Acyclic Discrete Phase Type Distributions: 
Part 1: Properties and Canonical Forms*

Andrea Bobbio¹, András Horváth², Marco Scarpa³, Miklós Telek²†

¹ Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale  
² Department of Telecommunications, Technical University of Budapest  
³ Istituto di Informatica e Telecomunicazioni, Università di Catania  
bobbio@di.unito.it, {horvath,telek}@pyxis.hit.bme.hu, mscarpa@sun195.ilt.unict.it

Abstract

The main goal of this paper is to derive properties about Discrete Phase Type (DPH) distributions and to provide results which make convenient their use in the approximate analysis of discrete state non-Markovian stochastic systems. To this end, similarities and differences between DPH and Continuous Phase Type (CPH) distributions are investigated. An immediate difference is that DPH distributions can represent in an exact way a number of distributions with finite support, like the deterministic. A minimal representation, called canonical form, for the subclass of Acyclic Discrete Phase Type distributions (ADPH) is provided. It is then shown that below a given order (that is a function of the expected value) the minimal coefficient of variation of the ADPH family is always less than the minimal coefficient of variation of the CPH family.

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

1 Introduction

Discrete Phase Type (DPH) distributions have been introduced and formalized in [9], 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 [10].

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 [14, 15]. Moreover, new emphasis has been put on discrete stochastic Petri Nets [4, 5, 16]. Finally, DPHs

---

*This work has been performed under a cooperative visiting exchange program supported by the Italian Ministry of Foreign Affairs and the Hungarian Committee for Technological Development (OMFB).
†M. Telek was partially supported by Hungarian Scientific Research Fund (OTKA) under Grant No. F-23971.
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 minimum coefficient of variation for the CPH family depends only on the order $n$ and is attained by the Erlang distribution [1], 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 [3].

The DPH representation of a given distribution function is, in general, non unique [12] 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 [7], the ADPH class admits a minimal representation, called canonical form.

Utilizing the canonical form, we first derive a theorem to evaluate the minimal coefficient of variation as a function of the order and of the mean, and we show that below a given order (that is a function of the mean) the minimum coefficient of variation of the ADPH family is always less than the minimum coefficient of variation of the CPH family.

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 [8]. Section 3 derives the canonical form (and their main properties) for the class of Acyclic-DPH (ADPH). Section 4 investigates the question of the minimal coefficient of variation for the ADPH class as a function of the order and of the mean and derives the shape of the structures with minimal coefficient of variation. Finally, Section 5 concludes the first part of the paper.

## 2 Definition and Notation

A DPH distribution [9, 10] 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:

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

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

The initial probability vector is an \((n + 1)\) dimensional vector \(\alpha = [\alpha, \alpha_{n+1}]\), with \(\sum_{j=1}^{n} \alpha_j = 1 - \alpha_{n+1} = 1 - \bar{\alpha}_{n+1}.\) In the present paper, we only consider the class of DPH distributions for which \(\alpha_{n+1} = 0\), but the extension to the case when \(\alpha_{n+1} > 0\) is straightforward.

Let \(\tau\) be the time till absorption into state \((n + 1)\) in the DTMC. We say that \(\tau\) is a DPH r.v. of order \(n\) and representation \((\alpha, B)\) [10]. Let \(f(k), F(k)\) and \(\mathcal{F}(z)\) be the probability mass, cumulative probability and probability generating function of \(\tau\), respectively. It follows:

\[
f(k) = Pr\{\tau = k\} = \alpha \ B^{k-1} b \quad \text{for } k > 0
\]

\[
F(k) = Pr\{\tau \leq k\} = \alpha \sum_{i=1}^{k-1} B^i b = 1 - aB^k e
\]

\[
\mathcal{F}(z) = E\{z^\tau\} = z \alpha (I - zB)^{-1} b = \frac{N(z)}{D(z)}
\]

where \(e\) is an \(n\)-dimensional column vector with all the entries equal to 1 and \(I\) 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 \(D(z)\) in (3) is equal to the order \(n\), and the degree of the numerator \(N(z)\) is at most \((n - 1)\), it turns out that a DPH has at most \(N_G = n(n - 1)\) degrees of freedom. However, its representation has \(N_F = n^2(n - 1)\) free parameters \((n^2\) in matrix \(B\) and \(n - 1\) in vector \(\alpha\)). Therefore, the matrix representation is very redundant with respect to the degrees of freedom, and it is reasonable to look for minimal representations.

Let \(\phi_i\) be the sojourn time the DPH (the DTMC) spends in phase \(i\). \(\phi_i\) is geometrically distributed, i.e., \(Pr\{\text{sojourn time in } i = k\} = b_{ii}^{k-1}(1 - b_{ii})\). The generating function of the sojourn time in \(i\) is:

\[
\mathcal{F}_i(z) = E\{z^{\phi_i}\} = \frac{(1 - b_{ii}) z}{1 - b_{ii} z}
\]

Note that \(E\{\phi_i\} = 1/(1 - b_{ii})\), \(E\{\phi_i^2\} = (1 + b_{ii})/(1 - b_{ii})^2\), \(E\{\phi_i^2\} - E\{\phi_i\}^2 = b_{ii}/(1 - b_{ii})^2\) and \(cv_{\phi_i}^2 = b_{ii}\). We use the terminology that phase \(i\) is faster than phase \(j\) if \(b_{ii} < b_{jj}\), i.e., the mean time spent in phase \(i\) is less than the mean time spent in phase \(j\).

### 2.1 Properties of DPHs different from CPHs

A number of properties of the DPH family of distributions have been derived in [9]. Moreover, the DPH family inherits many properties from the CPH family [8], for which a more extensive literature exists [2, 13, 11, 6].

However, when DPHs and CPHs 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($\lambda, n$), (or Gamma($\lambda, n$)) distribution of order $n$ (independently of the parameter $\lambda$, and hence independently of the mean of the distribution).

The simple relation, established in [8], 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 prove that the mentioned property for CPHs does not hold for DPHs. In fact, it is clear that for any order $n$, the DPH with representation $\alpha$ given by:

$$\alpha = [1, 0, \ldots, 0]$$

represents a deterministic time to absorption $\tau = n$, with $cv = 0$. Hence for any order $n$ there exists at least one DPH with $cv = 0$. Moreover, in a DPH, the minimal $cv$ depends on the parameter values while in a CPH the minimal $cv$ is attained only by the Erlang distribution. In order to emphasise these differences, we carry on a simple comparative example on a 2-CPH versus a 2-DPH.

**Example 1** - Let $\tau_C$ and $\tau_D$ be the CPH and DPH r.v. shown in Figure 1, with representations $(\gamma, \Lambda)$ and $(\alpha, B)$, respectively:

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

![Figure 1: Two-state CPH and DPH structures](image)

The mean $m$, the variance $\sigma^2$ and the squared coefficient of variation $cv^2$ of $\tau_C$ and $\tau_D$ are given below.

$$m_C = \frac{1}{\lambda_1} + \frac{1}{\lambda_2}, \quad m_D = \frac{1}{\beta_1} + \frac{1}{\beta_2}$$

$$\sigma^2_C = \frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2}, \quad \sigma^2_D = \frac{1}{\beta_1^2} + \frac{1}{\beta_2^2} + \frac{1}{\beta_1 \beta_2}$$

$$cv^2_C = \frac{\sigma^2_C}{m_C^2} = \frac{\lambda_1^2 + \lambda_2^2}{(\lambda_1 + \lambda_2)^2}, \quad cv^2_D = \frac{\sigma^2_D}{m_D^2} = \frac{\beta_1^2 - \beta_1 \beta_2 + \beta_2^2 - \beta_1 \beta_2}{(\beta_1 + \beta_2)^2}$$
Both distributions are characterized by two free parameters: the CPH by \((\lambda_1, \lambda_2)\), the DPH by \((\beta_1, \beta_2)\). First, we suppose to fix the value of \(\lambda_1\) and \(\beta_1\), and to find the value \(\lambda_2^{\text{min}}\) and \(\beta_2^{\text{min}}\) that minimize the squared coefficient of variation in (5). The values \(\lambda_2^{\text{min}}\) and \(\beta_2^{\text{min}}\) are obtained by equating to 0 the derivative of \(cv^2\) with respect to \(\lambda_2\) and \(\beta_2\), and are given by:

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

where \(0 \leq \beta_2, \beta_1 \leq 1\). The minimal coefficient of variation of the CPH structure is obtained when the parameters \(\lambda_1\) and \(\lambda_2\) are equal, while the DPH structure exhibits the minimal coefficient of variation when, in general, \(\beta_1\) differs from \(\beta_2\).

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 \((\lambda_2, m_C)\) and \((\beta_2, m_D)\). Rearranging Equations (5), we have:

\[
\lambda_1 = \frac{\lambda_2}{m_C \lambda_2 - 1} ; \quad \beta_1 = \frac{\beta_2}{m_D \beta_2 - 1} ; \quad cv^2 = \frac{2 - 2m_C \lambda_2 + m_C^2 \lambda_2^2}{m_C^2 \lambda_2^2} \quad \text{and} \quad cv^2_D = \frac{2 - 2m_D \beta_2 - m_D \beta_2^2 + m_D^2 \beta_2^3}{m_D^2 \beta_2^3}.
\]

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 \(\lambda_2\) and \(\beta_2\), respectively. It is obtained:

\[
\lambda_2^{\text{min}} = \frac{2}{m_C} ; \quad \beta_2^{\text{min}} = \frac{2}{m_D},
\]

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

\[
\lambda_1 = \frac{2}{m_C}, \quad cv^2 = \frac{1}{2} ; \quad \beta_1 = \frac{2}{m_D}, \quad cv^2_D = \frac{1}{2} - \frac{1}{m_D}.
\]

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(\(\frac{2}{m_D}, 2\)), 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 \(B\) is an upper triangular matrix.

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

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 \(2^n - 1\).
A path $r_k$ of length $\ell \leq n$, is characterized by a set of indices, representing the states visited before absorption:

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

where the last two conditions mean that in a path any two subsequent indices represent states that must be connected by a direct arc (non-zero entry in the $B$ matrix), and the last index represents a state that must be connected by a direct arc to the absorbing state $(n + 1)$.

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

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

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

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

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 [7]):

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

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

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

Three different paths can be identified to reach the absorbing state. The paths are depicted in Figure 3 and have the following description:

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

$$P(r_1) = \frac{b_{13}}{1 - b_{11}} = \frac{2}{7}; \quad \mathcal{F}(z, r_1) = \frac{(1 - b_{11})z}{1 - b_{11}z} = \frac{0.7z}{1 - 0.3z}$$
Figure 2: An example of ADPH.

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

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

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

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

Figure 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: $p_1 \leq p_2 \leq \ldots \leq p_n$. 

7
Any path $r_k$ can be described as a binary vector $u_k = [u_i]$ 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 $\ell$ has $\ell$ ones in the vector $u_k$. With this representation any path is characterized by a unique binary number $1 \leq \#u_k \leq 2^n - 1$.

**Definition 2** A path $r_k$ of length $\ell$ of an ADPH is called a basic path (basic series [7]) if it contains the $\ell$ 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 $\ell$ terminal 1’s, so that the unique binary number of a basic vector is $\#u_k = 2^\ell - 1$.

**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:

$$\frac{p_i z}{1 - q_i z} = w_i \frac{p_{i+1} z}{1 - q_{i+1} z} + (1 - w_i) \frac{p_i z}{1 - q_i z} \frac{p_{i+1} z}{1 - q_{i+1} z}$$

(12)

where $w_i = p_i/p_{i+1}$, hence $0 \leq w_i \leq 1$.

A path $r_k$ is composed by geometric phases according to its associated binary vector $u_k$. Starting from the rightmost component of $u_k$ 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 [7] 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 $u_k = [0, 1, 0, 1, 0]$ (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 $u_k$ 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:

$$u_k = [0, 1, 0, 1, 0] \implies \begin{cases} [0, 1, 0, 0, 1] \\ [0, 1, 0, 1, 1] \end{cases}$$

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.
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 \((a, P)\) takes the form:

\[
\begin{align*}
\mathbf{u}_k &= [0, 1, 0, 1, 0] \\
&\implies \left\{ \\
&\quad [0, 1, 0, 0, 1] \\
&\quad [0, 1, 1, 0, 1] \\
&\quad [0, 1, 0, 1, 1] \\
&\quad [0, 1, 1, 1, 1] \\
&\right\} \\
&\implies \left\{ \\
&\quad [0, 0, 0, 1, 1] \\
&\quad [0, 0, 1, 1, 1] \\
&\quad [0, 1, 0, 1, 1] \\
&\quad [0, 1, 1, 1, 1] \\
&\right\}
\end{align*}
\]

\[a = [a_1, a_2, \ldots, a_n], \quad P = \begin{bmatrix} q_1 & p_1 & 0 & 0 & \cdots & 0 \\ 0 & q_2 & p_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q_n \end{bmatrix} \quad (13)\]

with: \(\sum_{i=1}^{n} a_i = 1\) and: \(p_1 \leq p_2 \leq \cdots \leq p_n\)

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{CF1_diagram}
\caption{Canonical Form CF1.}
\end{figure}

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 \(p\) containing the \(p_i\)'s, so that we will use the notation \((a, p)\) as the representation of a CF1, where \(a\) and \(p\) are \(n\)-dimensional vectors (where \(0 \leq a_i \leq 1, 0 < p_i \leq 1\)).

Example 4 - The transformation of the ADPH introduced in Example 3 (Figure 2) into the canonical form CF1 proceeds along the following steps. We first order the eigenvalues of the matrix \(B\) into a decreasing sequence to obtain: \(q_1 = b_{22} = 0.6\) and \(q_2 = b_{11} = 0.3\) with \(q_1 \leq q_2\). 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 \(b_1 = [0, 1]\) and \(b_2 = [1, 1]\).
Path $r_1$ - The associated binary vector is $u_1 = [0, 1]$ and is coincident with the basic path $b_1$. Hence:

$$\mathcal{F}(z, r_1) = \mathcal{F}(z, b_1)$$

Path $r_2$ - The associated binary vector is $u_2 = [1, 1]$ and is coincident with the basic path $b_2$. Hence:

$$\mathcal{F}(z, r_2) = \mathcal{F}(z, b_2)$$

Path $r_3$ - The associated binary vector is $u_3 = [1, 0]$ 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:

$$\mathcal{F}(z, r_3) = w_1 \mathcal{F}(z, b_1) + (1 - w_1) \mathcal{F}(z, b_2)$$

with $w_1 = \frac{p_1}{p_2} = \frac{4}{i}$.

![Diagram of transformation of path $r_3$.](image)

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

$$\mathcal{F}(z) = \alpha_1 [P(r_1)\mathcal{F}(z, r_1) + P(r_2)\mathcal{F}(z, r_2)] + \alpha_2 P(r_3)\mathcal{F}(z, r_3)$$

$$= \alpha_1 P(r_1)\mathcal{F}(z, b_1) + \alpha_1 P(r_2)\mathcal{F}(z, b_2) +$$

$$\alpha_2 P(r_3)w_1\mathcal{F}(z, b_1) + \alpha_2 P(r_3)(1 - w_1)\mathcal{F}(z, b_2)$$

(14)

Equation (15) can be rearranged in the following $CF1$ form, with $a_1 = \left(\frac{2}{i} \alpha_1 + \frac{4}{i} \alpha_2\right)$, and $a_2 = \left(\frac{2}{i} \alpha_1 + \frac{2}{i} \alpha_2\right)$:

$$\mathcal{F}(z) = a_1 \mathcal{F}(z, b_1) + a_2 \mathcal{F}(z, b_2)$$

(15)

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

$$a = \begin{bmatrix} a_1 & a_2 \end{bmatrix}, \quad p = \begin{bmatrix} 0.4 & 0.7 \end{bmatrix}$$

(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 $N_F = n(n - 1)$ 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 $(a, p)$ (Figure 4), the mean, the second moment and the probability generating function are expressed as:

$$m = \sum_{i=1}^{n} a_i \sum_{j=1}^{n} \frac{1}{p_j}$$  \hspace{1cm} (17)

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

$$\mathcal{F}(z) = \sum_{i=1}^{n} a_i \prod_{j=i}^{n} \frac{p_j z}{1 - (1 - p_j) z}$$  \hspace{1cm} (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 [7], 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:

$$\alpha = [1 \ 0 \cdots 0], \quad B = \begin{bmatrix} q_n & c_1 & c_2 & c_3 & \cdots & c_{n-1} \\ 0 & q_1 & p_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q_{n-1} \end{bmatrix}$$  \hspace{1cm} (20)

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

$$c_k = a_k p_n$$  \hspace{1cm} (21)

![Figure 6: Canonical form of the ADPH of Figure 2.](image)
**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:

\[
\alpha = [1 \ 0 \ldots 0] , \quad B = \begin{bmatrix}
q_n & e'_{n} & 0 & 0 & \cdots & 0 & 0 \\
0 & q_{n-1} & e'_{n-1} & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & q_2 & e'_2 \\
0 & 0 & 0 & 0 & \cdots & 0 & q_1
\end{bmatrix}
\] (22)

It is trivial to verify that CF3 is a minimal form and that the relationships between CF1 and CF3 is established by the following relationships:

\[
s_i = \sum_{j=1}^{i} a_j \\
e'_i = \frac{a_i}{s_i} p_i \\
e_i = \frac{s_{i-1}}{s_i} p_i
\]

**3.2 Conditional absorption time**

If an ADPH r.v. \( \tau \) is expressed in CF1, its mean, second moment and generating function can be evaluated according to a recursive algorithm. To do this, we need the following notations and definition.
**Definition 5** Given an ADPH r.v. \( \tau \) of order \( n \) with CF1 representation \((a, p)\), the conditional absorption time \( \tau_i \) is defined as the time to reach state \( i + 1 (\leq n) \) conditioned to the fact that the DTMC started from one of the preceding phases (from phase 1 to \( i \)).

The conditional absorption time \( \tau_i \) is obtained by making state \( i + 1 \) absorbing and renormalizing the initial probabilities. The conditional absorption time \( \tau_i \) is a CF1 of order \( i \) and representation \((a^{(i)}/s_i, p^{(i)})\), where:

- \( a^{(i)} \) is the vector grouping the first \( i \) elements of \( a \), i.e. \( a^{(i)} = \{a_1, ..., a_i\} \);
- \( p^{(i)} \) is the vector grouping the first \( i \) elements of \( p \), i.e. \( p^{(i)} = \{p_1, ..., p_i\} \);
- \( s_i \) is a normalization factor and is equal to the sum of the first \( i \) elements of the initial probability vector, i.e. \( s_i = \sum_{j=1}^{i} a_j, s_n = 1 \).

Let us denote by \( m_i, d_i \) and \( F_i \) the mean, the second moment and the generating function of the conditional absorbing time \( \tau_i \), respectively. The following recursive relations hold:

\[
\begin{align*}
m_{i+1} &= \frac{s_i}{s_{i+1}} m_i + \frac{1}{p_{i+1}}, \\
d_{i+1} &= \frac{s_i}{s_{i+1}} \left( d_i + \frac{2m_i}{p_{i+1}} \right) + \frac{2 - p_{i+1}}{(p_{i+1})^2}, \\
F_{i+1}(z) &= \frac{s_i}{s_{i+1}} F_i(z) \frac{p_{i+1} z}{1 - (1 - p_{i+1}) z} + \frac{a_{i+1}}{s_{i+1}} \frac{p_{i+1} z}{1 - (1 - p_{i+1}) z}.
\end{align*}
\]

**4 Minimal coefficient of variation for ADPHs**

It has been shown in Section 2.1, that the deterministic distribution with \( cv = 0 \) is a member of the ADPH class, and moreover that the \( cv \) 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 explore the question of the minimal coefficient of variation for the class of ADPH, given the order \( n \) and the mean \( m \). This is done in the following Theorem 2.

To state the theorem, the following notation is introduced. \( \tau_n(m) \) is an ADPH 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, \( cv_{min}^2 \), of an ADPH r.v. \( \tau_n(m) \) of order \( n \) with mean \( m \) is:

\[
\begin{align*}
\text{cv}_{min}^2(\tau_n(m)) &= \begin{cases} 
\frac{R(m)(1 - R(m))}{m^2} & \text{if } m \leq n, \\
\frac{1}{n} - \frac{1}{m} & \text{if } m > n.
\end{cases}
\end{align*}
\]

The ADPH which exhibits this minimal \( cv^2 \) is referred to as MDPH, and has the following structure (in CF1 form):
• if \( m \leq n \):
  \[ p_i = 1, \forall i \text{ and the nonzero initial probabilities are } a_{n-I(m)} = R(m) \text{ and } a_{n-I(m)+1} = 1 - R(m) \]  (Figure 9);

• if \( m > n \):
  \[ p_i = n/m, \forall i \text{ and the only nonzero initial probability is } a_1 = 1 \]  (discrete Erlang distribution \( DE(n/m, n) \)) (Figure 10).

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 \( CF1-ADPHs \) 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 \( m \geq n \) is more similar to the CPH case, and tends to be equal to the CPH case as \( m \to \infty \).

Proof - The first part of the proof consists of showing that the MDPH structures have the coefficients of variation defined in (26). But this part derives immediately by applying Equations (17) and (18) to the MDPH structures of Figures 9 and 10.

The second part of the proof consists of showing that the coefficients of variation defined in (26) are minimal for any \( m \) and \( n \). The line of the proof is based on transforming a generic \( CF1-ADPH \) into an MDPH phase by phase starting from phase 1 to phase \( n \) preserving the mean, and showing that at each step of the transformation the coefficient of variation cannot increase. In order to prove this, we need the following intermediate results:

- If two \( CF1-ADPH \) distributions have the same mean \( m \) and are “identical” from phase \( k \) to phase \( n \) (the last \( n-k \) phases), then the two coefficient of variations are in the same order (\( <, =, > \)) as the second moments measured on the first \( k \) phases only (this property is formally stated and proved in Lemma 2).
• Given a CF1-ADPH of order $n$ and mean $m$, such that the $n - 1$ phases (from 1 to $n - 1$) are in MDPH form, its coefficient of variation cannot be less than the coefficient of variation of the MDPH of same order $n$ and same mean $m$ (this property is formally stated and proved in Lemma 3).

• The $cv$ of any CF1-ADPH distribution of order $n$ and mean $m$ is not less than the $cv$ of MDPH($n, m$) (Lemma 4).

**Lemma 2** Let $\tau_n(m)$ and $\theta_n(m)$ be two CF1 r.v.’s of order $n$, with equal mean $m^* = m$ and representation $(a, p)$ and $(\alpha, \pi)$, respectively. Let us further assume that there exists a number $k$ ($k < n$) for which $a_i = \alpha_i$, $p_i = \pi_i$ for any $i > k$ ($i = k + 1, \ldots, n$). Let $\tau_k$ and $\theta_k$ be the conditional absorption time in phase $k+1$ in the two ADPH’s, respectively, and let $m^*_k$, $m^*_k$, $d^*_k$, $a^*_k$ and $cv^*_k$, $cv^*_k$ be their mean values, second moments, and coefficients of variation, respectively. By construction, $m^*_k = m^*_k$. Then the coefficients of variation of the original r.v.’s $\tau_n(m)$ and $\theta_n(m)$ are in the same relation $(<, =, >)$ as the coefficients of variation of the first $k$ phases.

**Proof of Lemma 2:** According to Definition 5, $\tau_k$ has representation $(a^{(k)}/s^*_k, p^{(k)})$ and $\theta_k$ has representation $(\alpha^{(k)}/s^*_k, p^{(k)})$ (with $s^*_k = s^*_k = s_k$ by assumption). We can assume, without loss of generality that:

$$cv(\tau_k) \leq cv(\theta_k) \quad (27)$$

Since the mean values are equal, Equation (27) also holds for the second moments, and from Equation (18) the relation $d^*_k \leq d^*_k$ can be written as:

$$\sum_{i=1}^{k} \frac{\alpha^{(k)}_i}{s_k} \left( \sum_{j=1}^{i} \left( \frac{1}{(p^*_j)^2} - \frac{1}{p^*_j} \right) + \left( \sum_{j=1}^{i} \frac{1}{p^*_j} \right)^2 \right) \leq \sum_{i=1}^{k} \frac{\alpha^{(k)}_i}{s_k} \left( \sum_{j=1}^{i} \left( \frac{1}{(\pi^*_j)^2} - \frac{1}{\pi^*_j} \right) + \left( \sum_{j=1}^{i} \frac{1}{\pi^*_j} \right)^2 \right),$$

Multiplying both sides by $s_k$, and adding two equal terms (from phase $k + 1$ to $n$), we can write:

$$\sum_{i=k+1}^{n} a_i \left[ \sum_{j=1}^{n} \left( \frac{1}{p^*_j} \right)^2 - \frac{1}{p^*_j} \right] + \left( \sum_{j=1}^{n} \frac{1}{p^*_j} \right)^2 \right] \leq \sum_{i=k+1}^{n} \frac{\alpha_i}{s_k} \left[ \sum_{j=1}^{n} \left( \frac{1}{\pi^*_j} \right)^2 - \frac{1}{\pi^*_j} \right] + \left( \sum_{j=1}^{n} \frac{1}{\pi^*_j} \right)^2$$

The l.h.s. and the r.h.s. are the expressions for the second moments of $\tau_n(m)$ and $\theta_n(m)$, respectively. Indeed:

$$\sum_{i=1}^{n} a_i \left[ \sum_{j=1}^{n} \left( \frac{1}{p^*_j} \right)^2 - \frac{1}{p^*_j} \right] \leq \sum_{i=1}^{n} \frac{\alpha_i}{s_k} \left[ \sum_{j=1}^{n} \left( \frac{1}{\pi^*_j} \right)^2 - \frac{1}{\pi^*_j} \right] + \left( \sum_{j=1}^{n} \frac{1}{\pi^*_j} \right)^2$$

which leads to:

$$cv(\tau_n(m)) \leq cv(\theta_n(m)).$$

This concludes the proof of Lemma 2.

To formulate Lemma 3 we define $\tau^M_n(m_n)$ as the MDPH of order $n$ and mean $m_n$ (which is unique).
Lemma 3 Let us construct a r.v. $\tau_n(m_n)$ of order $n$ and mean $m_n$ by adding to $\tau^{M}_{n-1}(m_{n-1})$ a further phase (phase $n$) with initial probability $a_n$ and exit probability $p_n$, so that $\tau^{M}_{n-1}(m_{n-1})$ results to be the conditional absorption time (Definition 5) of $\tau_n(m_n)$ in state $n$. Then:

$$cv(\tau_n(m_n)) \geq cv(\tau^{M}_n(m_n)).$$  

Comment: Lemma 3 states that any ADPH of order $n$ and mean $m_n$, obtained by adding a new phase of any value to the MDPH of order $n-1$ (and renormalizing the initial probabilities), has a coefficient of variation which is not less than the coefficient of variation of the MDPH of equal order ($n$) and mean ($m_n$).

The proof of Lemma 3 requires a detailed evaluation of several cases, and is deferred to the Appendix.

Lemma 4 Let $\tau_n(m_n)$ be a CF1-ADPH r.v. of order $n$ with mean $m_n$ then

$$cv(\tau_n(m_n)) \geq cv(\tau^{M}_n(m_n)).$$

Proof of Lemma 4 - The proof is based on iterating $n$ times the following algorithm:

**Step 1** - Let us consider the conditional absorption time of $\tau_n(m_n)$ (Definition 5) in state $i + 1 = 2$, say $\tau_1(m_1)$. By the definition of $\tau^M_1(m_1)$

$$\tau_1(m_1) = \tau^{M}_1(m_1)$$  

(Note that for any $\tau_n(m_n)$, $\tau_1(m_1)$ is such that $m_1 \geq 1$).

**Step 2** - Let us now consider the conditional absorption time of $\tau_n(m_n)$ (Definition 5) in state $i + 1 = 3$, say $\tau_2(m_2)$. We know from Lemma 3 that $cv(\tau_2(m_2)) \geq cv(\tau^{M}_2(m_2))$.

We replace $\tau_2(m_2)$ with $\tau^{M}_2(m_2)$ in $\tau_n(m_n)$ and we obtain a new random variable $\tau^{(2)}_n(m_n)$ of order $n$ and mean $m_n$ that, by Lemma 2, has a coefficient of variation:

$$cv(\tau_n(m_n)) \geq cv(\tau^{(2)}_n(m_n))$$

**Step i** - We proceed iteratively with the same scheme for phases $i = 3, 4, \ldots, n$. The generic $i$-th step of the transformation is depicted in Figure (11). Figure (11a) shows the transition graph of the random variable $\tau^{(i-1)}_n(m_n)$ that is obtained after $i - 1$ steps of the transformation and in which the conditional absorption time in phase $i$ is an MDPH of order $i - 1$. Now we consider the $i$-th phase. By Lemma 3 the conditional absorption time of $\tau^{(i-1)}_n(m_n)$ in phase $i + 1$ has a coefficient of variation which is not less than the MDPH of order $i$ with the same mean. We replace this MDPH of order $i$ in the original structure to obtain the random variable $\tau^{(i)}_n(m_n)$ for which Lemma 2 states that:

$$cv(\tau_n(m_n)) \geq cv(\tau^{(i)}_n(m_n))$$

**Step n** - At step $n$, the original $\tau_n(m_n)$ is transformed into the MDPH with the same mean and the proof of Lemma 4 is concluded.
Figure 11: One step of the transformation

The proof of Theorem 2 is given by combining Lemma 4, with Theorem 1. Lemma 4 proves the assertion for the CF1-ADPH class, while Theorem 1 ensures that the obtained result is valid over the whole ADPH class. □

Comment: We conjecture that Theorem 2 holds for the whole DPH class but our proof covers only the ADPH restriction.

Example 5 - Figure 12 shows the steps of the transformation that, according to Lemma 4, leads a generic CF1-ADPH of order \( n = 3 \) into a MDPH of equal order and equal mean.

Starting from Figure 12-1, that shows the original 3-CF1-ADPH, we follow step by step the procedure presented in Lemma 4.

Step 1. - No action is required.

Step 2. - The first two phases are isolated and the initial probabilities re-normalized in order to compute the conditional absorption time in phase 2: \( \tau_2(m_2) \) (Figure 12-2).

- The MDPH of order 2 with the same mean \( m_2 \) is evaluated: \( \tau_2^N(m_2) \) (Figure 12-3). Note that since \( m_2 < 2 \), the first MDPH structure of (26) is used (\( m \leq n \)).

- The obtained MDPH is plugged into the original ADPH by substituting the first two phases, to get the new r.v. \( \tau_2^{(2)}(m_n) \) (Figure 12-4).

Step 3. - The r.v. \( \tau_2^{(2)}(m_n) \) obtained at the end of Step 2 is a 3-ADPH that is then transformed into the final 3-MDPH (Figure 12-5). Note that, in this case, \( m_3 > n \) and the second structure in (26) is used.

The mean of the structures of Figure 12-1, 12-4. and 12-5. is 3.17, and the \( cv^2 \) of the same structures are, respectively:

\[
\begin{align*}
  cv_1^2 &= 0.285, \\
  cv_4^2 &= 0.098, \\
  cv_5^2 &= 0.018.
\end{align*}
\]
Figure 12: Transformation of a $CF1-ADPH$ into the $MDPH$ with the same mean

4.1 Comparing the minimal $cv^2$ for CPH and DPH

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 $ADPH$ families of the same order.

For fixed $m$, as the order $n$ increases beyond $n > m$ the minimal $cv^2$ of the $ADPH$ 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 $ADPH$ of the same order. Recalling Equation (26), the value of $n_C$ is the smallest positive integer which satisfies:

$$\frac{1}{n_C} < \frac{R(m)(1 - R(m))}{m^2}. \quad (30)$$

It is clear from (30) that if $m$ is integer, $R(m) = 0$ and $n_C \to \infty$. Using the relation $m = I(m) + R(m)$, in Equation (30), we can find the value of $R(m)$ that minimizes (30), 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 $I(m) = \text{const}$, we get:

$$\frac{dn_C}{dR(m)} = 0 \quad \text{iff} \quad R(m) = \frac{I(m)}{1 + 2I(m)}.$$

From which the minimal value of $n_C$, corresponding to any mean with integer part equal to $I(m)$, is given by:

$$n_{C_{\text{min}}} = 4I(m)(1 + I(m)) \quad (31)$$
From Equation (31) we can get Table 1 which gives us the minimal order \( n_{\text{Cmin}} \) as a function of the integer part of the mean \( I(m) \) for which the \( \text{CDH} \) class provides a minimal \( cv^2 \) less than the \( \text{ADPH} \) of the same order.

Since for an \( \text{ADPH} \) the lowest significant value of \( I(m) \) is \( I(m) = 1 \), we obtain from Table 1 that for an order \( n < 8 \) no member of the \( \text{CPH} \) family exists whose minimal \( cv^2 \) is less than the one of the \( \text{ADPH} \).

**Example 6** - Figure 13 shows the minimal \( cv^2 \) as a function of the number of phases for the \( \text{ADPH} \) family versus the \( \text{CPH} \) family, when the mean is \( m = 4.5 \). (Note that in Figure 13, \( m < n \) when \( n \geq 5 \).) According to Theorem 2 the minimal \( cv^2 \) for the \( \text{DPH} \) class remains unchanged (\( cv^2_{\text{min}} = 1/81 \)) for \( n \geq 5 \), while the the minimal \( cv^2 \) for the \( \text{CPH} \) class (\( cv^2_{\text{min}} = 1/n \)) decreases to 0 as \( n \to \infty \).

Application of Equation (31) 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 \( \text{CPH} \) has a \( cv^2 \) less than the \( \text{ADPH} \) is \( n_{\text{Cmin}} = 80 \), 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 \( n < n_{\text{Cmin}} \) no \( \text{CPH} \) can have a minimal \( cv^2 \) less than the \( \text{ADPH} \). However, if \( m \) increases with fixed \( n \), we arrive in a situation in which \( m > n \), and applying the second part of (26) we see that \( cv^2_{\text{min}} \to 1/n \) as \( m \to \infty \). Hence, as \( m \) increases, the behavior of the \( \text{ADPH} \) class tends to be similar to the one of the \( \text{CPH} \) class.

**Example 7** - Figure 14 shows the minimal \( cv^2 \) as a function of the mean for a \( \text{CF1-ADPH} \) of order \( n = 5 \). Note that for \( m \leq n (= 5) \), \( cv^2_{\text{min}} \) equals zero for any \( m \) integer, and \( cv^2_{\text{min}} \) tends to the value of the \( \text{CPH} \) class \((1/n)\) as \( m \to \infty \).

## 5 Conclusion

The properties of the \( \text{DPH} \) distributions are investigated and compared with the known properties of the \( \text{CPH} \) distributions. Similarly to the continuous family, acyclic \( \text{DPH} \) 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 \( \text{ADPH} \) family is less than the one for the \( \text{CPH} \) family of the same order. Hence, the famous theorem of Aldous and Shepp about the minimal coefficient of variation for the \( \text{CPH} \) class does not hold for the \( \text{DPH} \) class.
The results about the wider variability of the \textit{ADPH} class can be very relevant in stochastic modeling. When \textit{PH} distributions are used in modeling, then 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 \textit{ADPH} class, the use of \textit{DPH} distributions offers a viable technique to handle random execution times and constant durations inside the same formalism. This last feature has originated new research efforts aimed at combining the flexibility of \textit{ADPH} with the modeling power of other formalisms, like Petri nets.

Finally, the possibility of representing any \textit{ADPH} into a canonical form with the minimal number of free parameters, favours the implementation of estimation algorithms (as will be shown in the companion paper Part 2.)

\textbf{References}


A Proof of Lemma 3

In order to prove Lemma 3, we have to compare two ADPHs, in CF1 form, of order $n$ with mean $m_n$, namely: $\tau_n(m_n)$ and $\tau_n^M(m_n)$. By assumption, $\tau_n(m_n)$ is such that its conditional absorption time in phase $n$ is a MDPH ($\tau_{n-1}^M(m_{n-1})$); hence, $\tau_n(m_n)$ is completely specified by the values of $n-1, m_{n-1}, a_n, p_n$. On the other hand, $\tau_n^M(m_n)$ is a MDPH; hence, its second moment (and cv) is uniquely specified by the values of $n$ and $m_n$.

Since we compare random variables of equal mean, their coefficient of variations are in the same relation ($<, =, >$) as their second moments. Hence, all the following considerations are based on second moments, only. The proof consists of two steps: first we derive the conditions to minimize the second moment $d_n$ of $\tau_n(m_n)$, then we show that the obtained minimal second moment of $\tau_n(m_n)$ is, in any case, not less than the second moment of $\tau_n^M(m_n)$.

First part: conditions that minimizes the second moment of $\tau_n(m_n)$.

Since the structure of the MDPH changes when the mean is equal to the order (26), we need to consider two different possible structures for $\tau_n(m_n)$ according to the two possible structures of the MDPH defined over the $n-1$ phases. These two structures are depicted in Figure 15 and depend whether $m_{n-1}$ is less or greater than $n-1$:

Structure I. - $m_{n-1} \leq n-1$: $\tau_n(m_n)$ for this case is shown in Figure 15-I.

Structure II. - $m_{n-1} > n-1$: $\tau_n(m_n)$ for this case is shown in Figure 15-II.

Now we have to find the values of $a_n$ and $p_n$ that minimizes $d_n$ for the possible two possible structures of Figure 15. Since $s_n = 1$, (23) and (24) become:

$$m_n = (1-a_n)m_{n-1} + \frac{1}{p_n}, \quad (32)$$

$$d_n = (1-a_n) \left( d_{n-1} + \frac{2m_{n-1}}{p_n} \right) + \frac{2-p_n}{p_n^2}. \quad (33)$$
Rearranging (32), we obtain:

$$a_n = 1 - \frac{m_n - 1}{m_{n-1}}.$$  \hspace{1cm} (34)

From (34), we can establish the range of variability of $p_n$ given $n, m_{n-1}$ and $m_n$. Since $a_n$ and $p_n$ are probabilities (i.e., $0 \leq a_n \leq 1$ and $0 \leq p_n \leq 1$), and $a_n$ is a monotonic function of $p_n$, we obtain:

$$\frac{1}{m_n} \leq p_n \leq \begin{cases} 1 - \frac{1}{m_n - m_{n-1}} & \text{if } m_n - m_{n-1} \leq 1 \\ \frac{1}{m_n - m_{n-1}} & \text{if } m_n - m_{n-1} > 1 \end{cases}.$$  \hspace{1cm} (35)
Substituting (34) into (33), we obtain, for the second moment:

$$d_n = \frac{m_n - 1}{p_n} \left( d_{n-1} + \frac{2m_{n-1}}{p_n} \right) + \frac{2 - p_n}{p_n^2} = \frac{d_{n-1}m_n}{m_{n-1}} + \frac{2m_n - 1 - \frac{d_{n-1}}{m_{n-1}}}{p_n}.$$ (36)

By examining (36), the following cases have to be distinguished:

- if $2m_n - 1 - \frac{d_{n-1}}{m_{n-1}} \leq 0 \rightarrow$ then $d_n$ decreases as $p_n$ decreases.

The lower bound for $p_n$ is $1/m_n$ with $a_n = 1$ at this lower limit (Equation 35). Hence, in this case, the structures I. and II. of Figure 15 with minimal $d_n$ transform into the structure 1. of Figure 16.

- if $2m_n - 1 - \frac{d_{n-1}}{m_{n-1}} > 0 \rightarrow$ then $d_n$ decreases as $p_n$ increases.

To determine the upper limit for $p_n$ two cases have to be considered:

- if $m_n - m_{n-1} \geq 1$: the upper limit for $p_n$ is $1/(m_n - m_{n-1})$ and the associated initial probability is $a_n = 0$ (Equation 35). So the structure I. (II.) of Figure 15 with minimal $d_n$ transform into the structure 2. (3.) of Figure 16.

- if $m_n - m_{n-1} < 1$: the upper bound of $p_n$ is 1, and the associated initial probability is $a_n = 1 - (m_n - 1)/m_{n-1}$ (Equation 35). So the structure I. (II.) of Figure 15 with minimal $d_n$ transform into the structure 4. (5.) of Figure 16.

Now, we have found the various structures that minimize the second moment $d_n$ of $\tau_n(m_n)$, as it is summarized on Figure 16.

**Second part: comparing the minimal structures for $\tau_n(m_n)$ with $\tau_n^M(m_n)$.**

A MDPH is uniquely defined (and hence also its second moment) by $m$ and $n$; however, its shape changes when the mean is equal to the order, and hence $\tau_n^M(m_n)$ can have two different possible structures:

**Structure A.** - $m_n \leq n$: $\tau_n^M(m_n)$ for this case is shown in Figure 17-A.

**Structure B.** - $m_n > n$: $\tau_n^M(m_n)$ for this case is shown in Figure 17-B.

In order to complete the proof, we have to compare all the five possible structures of $\tau_n(m_n)$ (Figure 16), that minimizes its second moment $d_n$, with the two possible structures of $\tau_n^M(m_n)$ (Figure 17), and to show that, in any case, the second moment of $\tau_n(m_n)$ is not less than the second moment of $\tau_n^M(m_n)$. Table 2 summarizes all the combinations to be examined together with some conditions useful for the comparison.

Note that Cases 3-A and 4-B are not possible; while for cases 1-A and 1-B the condition for $m_{n-1}$ does not play any role because $a_n = 1$ and hence the initial probabilities of all the preceding phases are 0 (see structure 1 in Figure 16).

Now we consider all the cases of Table 2 one by one, and we prove that the second moment of the structure corresponding to $\tau_n(m_n)$ (first digit in the case number) is always not less than the second moment of the structure corresponding to $\tau_n^M(m_n)$ (second capital
Figure 17: Different structures for $\tau_n^M(m_n)$

<table>
<thead>
<tr>
<th>Case</th>
<th>Conditions for $2m_n - 1 - \frac{d_{n-1}}{m_{n-1}}$, $m_n - m_{n-1}$, $m_{n-1}$, $m_n$ to compare</th>
<th>Structures to compare</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-A</td>
<td>$\leq 0$</td>
<td>$\leq n$</td>
</tr>
<tr>
<td>1-B</td>
<td>$\leq 0$</td>
<td>$&gt; n$</td>
</tr>
<tr>
<td>2-A</td>
<td>$&gt; 0$</td>
<td>$\geq 1$</td>
</tr>
<tr>
<td>2-B</td>
<td>$&gt; 0$</td>
<td>$\geq 1$</td>
</tr>
<tr>
<td>3-A</td>
<td>$&gt; 0$</td>
<td>$\geq 1$</td>
</tr>
<tr>
<td>3-B</td>
<td>$&gt; 0$</td>
<td>$\geq 1$</td>
</tr>
<tr>
<td>4-A</td>
<td>$&gt; 0$</td>
<td>$&lt; 1$</td>
</tr>
<tr>
<td>4-B</td>
<td>$&gt; 0$</td>
<td>$&lt; 1$</td>
</tr>
<tr>
<td>5-A</td>
<td>$&gt; 0$</td>
<td>$&lt; 1$</td>
</tr>
<tr>
<td>5-B</td>
<td>$&gt; 0$</td>
<td>$&lt; 1$</td>
</tr>
</tbody>
</table>

Table 2: Combination of cases for comparing the second moments of $\tau_n(m_n)$ and $\tau_n^M(m_n)$

letter in the case number). In each one of the following cases, the second moment of $\tau_n(m_n)$ is on the l.h.s and the second moment of $\tau_n^M(m_n)$ is on the r.h.s.

Case 1-A:

$$m_n^2 - m_n + m_n^2 \geq R(m_n)(1 - R(m_n)) + m_n^2$$

$$\frac{(I(m_n) + R(m_n))^2 - (I(m_n) + R(m_n)) \geq R(m_n)(1 - R(m_n))}{\geq 0} \quad \frac{2R(m_n)^2 + 2R(m_n)(I(m_n) - 1) \geq 0}{\geq 0}$$

Case 1-B:

$$m_n^2 - m_n + m_n^2 \geq (n) \left[ \left( \frac{m_n}{n} \right)^2 - \frac{m_n}{n} \right] + m_n^2$$

$$m_n^2 \geq \frac{m_n^2}{n} \quad \implies \quad 1 \geq \frac{1}{n}$$

Case 2-A:
Figure 18: Case 2-A: \( f_{2-A}(m_{n-1}, m_n) \) (solid line) and \( \frac{d f_{2-A}(m_{n-1}, m_n)}{d m_{n-1}} \) (hashed line) as a function of \( m_{n-1} \) when \( m_n = 5 \).

\[
R(m_{n-1})(1-R(m_{n-1}))+ (m_n-m_{n-1})^2 - (m_n-m_{n-1}) + m_n^2 \geq R(m_n)(1-R(m_n)) + m_n^2 \tag{37}
\]

\[
R(m_{n-1})(1 - R(m_{n-1})) + (m_n - m_{n-1})^2 - (m_n - m_{n-1}) - R(m_n)(1 - R(m_n)) \geq 0
\]

Let us denote the l.h.s. by \( f_{2-A}(m_{n-1}, m_n) \). For any value of \( m_n \), \( f_{2-A}(m_{n-1}, m_n) \) has the following properties:

- \( f_{2-A}(m_{n-1}, m_n) \) is continuous, since all of its terms are continuous.
- \( f_{2-A}(1, m_n) \geq 0 \), since the inequality
  \[
  f_{2-A}(1, m_n) = (m_n - 1)^2 - (m_n - 1) - R(m_n)(1 - R(m_n)) \geq 0,
  \]
  holds for any \( m_n \geq 2 \) (which is a condition of case 2-A).
- \( f_{2-A}(m_n - 1, m_n) = 0 \).
- The derivative of \( f_{2-A}(m_{n-1}, m_n) \) with respect to \( m_{n-1} \)
  \[
  \frac{d f_{2-A}(m_{n-1}, m_n)}{d m_{n-1}} = 2(I(m_{n-1}) + 1 - m_n)
  \]
  is negative under the conditions of case 2-A.

Hence \( f_{2-A}(m_{n-1}, m_n) \) can not be negative for \( m_{n-1} \in (1, m_n - 1) \) (see also Figure 18).

Case 2-B:

\[
R(m_{n-1})(1 - R(m_{n-1}))(n - m_{n-1})^2 - (m_n - m_{n-1}) \geq m_n^2 - m_n + m_n^2
\]

\[
R(m_{n-1})(1 - R(m_{n-1}))(m_n^2 n - 1 - 2m_{n-1} m_n + m_{n-1}) \geq 0
\]

\[
\left(\frac{n - 1}{n} m_n - \frac{n}{n-1} m_{n-1}\right)^2 + \left(1 - \frac{m_{n-1}}{n-1}\right)^2 \geq 0
\]

26
Case 3-B:

\[ (n-1) \left( \frac{m_{n-1}^2}{n-1} - \frac{m_{n-1}}{n-1} \right) + (m_n - m_{n-1})^2 - (m_n - m_{n-1}) + m_n^2 \geq \frac{m_n^2}{n} - m_n + m_n^2 \]

\[ \frac{n-1}{n} m_n^2 - 2m_{n-1}m_n + \frac{n}{n-1} m_{n-1}^2 \geq 0 \]

\[ \left( \sqrt{\frac{n-1}{n} m_n} - \sqrt{\frac{n}{n-1} m_{n-1}} \right)^2 \geq 0 \]

Case 4-A:

![Figure 19: Case 4-A: \( f_{4-A}(m_{n-1}, m_n) \) (solid line) and \( \frac{d f_{4-A}(m_{n-1}, m_n)}{dm_n} \) (hashed line) as a function of \( m_n \) when \( m_{n-1} = 5 \)](image)

\[ \frac{m_n - 1}{m_{n-1}} R(m_{n-1})(1 - R(m_{n-1})) + (m_n)^2 + 1 - \frac{m_n - 1}{m_{n-1}} \geq R(m_n)(1 - R(m_n)) + m_n^2 \]  (38)

Moving everything to the left side, using the function \( I(m) \) instead of \( R(m) \) we have

\[ \frac{1}{m_{n-1}} (2m_{n-1}m_n - 2m_{n-1} + I(m_{n-1})(1 - 2m_{n-1} + I(m_{n-1})) - m_n + 2m_{n-1}m_n \]

\[ + I(m_n)(m_{n-1} - I(m_{n-1})m_n) - 2m_{n-1}m_n + m_{n-1}I(m_n)) \geq 0 \]

Following a similar approach as in case 2-A, let us denote the left hand side by \( f_{4-A}(m_{n-1}, m_n) \).

It is easy to check that for any value of \( m_{n-1} \)

- \( f_{4-A}(m_{n-1}, m_n) \) is continuous,
- \( f_{4-A}(m_{n-1}, 1) \geq 0, \)
- \( f_{4-A}(m_{n-1}, m_{n-1} + 1) \geq 0, \)
- the derivative of \( f_{4-A}(m_{n-1}, m_n) \) with respect to \( m_n \) is positive at \( m_n = 1 \) and is monotonically decreasing under the conditions of case 4-A (Figure 19).
Figure 20: Case 5-A: $f_{5-A}(m_{n-1}, m_n)$ (solid line) and $\frac{d}{d m_n} f_{5-A}(m_{n-1}, m_n)$ (hashed line) as a function of $m_n$ when $n = 5$ and $m_{n-1} = 5$

So that the inequality (38) holds under the conditions of case 4-A.

**Case 5-A:**

$$\frac{m_n - 1}{m_{n-1}} \left( (n - 1) \left( \frac{m_{n-1}}{n - 1} \right)^2 + 1 - (m_{n-1} + 1) + (m_{n-1} + 1)^2 \right) + 1 - \frac{m_n - 1}{m_{n-1}} \geq R(m_n) (1 - R(m_n)) + m_n^2$$

$$m_n + m_{n-1} m_n + \frac{m_{n-1} m_n}{n - 1} - m_{n-1} - \frac{m_{n-1}}{n - 1} \geq R(m_n) (1 - R(m_n)) + m_n^2 \quad (39)$$

Using the same technique as for case 4-A, we have

$$m_n^2 + m_{n-1} m_n + \frac{m_{n-1} m_n}{n - 1} - m_{n-1} - \frac{m_{n-1}}{n - 1} + I(m_n) - 2 m_n I(m_n) \geq 0$$

Denoting the left hand side by $f_{5-A}(m_{n-1}, m_n)$ it is easy to show that

- $f_{5-A}(m_{n-1}, m_n)$ is continuous,
- $f_{5-A}(m_{n-1}, 1) \geq 0$,
- $f_{5-A}(m_{n-1}, n) \geq 0$,
- the derivative of $f_{5-A}(m_{n-1}, m_n)$ with respect to $m_n$ is positive for $m_n = 1$ and monotone decreasing under the conditions of case 5-A (Figure 20).

As a consequence (39) holds.

**Case 5-B:**

$$m_n + m_{n-1} m_n + \frac{m_{n-1} m_n}{n - 1} - m_{n-1} - \frac{m_{n-1}}{n - 1} \geq \frac{m_n^2}{n} - m_n + m_n^2$$

$$- m_{n-1} - \frac{m_{n-1}}{n - 1} + 2 m_n + m_{n-1} m_n + \frac{m_{n-1} m_n}{n - 1} - m_n^2 - \frac{m_n^2}{n} \geq 0$$

$$\left( m_n - 1 \right) \left( 1 + \frac{1}{n - 1} \right) (m_{n-1} + 1 - m_n) + \frac{(m_n - (n-1) - 1)^2}{(n-1)n} \geq 0$$

Which concludes the proof of Lemma 3.