm is much simpler to be implemented, more stable and faster. Only when the time span k over which the solution is required becomes very high, the use of the z-transform algorithm may be justified.

We have implemented and experimented both algorithms, but, since the time-domain computation proved to be more applicable for high number of phases, the results we show in the next section are all obtained by means of the time-domain algorithm.

6 Approximating continuous distributions

When using *ADPH* distributions to approximate random variables arising in practical problems, there are cases in which a discrete sample of data points is directly derived from the application. But there are also cases in which the distributions to be approximated are not discrete. For example, *ADPH* distributions can be utilized to approximate continuous distributions.

The *ADPH* approximation of a continuous distribution requires two steps:

(1) The distribution is discretized according to a given discretization step. Indeed, discrete samples and associated mass probability values are generated.
(2) The *ADPH* estimation algorithm is run over the discrete sample provided in the previous step.

The discretization of a continuous distribution is a delicate step that introduces errors, and the amplitude of the introduced errors is mainly related to the size of the discretization interval. Therefore, the role of the discretization interval and its impact on the goodness of fit of *DPH* estimation algorithms is investigated in the following sections.

6.1 The role of the discretization interval

There are several ways to "discretize" a general distribution, i.e., to assign a probability mass to the elements of a discrete, finite (ordered) set $\mathcal{S} = \{x_1, x_2, x_3, \ldots\}$ (where $x_1 < x_2 < x_3 < \ldots$). The most common case of discretization is when the elements of the discrete set are integer multiples of a discretization interval (δ), i.e., $x_i = i\delta$.

Given a r.v. X whose *cdf* is $F_X(x)$, a simple rule for discretizing $F_X(x)$ over the discrete set $\mathcal{S} = \{x_1, x_2, x_3, \ldots\}$ is to use the following:

$$p_i = F_X\left(\frac{x_i + x_{i+1}}{2}\right) - F_X\left(\frac{x_{i-1} + x_i}{2}\right), \quad i > 1, \quad \text{and} \quad p_1 = F_X\left(\frac{x_1 + x_2}{2}\right) \tag{36}$$

where p_i is the probability associated with x_i. This discretization does not preserve the moments of the distribution.

Since there are various potential ways to discretize a given distribution function, here we try to provide some general guidelines.

In general, the smaller is the discretization interval the closer is the discretized distribution to the original one. Hence, on the one hand, the discretization error decreases by decreasing the discretization interval: this remark suggests the use of a small δ. On the other hand, the discretization interval changes the scale of the representation. Indeed, let X be the original (non-negative) random variable expressed in a natural time unit (e.g., seconds); its discretized counterpart X_d is expressed in δ unit. For any reasonable discretization procedure, we must have:

$$E(X^i) \sim \delta^i \; E(X_d^i), \qquad i \geq 1 \; , \tag{37}$$

being $E(X^i)$ and $E(X_d^i)$ the i-th moment of X and X_d^i, respectively.

Equation (37) shows that a discretization procedure modifies the mean of the distribution ($E(X) \sim \delta \; E(X_d)$) (since the mean of the discretized distribution is δ times lower than the mean of the original distribution), but leaves (almost) unchanged its coefficient of variation ($cv(X) \sim cv(X_d)$). Since the minimal cv of a *DPH* distribution is a function of its mean the chosen discretization interval may play a significant role in the variability of the *DPH* and, hence, in the goodness of the fit.

6.1.1 Bounds of the discretization interval

The following considerations provide practical upper and lower bounds to guide in the choice of a suitable discretization interval δ, and are mainly based on the dependence of the minimal coefficient of variation of a *ADPH* on the order n and on the mean m.

Since we only consider *DPH* distributions with no mass at zero, the mean of any *DPH* distribution is greater than 1, which means that, δ should be less than $E(X)$. However, given the number of phases n, in order to completely exploit the flexibility associated with the n phases, a better upper bound is:

$$\delta \leq \frac{E(X)}{n-1} \; . \tag{38}$$

If the squared coefficient of variation of the distribution to be approximated ($cv^2(X_d)$) is greater than $1/n$ (i.e., $cv^2(X) \sim cv^2(X_d) > 1/n$), any small value of δ provides a suitable discretization interval. Instead, if $cv^2(X) \sim cv^2(X_d) \leq 1/n$, in order to allow the *ADPH* to reach this low coefficient of variation (lower than the bound of any *CPH* as established

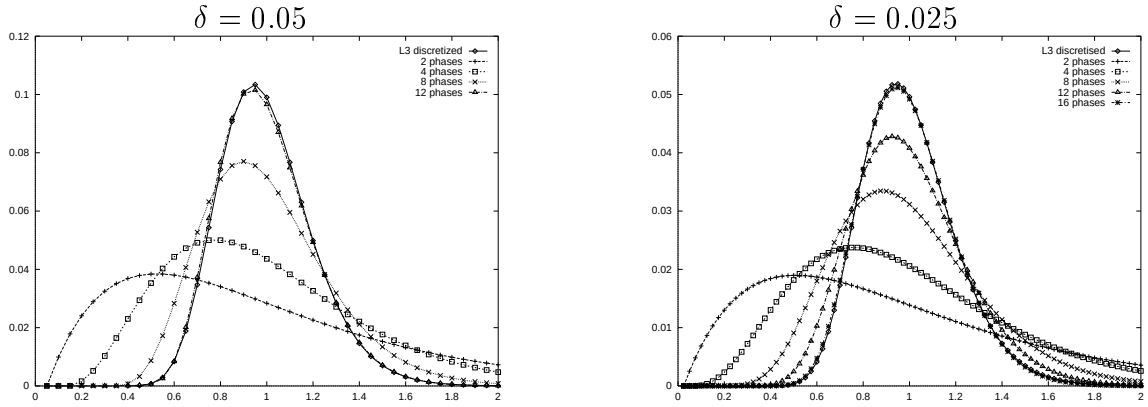


Fig. 14. Effect of changing the discretization step

in [1]), δ should satisfy the following relation

$$\delta > \left(\frac{1}{n} - cv^2(X)\right) E(X) \tag{39}$$

based on the theorem about the minimal cv of $ADPH$ distributions (Section 4).

The effect on the goodness of the attainable approximation of different discretization intervals (δ) is illustrated utilizing a Lognormal distribution with parameters $(1, 0.2)$, whose mean is 1 and cv^2 is 0.0408 (this Lognormal distribution is the test case L3 of the benchmark considered in Section 7).

Figure 14 reports the discretized Lognormal distribution together with the best fit $ADPH$'s of order $n = 2$, $n = 4$, $n = 8$, $n = 12$ and $n = 16$ (only for $\delta = 0.025$) obtained applying the ML algorithm, for two different values of the discretization interval $\delta = 0.05$ and $\delta = 0.025$. The discretized mass of the continuous Lognormal distribution at $k\delta$ is $(F(k\delta) - F((k-1)\delta))/\delta$, where $F(t)$ is the cdf of the Lognormal distribution. Note that the discretized distribution is a function of the discretization interval. The lower and upper bounds of δ, computed from Equations (39) and (38), are reported in Table 2 as a function of the order of the $ADPH$ (the same orders $n = 2, 4, 8, 12$ as in Figure 14 are used).

n	lower bound of δ equation (39)	upper bound of δ equation (38)
4	0.2092	0.333
8	0.0842	0.1429
12	0.0425	0.0909
16	0.0217	0.0667

Table 2
Upper and lower bound for δ as a function of the order

In Figure 14, it can be seen that when δ is less than its lower bound the required low

cv cannot be attained; while when δ is in the proper range (e.g. $n = 12; \delta = 0.05$ and $n = 16; \delta = 0.025$) a reasonably good fit is obtained.

Figure 15 depicts the *cdf* and the *pmf* of three PH distributions approximating distribution L3 using different discretization steps (0.1,0.05,0.025). The figure shows the *cdf* and the *pdf* of the original distribution and the approximating CPH as well. (When plotting the *pmf* the probabilities of the approximating PH distributions are multiplied by $1/\delta$ in order to have the values in the same range. This is done to illustrate in a single figure how the mass functions with different discretization steps follow the shape of the original continuous curve and where the CPH approximation is located compared to them.) All the PH distributions have 8 phases. Having 0.1 as discretization step 8 phases are enough to capture the low cv of the distribution L3 (Table 3), this DPH approximation follows the steep changes of the *pdf* and the *cdf* as well. As the discretization step is decreased the discrete approximation is getting worse and is approaching the continuous approximation.

6.1.2 The required number of phases with a given discretization step

In Figure 14, it is also visible that the lower δ we use (the higher the mean of the discretized distribution with respect to the discretization interval) the more phases are needed in order to achieve the same goodness of fit. In fact, according to the theorem given in Section 4 about the minimal cv of the *ADPH* family, more phases are needed to attain a given coefficient of variation. The minimal number of phases (n) that are needed to reach a given cv^2 when the mean is $E(X_d)$ is given by the next expression

$$n \geq \frac{E(X_d)}{cv^2(X_d)\ E(X_d) + 1}\ , \quad \text{if} \quad cv^2 > \frac{R(E(X_d))(1 - R(E(X_d)))}{E(X_d)^2}\ .$$

Table 3 reports, for the lognormal distribution of Figure 14, the mean $E(X_d)$ and the coefficient of variation $cv^2(X_d)$ of the discretized distribution together with the minimal number of phases needed to reach the coefficient of variation of the original distribution ($cv^2 = 0.0408$), as a function of different discretization steps.

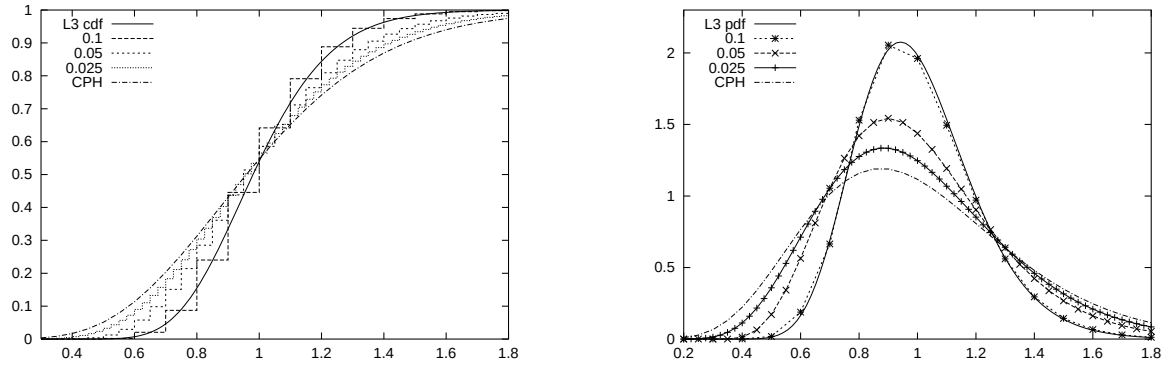


Fig. 15. DPHs with different discretization steps versus CPH

δ	$E(X_d)$	$cv^2(X_d)$	min. phases needed
0.1	9.9649	0.04164	8
0.05	19.9136	0.04101	12
0.025	39.8079	0.04086	16

Table 3
Minimal number of phases as a function of δ

Table 3 also shows how the discretization modifies the mean and the cv as a function of the discretization step.

7 Examples for the estimation process

This section reports the results of the numerical experiments that have been carried out to test the goodness of fit of the proposed ML fitting algorithm. The experiments are based on a benchmark (composed of continuous distributions only) already proposed in [4] to test the goodness of fit of algorithms for *CPH* distributions (the origin and the motivations behind the proposed benchmark are discussed in [4]). Hence, the present results allows us to compare the features of the discrete and the continuous phase type fitting.

Table 4 summarizes the distributions that compose the benchmark. In Table 4, the continuous exponential distribution has been added, which was not present in the original benchmark in [4], since the continuous exponential is not a *DPH* distribution.

Since in our experiments we have to approximate continuous distributions, we have to discretize them before approximation. In the present experiments, we have used the following discretization method. We conventionally assume that the largest sample in the discretized distribution corresponds to the discrete point closest to $\hat{x}$ where $F(\hat{x}) = 0.995$, and we assign a probability mass to all points from 1 to $\hat{x}$ based on the rule in (36). As mentioned in the previous section, this discretization rule does not preserve the moments, so that the moments of the discretized distribution (including the expected value) are not coincident with the ones of the original continuous distribution.

For further reference let us denote $F(\cdot)$ $f(\cdot), F_d(\cdot)$ $f_d(\cdot), \bar{F}(\cdot)$ $\bar{f}(\cdot)$ the *cdf* and *pmf* of the original distribution, the discretized distribution, and the one resulting from the *ML* estimation algorithm, respectively.

According to [4], five different measures have been chosen to evaluate the goodness of the fit. The five measures are defined on Table 5, where $c_1(F)$, $c_2(F)$ and $c_3(F)$ represent the first three centered moments of $F(\cdot)$.

While in [4], measures 4. and 5. were defined over continuous functions (as integrals over the support of the distribution), in Table 5 the discretized version has been reported.

Density	**Symbol**	**Numerical Cases**
Weibull		
$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-(\frac{t}{\eta})^{\beta}}$	*W1*	$\eta = 1 \quad \beta = 1.5$
	W2	$\eta = 1 \quad \beta = 0.5$
Lognormal		
$f(t) = \frac{1}{\sigma\, t\sqrt{2\,\pi}} \exp\left[-\frac{(\log(t/\phi)+\sigma^2/2)^2}{2\,\sigma^2}\right]$	*L1*	$\phi = 1 \quad \sigma = 1.8$
	L2	$\phi = 1 \quad \sigma = 0.8$
	L3	$\phi = 1 \quad \sigma = 0.2$
Uniform on (a, b)	*U1*	$a = 0 \quad b = 1$
	U2	$a = 1 \quad b = 2$
Shifted Exponential		
$f(t) = \frac{1}{2}\, e^{-t} + \frac{1}{2}\, e^{-(t-1)}\ I(t \geq 1)$	*SE*	
Matrix Exponential		
$f(t) = \left(1 + \frac{1}{(2\,\pi)^2}\right)\ (1 - \cos(2\,\pi\, t))\, e^{-t}$	*ME*	
Exponential		
$f(t) = \lambda e^{-\lambda t}$	*EX*	$\lambda = 1$

Table 4
Test cases of the benchmark

Hence, the first three measures in Table 5 are computed between the original and the *ML*-estimation, the last two measures are computed between the discretized and the *ML*-estimation.

7.1 Results

Figure 16 plots the results obtained for the 10 distributions of the benchmark in term of their *pmf*'s. For each distribution of Table 4, Figure 16 reports the discretized distribution $f_d(\cdot)$ (in solid line) and the *ML*-estimations $\bar{f}(\cdot)$, computed for *CF1* with 2, 4 and 8 phases, respectively. The discretization step is assumed $\delta = 0.1$ in all the plots.

A detailed description of the measures obtained for all the distributions in the benchmark is reported from Table 6 to Table 10. In each table, the measures are reported for *CF1* of

1.	Relative error in the 1st moment:	$\hat{e}_1 = \lvert c_1(F) - c_1(\bar{F})\rvert / c_1(F)$
2.	Relative error in the 2nd moment:	$\hat{e}_2 = \lvert c_2(F) - c_2(\bar{F})\rvert / c_2(F)$
3.	Relative error in the 3rd moment:	$\hat{e}_3 = \lvert c_3(F) - c_3(\bar{F})\rvert / c_3(F)$
4.	*pmf* absolute area difference:	$\hat{D} = \sum_{i=1}^{\infty} \lvert f_d(i) - \bar{f}(i)\rvert$
5.	Minus cross entropy:	$-\hat{H} = \sum_{i=1}^{\infty} f_d(i)\ log(\bar{f}(i))$

Table 5
Measures for evaluating the goodness of fit

order 2, 4 and 8 respectively, and for two discretization intervals, namely: $\delta = 0.1$ (as in Figure 16) and $\delta = 0.05$.

For most of the cases the results of the discrete approximation are comparable with the results obtained from the continuous approximation [4]. However, for the cases where the distribution has a low coefficient of variation the *DPH* approximation shows a better fit, which is in line with the result on the minimal cv of the *ADPH* class, discussed in Section 4. As the relation between the order n and the discretization interval δ fits the bounds established in Section 6, the *ADPH* approximation can attain lower coefficients of variations with respect to the *CPH*s of the same order. This can be seen for the test case L3, when $n = 8$ and $\delta = 0.1$.

In the benchmark, there are test cases whose discretized version is a *DPH* distribution. For example, the two uniform distributions (U1 and U2), using a discretization interval $\delta = 0.1$, can be represented as, respectively:

$$\mathbf{a}_1 = [0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1], \quad \mathbf{p}_1 = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1],$$

and

$$\mathbf{a}_2 = [0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0],$$

$$\mathbf{p}_2 = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1],$$

The proposed estimation algorithm is able the find exactly these forms. For example,

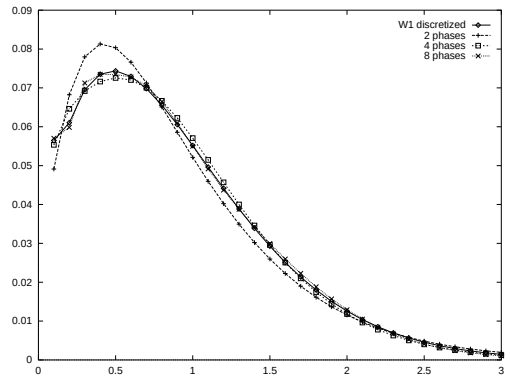
Dist.	$c_1(F)$	Relative error $\hat{e}_1$					
		2 phases		4 phases		8 phases	
		$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.05$
W1	0.9027	0.0144	0.0352	0.0018	0.0142	0.0031	0.0043
W2	2.0000	0.5148	0.6417	0.5203	0.6705	0.5297	0.6711
L1	1.0000	0.2195	0.3703	0.3215	0.4721	0.3224	0.4945
L2	1.0000	0.0918	0.0339	0.0751	0.0794	0.0744	0.0758
L3	1.0000	0.0194	0.0226	0.0107	0.0337	0.0036	0.0072
U1	0.5000	0.0904	0.0667	0.0996	0.0375	0.1000	0.0492
U2	1.5000	0.0308	0.0024	0.0215	0.0387	0.0014	0.0147
SE	1.5000	0.0148	0.0503	0.0238	0.0200	0.0355	0.0623
ME	1.0494	0.0706	0.0831	0.0709	0.1334	0.0709	0.0755
EX	1.0000	0.0181	0.0560	0.0170	0.0467	0.0176	0.0353

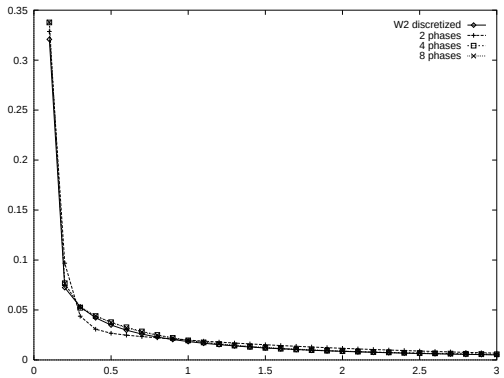
Table 6
Relative error in the 1st moment

Dist.	Original Dist.		Relative error $\hat{e}_2$					
	$c_2(F)$	$(c.v.)^2$	2 phases		4 phases		8 phases	
			$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.05$
W1	0.3756	0.4610	0.1704	0.0402	0.0136	0.0497	0.0288	0.0262
W2	20.000	5.0000	0.9161	0.9600	0.9057	0.9618	0.9128	0.9629
L1	24.534	24.534	0.9282	0.9604	0.9347	0.9714	0.9350	0.9736
L2	0.8964	0.8964	0.4661	0.3977	0.4194	0.4510	0.4286	0.4392
L3	0.0408	0.0408	0.9922	10.405	0.9926	4.2794	0.9928	0.8861
U1	0.0833	0.3333	0.6109	0.7544	0.1611	0.1915	0.0070	0.0522
U2	0.0833	0.0370	11.479	10.308	4.2027	4.3402	0.5621	1.5680
SE	1.2500	0.5555	0.0890	0.1224	0.1732	0.2044	0.1873	0.2765
ME	0.9530	0.8653	0.3267	0.3765	0.1358	0.3561	0.3080	0.3190
EX	1.0000	1.0000	0.0396	0.1953	0.0773	0.2075	0.1246	0.1884

Table 7
Relative error in the 2nd moment

with $\delta = 0.1$, $n = 10$ phases are needed to represent exactly the discretized version of U1; however, it is interesting to observe by a visual inspection of Figure 16 how the approximating *ADPH* improves the fit passing from $n = 2$ to $n = 8$.





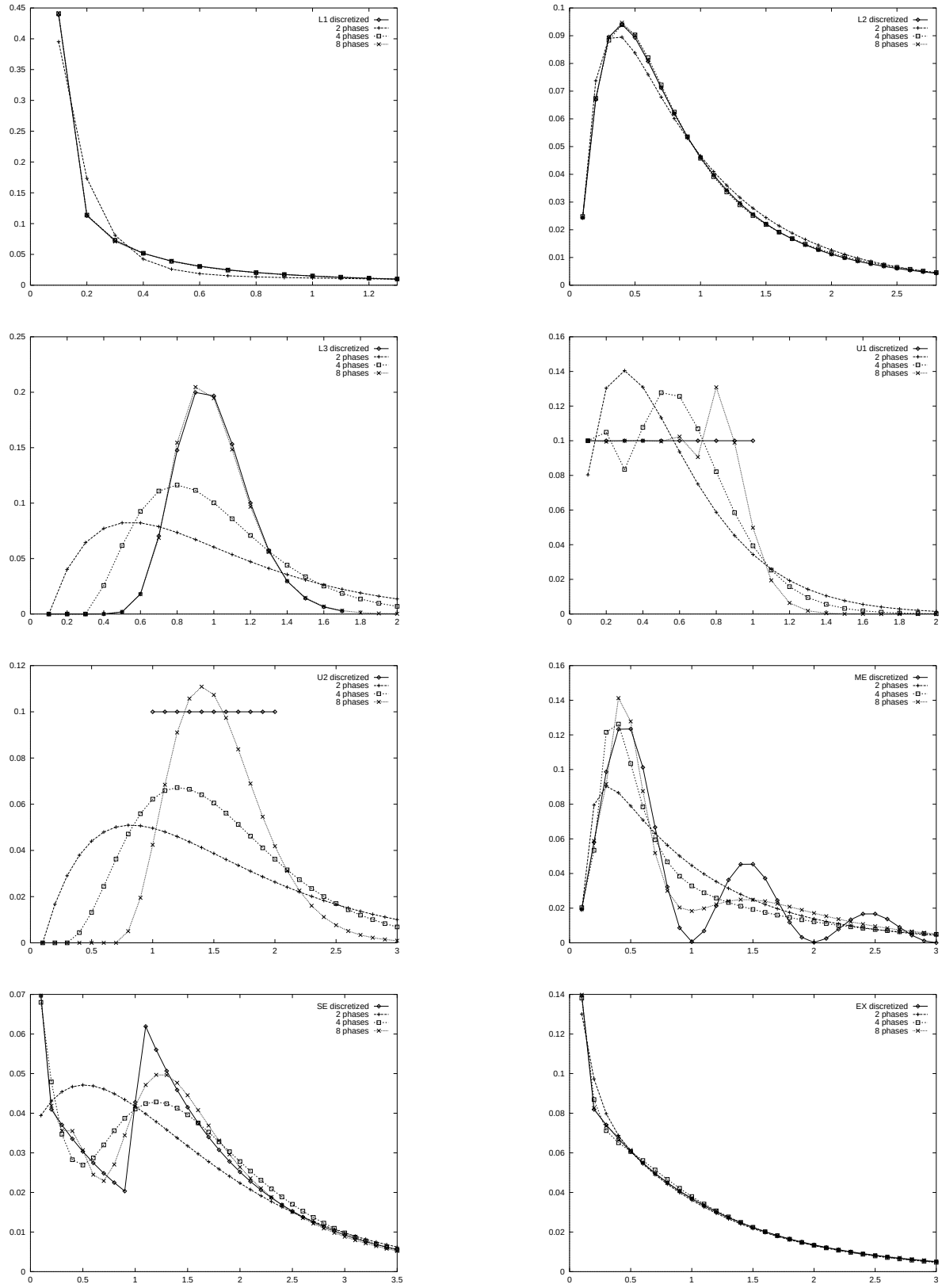


Fig. 16. Probability mass functions of the discretized and the approximating $ADPH$ distributions

Dist.	$c_3(F)$	Relative error $\hat{e}_3$					
		2 phases		4 phases		8 phases	
		$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.05$
W1	0.2468	0.7011	0.3739	0.0263	0.1875	0.0899	0.0723
W2	592.00	0.9917	0.9977	0.9889	0.9976	0.9902	0.9977
L1	16573	0.9995	0.9998	0.9995	0.9999	0.9995	0.9999
L2	3.1315	0.8061	0.7974	0.7833	0.8149	0.8053	0.8152
L3	0.0050	74.82	83.019	11.362	18.847	0.1713	2.1075
U1	0.0000	0.3949	0.4090	0.2826	0.2987	0.2482	0.2625
U2	0.0000	8.5742	7.4103	4.8749	5.3990	3.3664	3.8605
SE	2.0000	0.1740	0.4032	0.4917	0.6468	0.4752	0.1260
ME	1.9929	0.4958	0.6240	0.1261	0.5418	0.5922	0.6278
EX	2.0000	0.0490	0.4366	0.1249	0.4579	0.3114	0.4689

Table 8
Relative error in the 3rd moment

Dist.	H	Cross Entropy $\hat{H}$					
		2 phases		4 phases		8 phases	
		$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.05$
W1	0.7869	0.7740	0.7868	0.7688	0.7810	0.7686	0.7803
W2	1.1546	0.8707	0.5096	0.8478	0.5061	0.8577	0.5138
L1	0.3745	0.1620	0.0866	0.1315	0.0420	0.1314	0.0457
L2	0.8756	0.8940	0.8284	0.8910	0.8256	0.9034	0.8257
L3	-0.2104	0.5300	0.5845	0.0982	0.2141	-0.1978	-0.1252
U1	0.0000	0.1528	0.1822	0.0722	0.1016	0.0172	0.0562
U2	0.0000	1.0041	1.0071	0.5801	0.6525	0.2201	0.3280
SE	1.2950	1.3327	1.2820	1.2983	1.2382	1.2883	1.2287
ME	0.7277	0.9439	0.9049	0.9013	0.8712	0.8609	0.8201
EX	1.0000	0.9455	0.9317	0.9442	0.9312	0.9460	0.9286

Table 9
Cross entropy

7.2 Empirical guidelines of ADPH fitting

In this subsection, we try to draw a general conclusion about the applicability of ADPH distribution fitting based on our fitting experience.

The following specific properties of the class of ADPH distributions limit their applicability in distribution fitting:

- bounded moments;
- limited number of waves of the pmf;
- sharp changes of the pmf is not possible at high ($>> n$) time instances (where n is the order of the ADPH distribution);
- exponentially decaying tail distribution.

Dist.	Area Difference					
	2 phases		4 phases		8 phases	
	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.05$
W1	0.0710	0.0948	0.0248	0.0208	0.0088	0.0103
W2	0.1879	0.2095	0.1445	0.1976	0.1512	0.1972
L1	0.2259	0.2782	0.0099	0.0191	0.0063	0.0054
L2	0.0613	0.0951	0.0396	0.0358	0.0371	0.0306
L3	1.0233	1.0733	0.6274	0.7574	0.0248	0.3099
U1	0.4270	0.4162	0.2773	0.2886	0.1227	0.1892
U2	1.2994	1.2107	0.7871	0.8975	0.3091	0.4832
SE	0.2788	0.2770	0.1592	0.1824	0.0913	0.1260
ME	0.4698	0.5006	0.3845	0.4184	0.2668	0.3151
EX	0.0458	0.0275	0.0284	0.0341	0.0147	0.0254

Table 10
pmf absolute area difference

The bounds of the first two moments of ADPH distributions are already mentioned above ($m \geq 1$, and Theorem 2). These bounds already indicate that the ADPH class can not exhibit all the possible sets of moments that can be obtained by positive distributions. Similar bounds hold for higher moments as well. This means that the moments of the distribution to be fitted should be realizable by ADPH distributions of the given order.

The number of waves exhibited by the pmf of an ADPH distribution of order n is not greater than $n/2$.

The sharpest possible change of the pmf of a ADPH of order n at time $k, (k > n)$ is obtained by the discrete Erlang(n) distribution. Hence distributions with sharp changes at time $k >> n$ is not possible to approximate closely. E.g., the jumps of the discrete uniform distribution between a and b, with $a < n$ and $b >> n$, can not be equally well captured. It is possible to capture the sharp jump at time a, but the best ADPH fitting of the jump at b is distributed in a wide range.

With respect to the tail behavior, it is possible to approximate both heavier or lighter tail behavior than exponential decay to some upper limit, but after a limit all ADPH distributions have an exponentially decaying tail behavior.

As a conclusion, based on our experiences and the above considerations, we would predict a "close" ADPH fit when the distribution to be fitted has moments achievable with ADPH of order n, has less than $n/2$ waves, has a smoothly changing pmf especially after time n, and has an approximately exponentially decaying tail behavior.

8 Conclusion

Some previously not considered properties of the *DPH* distributions, which are essential for *DPH* fitting, are investigated and compared with the known properties of the *CPH* distributions. Similarly to the continuous family, acyclic *DPH* distributions admit a minimal representation called canonical form. Resorting to the canonical form, we have investigated the dependence of the minimal squared coefficient of variation on the mean and on the order, and we have established the conditions for which the minimal coefficient of variation for the *DPH* family is less than the one for the *CPH* family of the same order.

The results about the wider variability of the *DPH* class can be very relevant in stochastic modeling. When *PH* distributions are used in modeling, the number of states in the model depends multiplicatively on the number of phases. Keeping the order as low as possible increases the capability of the approach.

Furthermore, since the deterministic distribution is a member of the *ADPH* class, the use of *DPH* distributions offers a viable technique to handle random execution times and constant durations inside the same formalism.

A *DPH* fitting method is presented for the first time. Similar to the continuous case we used a maximum likelihood estimation procedure for the evaluation of the parameters of a *ADPH* distribution in canonical form *CF1*. While previous estimation algorithms for the *CPH* family were based on a transform domain analysis, we have shown that the time domain analysis is also possible, and the estimation algorithm based on time domain expressions is easier to implement,numerically simpler and more stable.

The goodness of fit of this new algorithm has been tested with respect to a benchmark composed of 10 different continuous distributions. However, in order to apply the proposed procedure to a continuous distribution, the continuous function must be discretized according to a given discretization interval. The role of the discretization interval has been discussed, and the way to chose a suitable discretization interval as a function of the mean and of the coefficient of variation has been indicated.

As it could have been expected from the properties of the *ADPH* family, the fitting algorithm performs better than the *CPH* one in the cases in which the coefficient of variation is low, and in the cases of distributions with finite support (like the uniform).

References

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

[2] D. Assaf and B. Levikson. Closure of phase type distributions under operations arising in reliability theory. *The Annals of Probability*, 10:265–269, 1982.

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

[5] G. Ciardo. Discrete-time markovian stochastic Petri nets. In *Proceedings of the 2-nd International Workshop on Numerical Solution of Markov Chains*, pages 339–358, 1995.

[6] G. Ciardo and R. Zijal. Discrete deterministic and stochastic Petri nets. In *ICASE Technical Report No. 96-72*, pages 1–25. NASA, 1996.

[7] C. Commault and J.P. Chemla. An invariant of representations of phase-type distributions and some applications. *Journal Applied Probability*, 33:368–381, 1996.

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

[9] R.S. Maier. The algebraic construction of phase-type distributions. *Stochastic Models*, 7:573–602, 1991.

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

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

[12] M.F. Neuts. Two further closure properties of PH-Distributions. *Asia-Pacific Journal of Operational Research*, 9:459–477, 1992.

[13] C.A. O'Cinneide. On non-uniqueness of representations of Phase-type distributions. *Stochastic Models*, 5:247–259, 1989.

[14] C.A. O'Cinneide. Characterization of Phase-type distributions. *Stochastic Models*, 6(1):1–57, 1990.

[15] W.J. Padgett and J.D. Spurrier. On discrete failure models. *IEEE Transactions on Reliability*, 34:253–256, 1985.

[16] A.A. Salvia. Some results on discrete mean residual life. *IEEE Transactions on Reliability*, 45:359–361, 1996.

[17] M. Scarpa and A. Bobbio. Kronecker representation of stochastic Petri nets with discrete PH distributions. In *International Computer Performance and Dependability Symposium - IPDS98*, pages 52–61. IEEE CS Press, 1998.

[18] M. Telek. The minimal coefficient of variation of discrete phase type distributions. In *Proceedings of The Third Iinternational Conference on Matrix-Analytic Methods in Stochastic Models*, July 2000.