The Minimal Coefficient of Variation of Discrete Phase Type Distributions

Miklós Telek[1]

Abstract: The paper provides the lower bound of the coefficient of variation of discrete phase type distributions that is a discrete time counterpart of the well known result by Aldous and Shepp [1].

Keywords: Discrete Phase Type Distribution, Minimal Coefficient of Variation.

1 Introduction

Discrete Phase Type (DPH) distributions have been used for some time [3], but they have received less attention than Continuous Phase Type *(CPH)* distributions because continuous time models were more popular in stochastic modelling. Recent attention toward discrete time stochastic models initiated new research on DPH distributions. It was well known that the DPH class has different properties with respect to the minimal coefficient of variation than the CPH class, since the deterministic distribution (with zero coefficient of variation) is a member of the DPH class, but to the best of our knowledge the bound of the coefficient of variation of DPH distributions was first considered in [2]. Indeed [2] provided a lower bound that is valid for the whole DPH class, as it shown in this paper, but that bound is proved to be valid only for the Acyclic DPH class in [2]. This paper focuses only on the proof of the bound, while a detailed study of the modeling power of the DPH class considering the bound on the coefficient of variation is provided in [2].

[1]Department of Telecommunications, Technical University of Budapest, Sztoczek u. 2., 1111 Budapest, Hungary. *telek@hit.bme.hu*

2 Model description and notations

Let $\boldsymbol{X} = \{X_k, k = 0, 1, \ldots\}$ be a time-homogeneous discrete-time Markov chain (DTMC) over $S = \{0, 1, \ldots, N\}$, with N transients states, $\{1, \ldots, N\}$, and an absorbing one, 0. The state transition probability matrix of $\boldsymbol{X}$ is $\boldsymbol{\Pi} = \{\pi_{ij}\}$. The unconditional and the conditional time to absorption is denoted as

$$\tau = \min\{k : X_k = 0\}$$

and

$$\tau_i = \min\{k : X_k = 0 \text{ given that } X_0 = i\}.$$

Let $\mu = \mathbb{E}\tau$ and $G(i) = \mathbb{E}\tau_i$. Without loss of generality we assume that (the states are numbered such that)

$$0 = G(0) < 1 \leq G(1) \leq G(2) \leq \ldots \leq G(N). \tag{1}$$

$G(i)$ satisfies

$$G(i) = 1 + \sum_{j \in S} \pi_{ij} G(j) \ . \tag{2}$$

The initial distribution of $\boldsymbol{X}$ is given as $p_i = \Pr(X_0 = i)$, hence τ is a DPH distribution of order N with mean

$$\mu = \sum_{i \in S} p_i G(i). \tag{3}$$

Further more, $\lfloor x \rfloor$ and $\langle x \rangle$ denote the integer and fraction part of x, respectively, i.e., $x = \lfloor x \rfloor + \langle x \rangle$, such that $\lfloor x \rfloor$ is an integer and $0 \leq \langle x \rangle < 1$.

3 Problem formulation

A significant difference between the minimal coefficient of variation of the DPH and the CPH class can be observed in the following example.

Example 1: The simplest DPH distribution, the DPH of order 1, i.e., the geometric distribution with pmf $Pr(\tau = k) = (1-\pi_{10})^{k-1}\pi_{10}$ *has the following properties:*

$$\mu = G(1) = \mathbb{E}\tau_1 = 1/\pi_{10},$$

and

$$cv^2(\tau_1) = 1 - \pi_{10} = 1 - 1/G(1) = 1 - 1/\mu.$$

That is, in contrast with the CPH class, the minimal coefficient of variation of DPH distributions is a function of its mean. Hence in the DPH case the following constrained minimizations have to be solved:

$$\min_{\mathbf{\Pi}}\{cv^2(\tau) \mid \mathbb{E}\tau\} \quad \text{and} \quad \min_{\mathbf{\Pi}}\{cv^2(\tau_i) \mid \mathbb{E}\tau_i\}$$

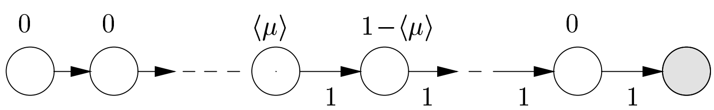
where τ, τ_i, $i \in S$ and $\mathbf{\Pi}$ are related through (2) and (3). Note also that the states are numbered according to (1) which plays an important role in the initial state dependent cases.

4 The minimal coefficient of variation of DPH distributions

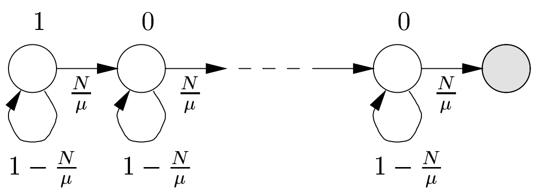
Theorem 1 *The squared coefficient of variation of τ, $cv^2(\tau)$, satisfies the inequality:*

$$cv^2(\tau) \geq \begin{cases} \dfrac{\langle\mu\rangle(1-\langle\mu\rangle)}{\mu^2} & \text{if } \mu < N \ , \\ \dfrac{1}{N} - \dfrac{1}{\mu} & \text{if } \mu \geq N \ . \end{cases} \tag{4}$$

- *a DPH distribution which satisfies the equality if $\mu \leq N$ is the following: the nonzero initial probabilities are $p_{N-\lfloor\mu\rfloor} = \langle\mu\rangle$, $p_{N-\lfloor\mu\rfloor+1} = 1-\langle\mu\rangle$ and the transition probabilities are $\Pr(X_1 = i-1 | X_0 = i) = 1, \forall i \in S$.*



- *the only DPH distribution which satisfies the equality if $\mu > N$ is the following:*
  the nonzero initial probability is $p_N = 1$ and the transition probabilities are $\Pr(X_1 = i-1|X_0 = i) = N/\mu, \Pr(X_1 = i|X_0 = i) = 1-N/\mu, \forall i \in S$ (discrete Erlang(N) distribution).



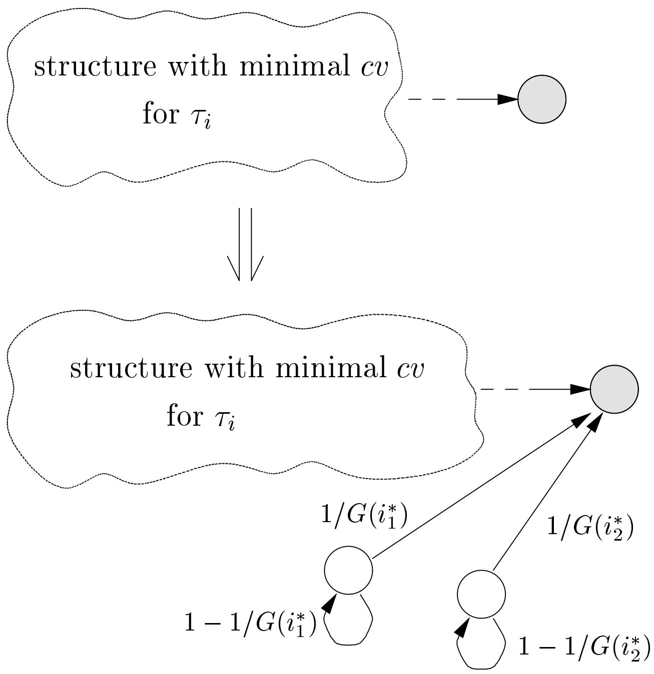
To prove the theorem we need the following lemmas.

Lemma 1 *The minimal coefficient of variation of τ_i does not increase when n extra states, $i_1^*, i_2^*, \ldots, i_n^*$, with arbitrary mean time to absorption, $G(i_1^*), \ldots, G(i_n^*) \geq 1$, is added to the DPH structure.*

Proof of Lemma 1:

We prove the lemma by providing a DPH structure of $N+n$ phases that has the same coefficient of variation as the minimal coefficient of variation of τ_i without the extra states:

- the state transition probabilities between the original states are the same as in the DPH that provides the minimal coefficient of variation of τ_i without the extra states
- from each extra state the non-zero transition probabilities are $\pi_{i_j^*,0} = 1/G(i_j^*)$ and $\pi_{i_j^*,i_j^*} = 1 - 1/G(i_j^*)$



Lemma 2 *The minimal squared coefficient of variation of τ_i is as follows:*

$$cv^2(\tau_i) \geq \begin{cases} \dfrac{\langle G(i)\rangle(1-\langle G(i)\rangle)}{G^2(i)} & \text{if } G(i) < i \ , \\[2ex] \dfrac{1}{i} - \dfrac{1}{G(i)} & \text{if } G(i) \geq i \ . \end{cases} \tag{5}$$

Note that Lemma 2 is valid for all $i \in S$ where the states are numbered according to (1).

Proof of Lemma 2:

The proof of Lemma 2 is composed by giving two lower bounds on the variance. According to our interpretation the first one is closely related to the degree of the considered DPH distribution while the second one is related to the structural properties of the DPH class. We refer to the bounds based on this classification. (A short explanation of these properties is provided after the proof.) The lower bound of Lemma 2, (5), is obtained as the larger of the two bounds.

Bound of variance related to the degree of DPH distributions

A lower bound of the variance of the DPH distributions can be obtained by applying the elegant martingale approach proposed by Aldous and Shepp [1]. The discrete time stochastic process $\boldsymbol{Y} = \{Y_k, k = 0, 1, \ldots\}$ is defined as

$$Y_k = G(X_k) + \min(k, \tau_i) - G(X_0) \tag{6}$$

assuming $X_0 = i$.

$\boldsymbol{Y}$ is a martingale since,

- if $k \geq \tau_i$ then $Y_{k+1} = Y_k = \tau_i - G(i)$ is constant; and
- if $k < \tau_i$ then

$$\begin{aligned} &\mathbb{E}(Y_{k+1}|X_k) = \mathbb{E}(G(X_{k+1})|X_k) + \min(k+1, \tau_i) - G(i) = \\ &\sum_{j \in S} \pi_{X_k,j} G(j) + \min(k+1, \tau_i) - G(i) = \\ &G(X_k) - 1 + (k+1) - G(i) = G(X_k) + k - G(i) \ , \end{aligned}$$

  where (2) has been applied in the second line.

By the definition of $\boldsymbol{Y}$, since $G(X_{\tau_i}) = 0$, we have

$$Y_{\tau_i} = \tau_i - G(i) \tag{7}$$

and

$$\mathbb{E} Y_{\tau_i}^2 = \mathrm{var}(\tau_i) \ . \tag{8}$$

For $k \leq \tau_i$, using martingale properties, we have:

$$\begin{aligned} &Y_k^2 = \sum_{i=1}^{k} (Y_i^2 - Y_{i-1}^2) = \sum_{i=1}^{k} (Y_i - Y_{i-1})^2 = \sum_{i=1}^{k} (G(X_i) - G(X_{i-1}) + 1)^2 = \\ &\sum_{i=1}^{k} (G(X_i) - G(X_{i-1}))^2 + 2 \ \sum_{i=1}^{k} (G(X_i) - G(X_{i-1})) + k \end{aligned}$$

We define $S_k = \sum_{s=1}^{k} (G(X_s) - G(X_{s-1}))^2$, which gives

$$Y_k^2 = S_k + 2(G(X_k) - G(i)) + k \ .$$

Note that $G(X_s) - G(X_{s-1})$ takes non-zero values only at state transitions. For $k = \tau_i$

$$Y_{\tau_i}^2 = S_{\tau_i} + 2(G(X_{\tau_i}) - G(i)) + \tau_i = S_{\tau_i} - 2G(i) + \tau_i \ ,$$

and

$$S_{\tau_i} = \sum_{s: X_s \neq X_{s-1}} (G(X_s) - G(X_{s-1}))^2 \geq \sum_{j=1}^{i} (G(j) - G(j-1))^2$$
$$\geq \frac{1}{i} \left(\sum_{j=1}^{i} G(j) - G(j-1) \right)^2 = \frac{1}{i} \ G^2(i) \ .$$

The first inequality says that the sequential path $(i, i-1, \ldots, 1, 0)$ results in the minimal squared differences and the second is Schwarz's inequality. Hence,

$$\mathbb{E} Y_{\tau_i}^2 = \text{var}(\tau_i) = \mathbb{E} S_{\tau_i} - 2G(i) + \mathbb{E}\tau_i = \mathbb{E} S_{\tau_i} - G(i) \geq \frac{1}{i} G^2(i) - G(i) \ , \quad (9)$$

results in a lower bound on the variance, that is:

$$cv^2(\tau_i) \geq \frac{1}{i} - \frac{1}{G(i)} \ . \quad (10)$$

<u>*Bound of variance provided by the structure of DPH distributions*</u>

An other lower bound on the variance of τ_i is obtained below by considering the structural properties of the DPH class and Lemma 1.

To simplify the notation we define

$$D(i) = \mathbb{E} S_{\tau_i} = \mathbb{E} \sum_{s=1}^{\tau_i} (G(X_s) - G(X_{s-1}))^2.$$

From Eq. (9) it follows that $\text{var}(\tau_i) = D(i) - G(i)$, (which implies, $D(i) \geq G(i), \forall i \in S$). $D(i)$ satisfies

$$D(i) = \sum_{j \in S} \pi_{ij} \left[D(j) + (G(i) - G(j))^2 \right] \ . \quad (11)$$

where

$$\begin{aligned}\sum_{j\in S}\pi_{ij}(G(i)-G(j))^2 &\geq \left(\sum_{j\in S}\pi_{ij}(G(i)-G(j))\right)^2 \\ &= \left(G(i)-\sum_{j\in S}\pi_{ij}G(j)\right)^2 = 1\ .\end{aligned} \tag{12}$$

Eq. (12) comes from Jensen's inequality and from (2). The equality

$$\sum_{j\in S}\pi_{ij}(G(i)-G(j))^2 = 1$$

holds when $\exists j^*$ such that $G(j^*) = G(i) - 1$ and $\pi_{ij^*} = 1$; i.e., equality can be attained only for $G(i) \geq 2$.

From (2) and (11) we have

$$\begin{aligned} var(\tau_i) &= D(i) - G(i) \\ &= \sum_{j\in S}\pi_{ij}(G(i)-G(j))^2 - 1 + \sum_{j\in S}\pi_{ij}(D(j)-G(j)) \\ &\geq \sum_{j\in S}\pi_{ij}(G(i)-G(j))^2 - 1 \ = \ \sum_{j\in S}\pi_{ij}G(j)^2 - (G(i)-1)^2 \\ &\geq \sum_{j\in S}\pi_{ij}G(j) - (G(i)-1)^2 \ = \ (G(i)-1) - (G(i)-1)^2\ , \end{aligned} \tag{13}$$

since $G(j)^2 \geq G(j); \forall j \in S$. For $1 \leq G(i) < 2$ (i.e., $G(i) - 1 = \langle G(i)\rangle$) Eq. (13) means that

$$D(i) - G(i) \geq \langle G(i)\rangle - \langle G(i)\rangle^2\ . \tag{14}$$

To show that (14) holds also for those states whose mean time to absorption is greater than 2 ($G(i) > 2$) we assume that there exists state i such that $n \leq G(i) < n+1$ and $D(i) - G(i) < \langle G(i)\rangle - \langle G(i)\rangle^2$. Knowing that state insertions do not increase the minimal variance of τ_i according to Lemma 1 we insert new states $i_1^*, i_2^*, \ldots, i_{n-1}^*$ to the Markov chain in the following way:

$$\pi_{i,i_1^*} = 1, \quad \pi_{i_j^*,i_{j+1}^*} = 1, j = 1, 2, \ldots, n-2\ ,$$

and the outgoing transition probabilities from i_{n-1}^* can be anything that fit with Eq. (2) (for the expanded Markov chain). Note that we maintain the numbering of the original states in the expanded Markov chain. This insertion of states results that $G(i_j^*) = G(i) - j$ (i.e., $\langle G(i)\rangle = \langle G(i_j^*)\rangle$) and $D(i_j^*) = D(i) - j$ for $\forall j \in \{1, 2, \ldots, n-1\}$, and hence

$$D(i_{n-1}^*) - G(i_{n-1}^*) < \langle G(i)\rangle - \langle G(i)\rangle^2\ .$$

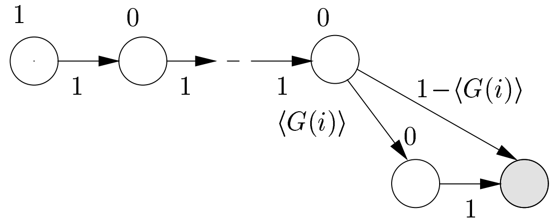
But this is in conflict with Eq. (14) since $1 \leq G(i^*_{n-1}) < 2$. Which means that for any i

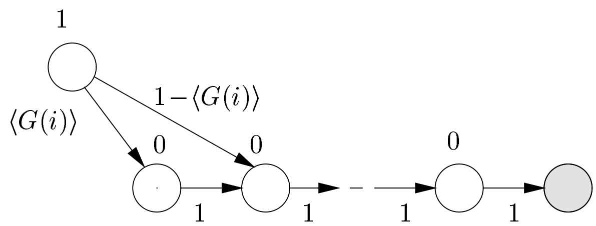
$$\mathrm{var}(\tau_i) = D(i) - G(i) \geq \langle G(i) \rangle - \langle G(i) \rangle^2$$

and

$$cv^2(\tau_i) \geq \frac{\langle G(i) \rangle \, (1 - \langle G(i) \rangle)}{G^2(i)} \; . \tag{15}$$

The two DPH distributions below exhibit the lower bound provided by the DPH structure, and the first one demonstrates the effect of the state insertion described above as well.





Comparing the bounds in (10) and (15) it can be seen that (10) is meaningless (i.e., negative) when $G(i) < i$ and the structural bound (15) dominates. In contrast, for $G(i) > i$ the structural bound is less and (10) dominates. For $G(i) = i$ both bounds equal to 0. □

Lemma 2 has the following consequences:

- The minimal coefficient of variation of τ_i is obtained by a DPH with only downward transitions, i.e., $\pi_{ij} > 0$, iff $i \geq j$. Hence the minimal DPH is acyclic.

- As a result of the previous point the minimal coefficient of variation of τ_i is independent of n (the degree of τ), and it is equivalent with the minimal coefficient of variation that can be obtained by i phases. Hence (10) provides a relation of the degree (i), the mean ($G(i)$) and the minimal coefficient of variation of τ_i.

- The lower bound in (15) is independent of the degree of τ_i, i. This bound comes from the structural properties of the DPH distributions.

Proof of Theorem 1:
On the one hand, from Lemma 2, we have

$$cv^2(\tau) = \frac{\sum_{i\in S} p_i\Big(cv^2(\tau_i)\cdot G^2(i) + G^2(i)\Big) - \mu^2}{\mu^2}$$

$$\geq \frac{\sum_{i\in S} p_i\left(\frac{G^2(i)}{i} - G(i) + G^2(i)\right) - \mu^2}{\mu^2}$$

where the inequality comes from (10). We further have:

$$\begin{aligned} cv^2(\tau) &\geq \frac{\sum_{i\in S} p_i\left(\left(\frac{1}{i}+1\right)G^2(i) - G(i)\right) - \mu^2}{\mu^2} \\ &\geq \frac{\sum_{i\in S}\left(\frac{1}{N}+1\right)p_iG^2(i) - \sum_{i\in S}p_iG(i) - \mu^2}{\mu^2} \\ &\geq \frac{\left(\frac{1}{N}+1\right)\mu^2 - \mu - \mu^2}{\mu^2} = \frac{1}{N} - \frac{1}{\mu} \end{aligned} \tag{16}$$

where $\sum_{i\in S} p_iG^2(i) \geq \mu^2$ by Jensen's inequality.
On the other hand, using (15), we have

$$cv^2(\tau) = \frac{\sum_{i\in S} p_i\Big(cv^2(\tau_i)\cdot G^2(i) + G^2(i)\Big) - \mu^2}{\mu^2}$$

$$\geq \frac{\sum_{i\in S} p_i\Big(\langle G(i)\rangle\ (1 - \langle G(i)\rangle) + G^2(i)\Big) - \mu^2}{\mu^2}$$

Considering the sum in the numerator

$$\sum_{i\in S} p_i\Big(\langle G(i)\rangle\Big(1-\langle G(i)\rangle\Big)+G^2(i)\Big) =$$

$$\sum_{i\in S} p_i\Big(\Big(1-\langle G(i)\rangle\Big)\ \lfloor G(i)\rfloor^2+\langle G(i)\rangle\ \Big(\lfloor G(i)\rfloor+1\Big)^2\Big) =$$

$$\sum_{i\in S}\Big(p_i\Big(1-\langle G(i)\rangle\Big)\ \lfloor G(i)\rfloor^2+p_i\langle G(i)\rangle\ \Big(\lfloor G(i)\rfloor+1\Big)^2\Big)$$

The last expression is the second moment of a random variable with mean μ and support on $\mathbb{N}$. Among the random variables with mean μ and support on $\mathbb{N}$ the one with the minimal second moment is $\bar{X}$, defined as $\Pr(\bar{X}=\lfloor\mu\rfloor)=1-\langle\mu\rangle$ and $\Pr(\bar{X}=\lfloor\mu\rfloor+1)=\langle\mu\rangle$ (i.e., the probability is concentrated around μ as much as possible), which means that:

$$\sum_{i\in S} p_i\Big(1-\langle G(i)\rangle\Big)\ \lfloor G(i)\rfloor^2+p_i\langle G(i)\rangle\ \Big(\lfloor G(i)\rfloor+1\Big)^2$$
$$\geq (1-\langle\mu\rangle)\ \lfloor\mu\rfloor^2+\langle\mu\rangle\ (\lfloor\mu\rfloor+1)^2=\mu^2+\langle\mu\rangle(1-\langle\mu\rangle)$$

from which

$$cv^2(\tau)\geq\frac{\langle\mu\rangle(1-\langle\mu\rangle)}{\mu^2}$$

Since $\frac{\langle\mu\rangle(1-\langle\mu\rangle)}{\mu^2}$ is greater (less) than $\frac{1}{N}-\frac{1}{\mu}$ when N is greater (less) than μ, the theorem is proved. □

References

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

[2] A. Bobbio, A. Horváth, M. Scarpa, and M. Telek, "Acyclic discrete phase type distributions. Part 1: Properties and canonical forms." Submitted for publication, 1999.

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