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

Andrea Bobbio[1], András Horváth[2], Marco Scarpa[3], Miklós Telek[2†]

[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.iit.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, *DPH*s

*This work has been performed under a cooperative visiting exchange program supported by the Italian Ministry of Foreign Affairs and the Hungarian Committee for Technologycal 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 emphasise 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:

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

where $\boldsymbol{B} = [b_{ij}]$ is the $(n \times n)$ matrix collecting the transition probabilities among the transient states, $\boldsymbol{b} = [b_{i,n+1}]^T$ 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 $\widehat{\boldsymbol{B}}$ 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 is an $(n+1)$ dimensional vector $\widehat{\boldsymbol{\alpha}} = [\boldsymbol{\alpha}, \alpha_{n+1}]$, with $\sum_{j=1}^{n} \alpha_j = 1 - \alpha_{n+1}$. In the present paper, we only consider the class of *DPH* distributions 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 $(\boldsymbol{\alpha}, \boldsymbol{B})$ [10]. Let $f(k)$, $F(k)$ and $\mathcal{F}(z)$ be the probability mass, cumulative probability and probability generating function of τ, respectively. It follows:

$$\begin{aligned}
f(k) &= Pr\{\tau = k\} = \boldsymbol{\alpha}\, \boldsymbol{B}^{k-1}\, \boldsymbol{b} \qquad \text{for } k > 0 & (1)\\
F(k) &= Pr\{\tau \le k\} = \boldsymbol{\alpha} \sum_{i=1}^{k-1} \boldsymbol{B}^{i}\, \boldsymbol{b} = 1 - \boldsymbol{a}\boldsymbol{B}^{k}\, \boldsymbol{e} & (2)\\
\mathcal{F}(z) &= E\{z^{\tau}\} = z\, \boldsymbol{\alpha}\, (\boldsymbol{I} - z\boldsymbol{B})^{-1}\, \boldsymbol{b} = \frac{N(z)}{D(z)} & (3)
\end{aligned}$$

where $\boldsymbol{e}$ is an n-dimensional column vector with all the entries equal to 1 and $\boldsymbol{I}$ 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 $\boldsymbol{B}$ and $n-1$ in vector $\boldsymbol{\alpha}$). Therefore, the matrix representation is very redundant with respect to the degrees of freedom, and it is reasonable to look for minimal representations.

Let ϕ_i be the sojourn time the *DPH* (the *DTMC*) spends in phase i. ϕ_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 *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 [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 *CPH*s does not hold for *DPH*'s. In fact, it is clear that for any order n, the *DPH* with representation $(\boldsymbol{\alpha}, \boldsymbol{J})$ given by:

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

represents a deterministic time to absorption $\tau = n$, with $cv = 0$. Hence for any order n there exists at least one *DPH* with $cv = 0$. 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 τ_C and τ_D be the *CPH* and *DPH* r.v. shown in Figure 1, with representations $(\boldsymbol{\gamma}, \boldsymbol{\Lambda})$ and $(\boldsymbol{\alpha}, \boldsymbol{B})$, respectively:

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

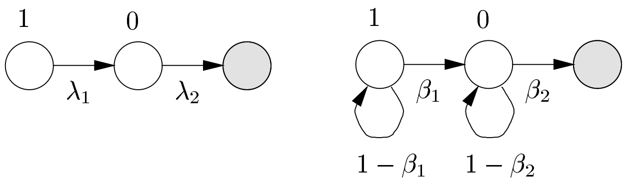

Figure 1: Two-state *CPH* and *DPH* structures

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

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

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:

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

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:

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

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* $\boldsymbol{B}$ *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 $\boldsymbol{B}$ and $n-1$ in the initial probability vector $\boldsymbol{\alpha}$).

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 $\ell \le n$, is characterized by a set of indices, representing the states visited before absorption:

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

where the last two conditions mean that in a path any two subsequent indices represent states that must be connected by a direct arc (non-zero entry in the $\boldsymbol{B}$ 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:

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

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

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

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]):*

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

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

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

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:

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

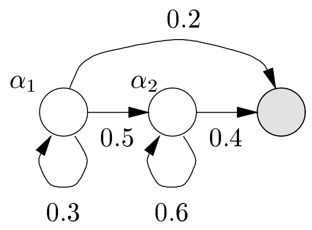

Figure 2: An example of *ADPH*.

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

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

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

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

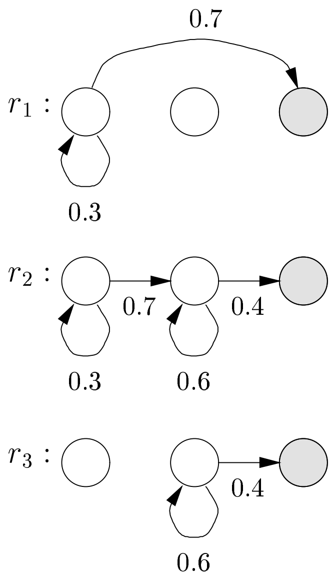

Figure 3: Possible paths of the *ADPH* of Figure 2.

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 $\boldsymbol{B}$ 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 $\boldsymbol{u}_k = [u_i]$ 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 $\boldsymbol{u}_k$. With this representation any path is characterized by a unique binary number $1 \leq \#\boldsymbol{u}_k \leq 2^n - 1$.

Definition 2 *A path r_k of length ℓ of an* ADPH *is called a* **basic path** *(basic series [7]) 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 $\#\boldsymbol{u}_k = 2^\ell - 1$.*

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:

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

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 $\boldsymbol{u}_k$. Starting from the rightmost component of $\boldsymbol{u}_k$ 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 $\boldsymbol{u}_k = [0, 1, 0, 1, 0]$ (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 $\boldsymbol{u}_k$ 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:

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

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.

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

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* $(\boldsymbol{a}, \boldsymbol{P})$ *takes the form:*

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

$$\text{with:} \qquad \sum_1^n a_i = 1 \quad \text{and:} \quad p_1 \leq p_2 \leq \ldots \leq p_n$$

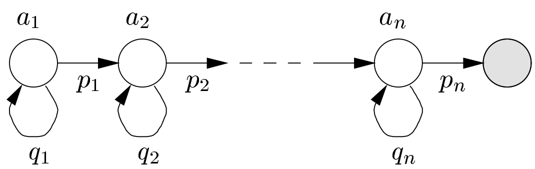

Figure 4: Canonical Form *CF1.*

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 $\boldsymbol{p}$ containing the p_i's, so that we will use the notation $(\boldsymbol{a}, \boldsymbol{p})$ as the representation of a *CF1*, where $\boldsymbol{a}$ and $\boldsymbol{p}$ 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 $\boldsymbol{B}$ 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 $\boldsymbol{b}_1 = [0, 1]$ and $\boldsymbol{b}_2 = [1, 1]$.

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

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

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

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

Path r_3 - The associated binary vector is $\boldsymbol{u}_3 = [1, 0]$ 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:

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

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

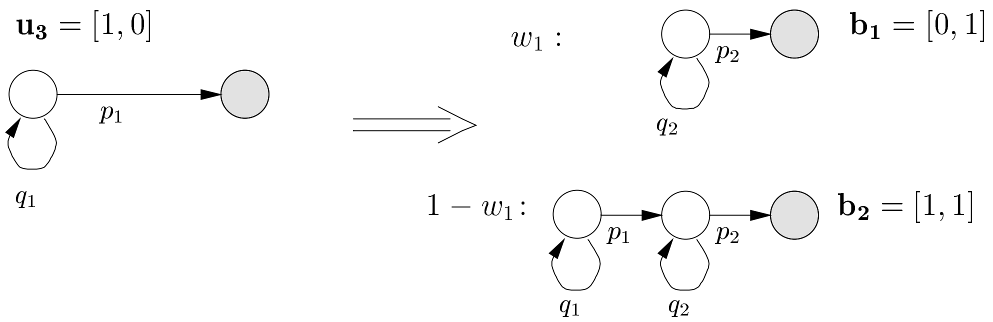

Figure 5: Transformation of the path r_3.

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

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

Equation (15) can be rearranged in the following *CF1* form, with $a_1 = \left(\frac{2}{7}\alpha_1 + \frac{4}{7}\alpha_2\right)$, and $a_2 = \left(\frac{5}{7}\alpha_1 + \frac{3}{7}\alpha_2\right)$:

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

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

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

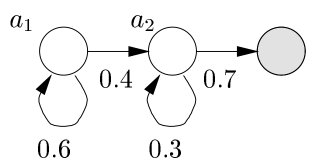

Figure 6: Canonical form of the ADPH of Figure 2.

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 $(\boldsymbol{a}, \boldsymbol{p})$ (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 [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:

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

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

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

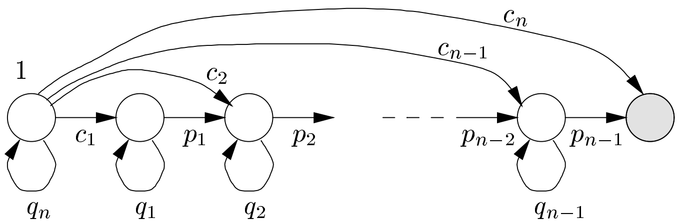

Figure 7: Canonical Form *CF2*

Definition 4 *Canonical Form CF3.*
An ADPH is in canonical form CF3 (Figure 8) if from any phase i ($1 \le i \le n$) 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:

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

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

$$\begin{aligned} s_i &= \sum_{j=1}^{i} a_j \\ e'_i &= \frac{a_i}{s_i} p_i \\ e_i &= \frac{s_{i-1}}{s_i} p_i \end{aligned}$$

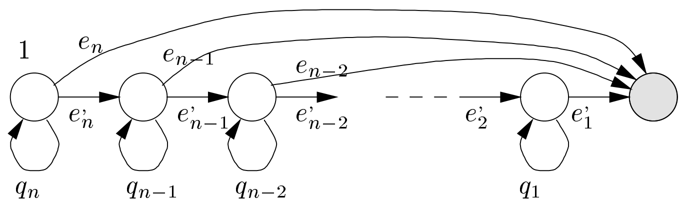

Figure 8: Canonical Form *CF3*

3.2 Conditional absorption time

If an *APDH* r.v. τ 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.* τ *of order* n *with CF1 representation* $(\boldsymbol{a}, \boldsymbol{p})$*, the conditional absorption time* τ_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 preceeding phases (from phase 1 to i).*

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

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

Let us denote by m_i, d_i and $\mathcal{F}_i$ the mean, the second moment and the generating function of the conditional absorbing time τ_i, respectively. The following recursive relations hold:

$$m_{i+1} = \frac{s_i}{s_{i+1}} m_i + \frac{1}{p_{i+1}} , \tag{23}$$

$$d_{i+1} = \frac{s_i}{s_{i+1}} \left(d_i + \frac{2m_i}{p_{i+1}} \right) + \frac{2 - p_{i+1}}{(p_{i+1})^2} , \tag{24}$$

$$\mathcal{F}_{i+1}(z) = \frac{s_i}{s_{i+1}} \mathcal{F}_i(z) \, \frac{p_{i+1} z}{1 - (1 - p_{i+1}) z} + \frac{a_{i+1}}{s_{i+1}} \, \frac{p_{i+1} z}{1 - (1 - p_{i+1}) z} . \tag{25}$$

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^2_{min}*, of an* ADPH *r.v.* $\tau_n(m)$ *of order* n *with mean* m *is:*

$$cv^2_{min}(\tau_n(m)) = \begin{cases} \dfrac{R(m)(1 - R(m))}{m^2} & \text{if } m \leq n , \\[2ex] \dfrac{1}{n} - \dfrac{1}{m} & \text{if } m > n , \end{cases} \tag{26}$$

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

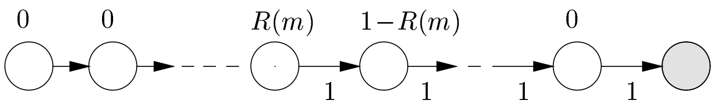

Figure 9: MDPH with $m \leq n$

- *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).

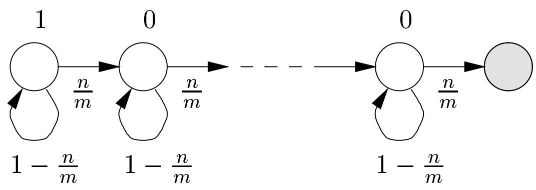

Figure 10: MDPH with $m > n$

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 $m \geq n$ 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 $m^\tau = m^\theta = m$ and representation $(\boldsymbol{a}, \boldsymbol{p})$ and $(\boldsymbol{\alpha}, \boldsymbol{\pi})$, 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 τ_k and θ_k be the conditional absorption time in phase $k+1$ in the two* ADPH*'s, respectively, and let m_k^τ, m_k^θ, d_k^τ, d_k^θ and cv_k^τ, cv_k^θ be their mean values, second moments, and coefficients of variation, respectively. By construction, $m_k^\tau = m_k^\theta$. 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, τ_k has representation $(\boldsymbol{a}^{(k)}/s_k^\tau, \boldsymbol{p}^{(k)})$ and θ_k has representation $(\boldsymbol{\alpha}^{(k)}/s_k^\theta, \boldsymbol{p}^{(k)})$ (with $s_k^\tau = s_k^\theta = s_k$ by assumption). We can assume, without loss of generality that:

$$cv(\tau_k) \leq cv(\theta_k) \tag{27}$$

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

$$\sum_{i=1}^{k} \frac{a_i^{(k)}}{s_k} \left[\sum_{j=i}^{k} \left(\frac{1}{(p_j^{(k)})^2} - \frac{1}{p_j^{(k)}} \right) + \left(\sum_{j=i}^{k} \frac{1}{p_j^{(k)}} \right)^2 \right] \leq \sum_{i=1}^{k} \frac{\alpha_i^{(k)}}{s_k} \left[\sum_{j=i}^{k} \left(\frac{1}{(\pi_j^{(k)})^2} - \frac{1}{\pi_j^{(k)}} \right) + \left(\sum_{j=i}^{k} \frac{1}{\pi_j^{(k)}} \right)^2 \right],$$

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

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

$$\sum_{i=k+1}^{n} \alpha_i \left[\sum_{j=i}^{n} \left(\frac{1}{\pi_j^2} - \frac{1}{\pi_j} \right) + \left(\sum_{j=i}^{n} \frac{1}{\pi_j} \right)^2 \right] + \sum_{i=1}^{k} \alpha_i \left[\sum_{j=k+1}^{n} \left(\frac{1}{\pi_j^2} - \frac{1}{\pi_j} \right) + \left(\sum_{j=k+1}^{n} \frac{1}{\pi_j} \right)^2 \right]$$

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:

$$\sum_{i=1}^{n} a_i \left[\sum_{j=i}^{n} \left(\frac{1}{p_j^2} - \frac{1}{p_j} \right) + \left(\sum_{j=i}^{n} \frac{1}{p_j} \right)^2 \right] \leq \sum_{i=1}^{n} \alpha_i \left[\sum_{j=i}^{n} \left(\frac{1}{\pi_j^2} - \frac{1}{\pi_j} \right) + \left(\sum_{j=i}^{n} \frac{1}{\pi_j} \right)^2 \right],$$

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 $\tau_n^M(m_n)$ 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:*

$$cv(\tau_n(m_n)) \geq cv(\tau^M_n(m_n)) \ . \tag{28}$$

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 $\tau^M_1(m_1)$

$$\tau_1(m_1) = \tau^M_1(m_1) \tag{29}$$

(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.

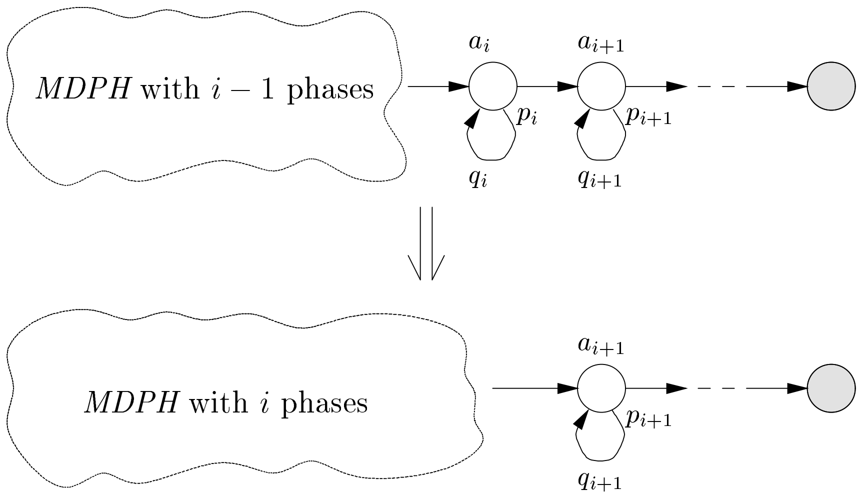

Figure 11: One step of the transformation

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: $\tau_2^M(m_2)$ (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. $\tau_n^{(2)}(m_n)$ (Figure 12-4).

Step 3. - The r.v. $\tau_n^{(2)}(m_n)$ 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:

$$cv_1^2 = 0.285, \qquad cv_4^2 = 0.098, \qquad cv_5^2 = 0.018.$$

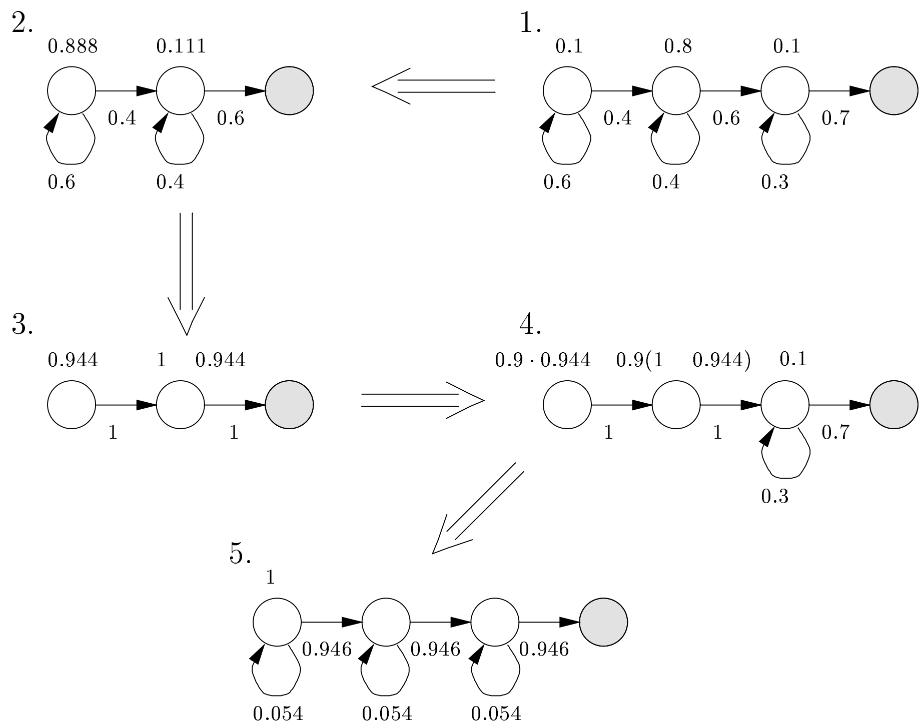

Figure 12: Transformation of a *CF1-ADPH* into the *MDPH* with the same mean

4.1 Comparing the minimal $\mathbf{cv^2}$ 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:

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

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

$$n_{Cmin} = 4\, I(m)\, (1 + I(m)) \tag{31}$$

$I(m)$	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 $I(m)$

From Equation (31) 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 *CDH* class provides a minimal cv^2 less than the *ADPH* of the same order.

Since for an *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 *CPH* family exists whose minimal cv^2 is less than the one of the *ADPH*.

Example 6 - Figure 13 shows the minimal cv^2 as a function of the number of phases for the *ADPH* family versus the *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 *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 (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 *CPH* has a cv^2 less than the *ADPH* 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 *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 $cv^2_{min} \to 1/n$ as $m \to \infty$. Hence, as m increases, the behavior of the *ADPH* class tends to be similar to the one of the *CPH* class.

Example 7 - Figure 14 shows the minimal cv^2 as a function of the mean for a *CF1-ADPH* 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 $m \to \infty$.

5 Conclusion

The properties of the *DPH* distributions 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 *ADPH* family is less than the one for the *CPH* family of the same order. Hence, the famous theorem of Aldous and Shepp about the minimal coefficient of variation for the *CPH* class does not hold for the *DPH* class.

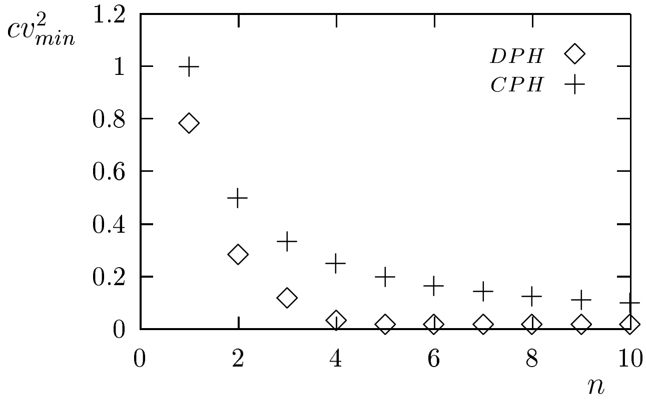

Figure 13: Minimal squared coefficient of variation for $m = 4.5$

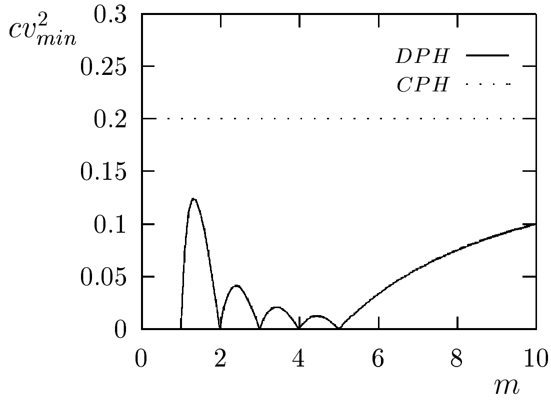

Figure 14: Minimal squared coefficient of variation for $n = 5$

The results about the wider variability of the *ADPH* class can be very relevant in stochastic modeling. When *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 *ADPH* class, the use of *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 *ADPH* with the modeling power of other formalisms, like Petri nets.

Finally, the possibility of representing any *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.)

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

[4] 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.

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

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

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

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

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

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

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

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

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

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

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

[16] 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.

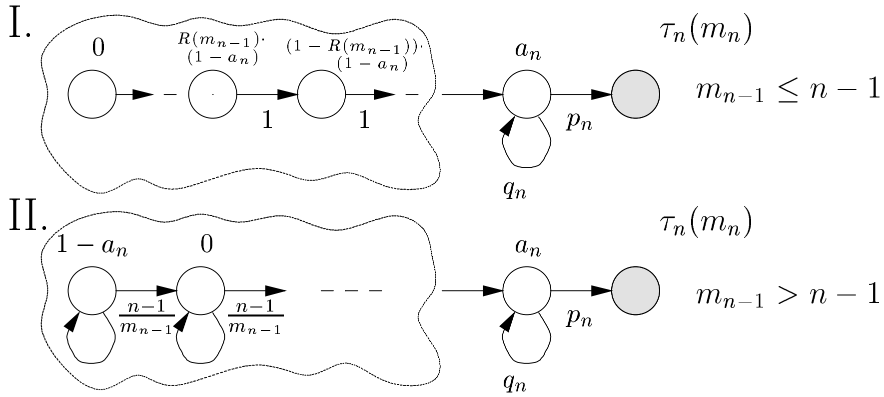

Figure 15: Different structures for $\tau_n(m_n)$

A Proof of Lemma 3

In order to prove Lemma 3, we have to compare two *ADPH*s, 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. - $m_{n-1} \le n-1$: $\tau_n(m_n)$ 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:

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

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

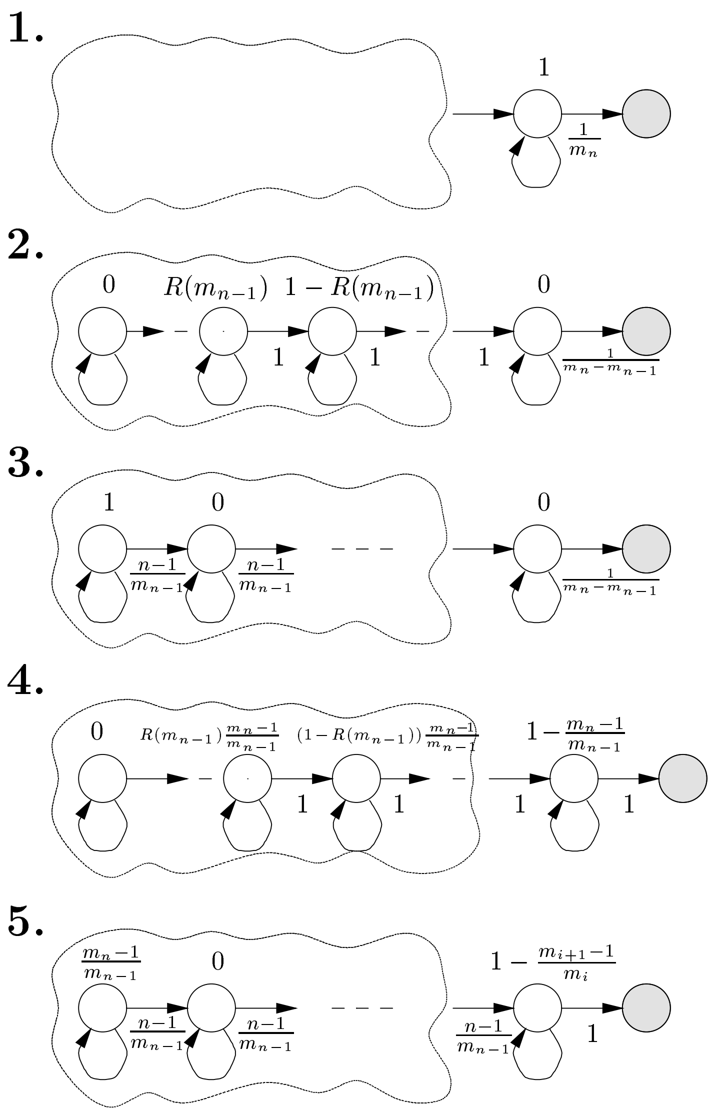

Figure 16: The structures of $\tau_n(m_n)$ with minimal second moment d_n

Rearranging (32), we obtain:

$$a_n = 1 - \frac{m_n - \dfrac{1}{p_n}}{m_{n-1}} \ . \tag{34}$$

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:

$$\frac{1}{m_n} \leq p_n \leq \begin{cases} 1 & \text{if} \quad m_n - m_{n-1} \leq 1 \\ \dfrac{1}{m_n - m_{n-1}} & \text{if} \quad m_n - m_{n-1} > 1 \end{cases} \tag{35}$$

Substituting (34) into (33), we obtain, for the second moment:

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

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

- if $2m_n - 1 - \dfrac{d_{n-1}}{m_{n-1}} \leq 0 \longrightarrow$ 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 $2m_n - 1 - \dfrac{d_{n-1}}{m_{n-1}} > 0 \longrightarrow$ 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 preceeding 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

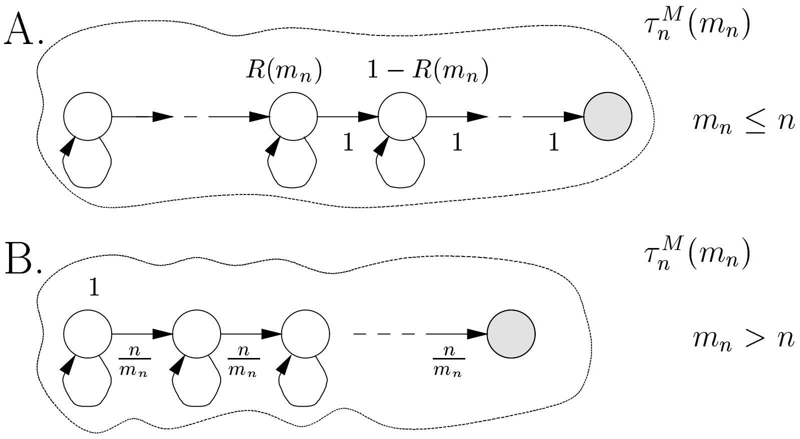

Figure 17: Different structures for $\tau_n^M(m_n)$

Case	Conditions for				Structures to compare	
	$2m_n - 1 - \frac{d_{n-1}}{m_{n-1}}$	$m_n - m_{n-1}$	m_{n-1}	m_n		
1-A	≤ 0			$\leq n$	1. of Fig. 16	A. of Fig 17
1-B	≤ 0			$> n$	1. of Fig. 16	B. of Fig 17
2-A	> 0	≥ 1	$\leq n-1$	$\leq n$	2. of Fig. 16	A. of Fig 17
2-B	> 0	≥ 1	$\leq n-1$	$> n$	2. of Fig. 16	B. of Fig 17
3-A	> 0	≥ 1	$> n-1$	$\leq n$	Not possible	
3-B	> 0	≥ 1	$> n-1$	$> n$	3. of Fig. 16	B. of Fig 17
4-A	> 0	< 1	$\leq n-1$	$\leq n$	4. of Fig. 16	A. of Fig 17
4-B	> 0	< 1	$\leq n-1$	$> n$	Not possible	
5-A	> 0	< 1	$> n-1$	$\leq n$	5. of Fig. 16	A. of Fig 17
5-B	> 0	< 1	$> n-1$	$> n$	5. of Fig. 16	B. of Fig 17

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:

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

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

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

Case 1-B:

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

$$m_n^2 \geq \frac{m_n^2}{n} \qquad \Longrightarrow \qquad 1 \geq \frac{1}{n}$$

Case 2-A:

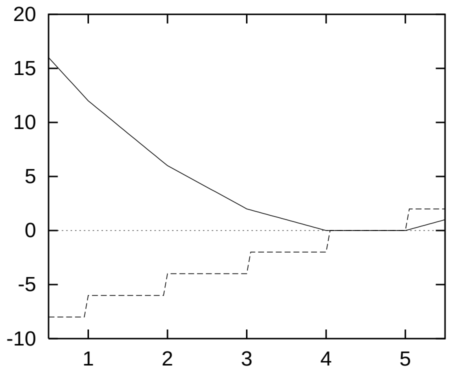

Figure 18: Case 2-A: $f_{2-A}(m_{n-1}, m_n)$ (solid line) and $\frac{d\ f_{2-A}(m_{n-1},m_n)}{dm_{n-1}}$ (hashed line) as a function of m_{n-1} when $m_n = 5$.

$$R(m_{n-1})(1-R(m_{n-1}))+(m_n-m_{n-1})^2-(m_n-m_{n-1})+m_n^2 \geq R(m_n)(1-R(m_n))+m_n^2 \quad (37)$$

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

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}

  $$\frac{d\ f_{2-A}(m_{n-1}, m_n)}{dm_{n-1}} = 2(I(m_{n-1}) + 1 - m_n)$$

  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:

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

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

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

Case 3-B:

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

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

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

Case 4-A:

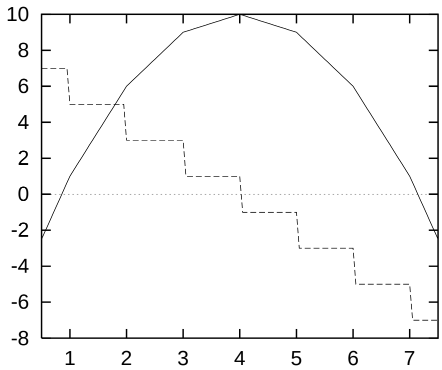

Figure 19: Case 4-A: $f_{4-A}(m_{n-1},m_n)$ (solid line) and $\frac{d\ f_{4-A}(m_{n-1},m_n)}{dm_n}$ (hashed line) as a function of m_n when $m_{n-1}=5$

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

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).

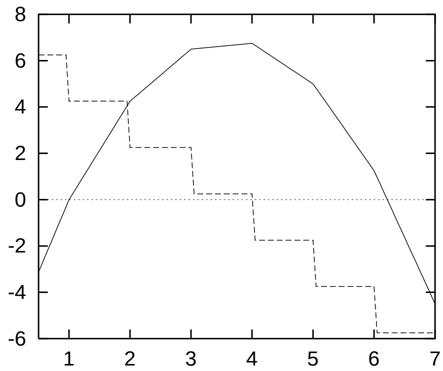

Figure 20: Case 5-A: $f_{5-A}(m_{n-1}, m_n)$ (solid line) and $\frac{d\ f_{5-A}(m_{n-1},m_n)}{dm_n}$ (hashed line) as a function of m_n when $n = 5$ and $m_{n-1} = 5$

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

Case 5-A:

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

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

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

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:

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

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

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

Which concludes the proof of Lemma 3.