



Classification and properties of acyclic discrete phase-type distributions based on geometric and shifted geometric distributions

Mohsen Varmazyar¹ · Raha Akhavan-Tabatabaei² · Nasser Salmasi¹ · Mohammad Modarres¹

Received: 30 November 2017 / Accepted: 12 November 2018 / Published online: 27 November 2018
© The Author(s) 2018

Abstract

Acyclic phase-type distributions form a versatile model, serving as approximations to many probability distributions in various circumstances. They exhibit special properties and characteristics that usually make their applications attractive. Compared to acyclic continuous phase-type (ACPH) distributions, acyclic discrete phase-type (ADPH) distributions and their subclasses (ADPH family) have received less attention in the literature. In this paper, we present the definition, properties, characteristics and PH representations of ADPH distributions and their subclasses with finite state space. Based on the definitions of geometric and shifted geometric distributions, we propose a distinct classification for the ADPH subclasses analogous to ACPH family. We develop the PH representation for each ADPH subclass and prove them through their closure properties. The advantage of our proposed classifications is in applying precise representations of each subclass and preventing miscalculation of the probability mass function, by computing the ADPH family based on geometric and shifted geometric distributions.

Keywords Phase-type distribution · Acyclic discrete phase-type distribution (ADPH) · Classification of ADPH · Representations of ADPH · Geometric and shifted geometric distribution

Introduction

Phase-type (PH) distributions, introduced by Neuts (1975, 1981), form a very general class of distributions that have been successfully applied in a wide variety of stochastic disciplines for the last few decades. Acyclic subsets of PH (APH) distributions, continuous and discrete, are the important and interesting subclass of PH distributions with triangular matrix representation. They constitute a versatile modeling tool for as much as firstly, they admit a unique minimal representation, called canonical form (Bobbio et al. 2003). Secondly, the canonical form would simplify the computation of the best approximation for given distribution, by not taking into account redundant parameters. Thirdly, the complexity of overall system model can be controlled by APH minimal representation (Cumani 1982). Special properties

and characteristics of APH distributions usually make the analysis easier and they are highly applicable in mathematical and analytical approaches. Two important applications of APH distributions in stochastic modeling, namely smaller-sized representations, and estimation of the APH distribution parameters are reviewed in the following paragraphs.

The first application, smaller-sized representations, is one of the most interesting theoretical research questions in the field of APH distributions. The size of the matrix representations has a strong effect on the computational efforts which is needed in analyzing this kind of distributions. These representations, however, are not unique, and two representations of the same distribution can differ drastically in size. The systematic study of representations for acyclic continuous phase-type (ACPH) distributions has been initiated by Cumani (1982). In particular, he proves that every ACPH representation has a bidiagonal representation of the same or less order. Aside from the bidiagonal representation, he also provides two other canonical forms and straightforward procedures to transform one to others. In O’Cinneide (1989), the result of Cumani (1982) is extended and restated that every PH representation with a triangular PH generator has a Coxian representation. O’Cinneide (1991, 1993) identifies

✉ Mohsen Varmazyar
varmazyar@ie.sharif.edu; mohsen.varmazyar@gmail.com

¹ Department of Industrial Engineering, Sharif University of Technology, P.O. Box 11365-8639 Tehran, Iran

² School of Management, Sabanci University, Istanbul, Turkey

liminality conditions without presenting algorithmic considerations. Mocanu and Commault (1999) introduce an extension of the triangular PH distributions (monocyclic distributions). They show that any PH distribution can be represented as a mixture of these simple sparse distributions. For two recent decades, most researchers have focused their attention on algorithms to construct the minimal representation of any ACPH distribution. He and Zhang (2008) provide an algorithm for computing minimal representations of APH distribution. This algorithm involves converting the given ACPH distribution to a representation that only contains the poles of the distribution and solving a system of nonlinear equations for each additional state. Pulungan and Hermanns (2008a) develop an algorithm to address the same problem which is considered by He and Zhang (2008). Their algorithm eliminates states from a representation until no further elimination is possible. He et al. (2011) present two new algorithms to find a Coxian representation for any PH representation with only real eigenvalues. Pulungan and Hermanns (2013) develop an algorithm that almost surely (i.e., with probability 1) finds the smallest possible representation of a given ACPH distribution. The algorithm is embedded in a simple, yet expressive calculus of delays, enabling the user to specify complex delay dependencies with the aid of convenient operations. The first exploration of acyclic discrete phase-type (ADPH) distributions is started by Bobbio et al. (2003), and they show that similar to the continuous case (Cumani 1982), the ADPH class admits a unique minimal representation, called canonical form. Bobbio et al. (2004) introduce a new parameter for DPH distributions named scale factor. This new parameter represents the time span associated with each step and can be viewed as a new degree of freedom since its choice largely impacts the shape and properties of a DPH distribution over the continuous time axis. They show that the case when the scale factor is strictly positive results in DPH distributions and if the scale factor is zero, the resulting class is the class of CPH distributions. New results on the canonical representation of DPH with 2 and 3 phases (DPH(2) and DPH(3)) as well as discrete MAP with 2 phases (DMAP(2)) are presented by Meszáros et al. (2014). They provide explicit formulas for parameter matching using these canonical forms, give moments and correlation bounds for these models and show their efficiency in fitting through numerical examples. The canonical representation of DPH distributions with 3 phases is investigated by Horváth et al. (2015). During the course of their investigation, they find that the problem of canonical representation of DPH distributions with 3 phases is far more complex than the one of CPH distribution with 3 phases. As a result, they distinguish 8 different subclasses of DPH distribution with 3 phases, while it is enough to distinguish 3 subclasses of CPH distribution with 3 phases for their canonical representation.

The second application, estimation of the PH distribution parameters, is a critical problem with several numerical limitations in practice. The difficulty of the fitting problem is largely related to the nonlinearity of the model and to the number of the parameters to be estimated (Bobbio and Telek 1994). Considerations of model parsimony have led many authors to constrain many of the PH transition rates to be the same or functionally related such as ACPH subclasses (Slud and Suntornchost 2014). One of the CPH subclass distributions, represented by the so-called Coxian distribution (Cox 1955), can be formally considered as resulting from a series of exponential stages with complex valued transition rates. Fitting a Coxian distribution of order n needs the estimation of $2n$ parameters. Even with the reduced number of parameters required for the Coxian distribution, estimation can still be problematic. This is due to the nonlinear expression and non-unique representations of the distribution which requires optimizing a number of parameters simultaneously (Marshall and Zenga 2012). In order to overcome these problems, various restrictions of the ACPH representation are defined by many authors.

A simple and popular restriction representation of ACPH distributions consists of mixtures of Erlang or hyper-Erlang distributions (HErD). Bux and Herzog (1977) develop a nonlinear estimation approach based on the matching of the first two moments coupled with the minimization of a distance measure with respect to the mixtures of Erlangs. Singh et al. (1977) consider series/parallel combinations of Erlang stages and estimated parameters by matching an equal number of moments by means of a Newton–Raphson numerical method. In some series of papers, Johnson and Taaffe explore the problem of matching the first three moments to a mixture of two Erlangs (Johnson and Taaffe 1989, 1990a; Johnson 1993). Thümmler et al. (2006) develop a new approach by an expectation–maximization (EM) algorithm for mixed-type distributions to compute MLEs of hyper-Erlang distributions (mixed-Erlang distributions). Since their approach focuses only on the hyper-Erlang distributions, the computation speed is improved over other algorithms such as Asmussen et al.’s EM algorithm (Asmussen et al. 1996).

Another popular restriction representation of ACPH distributions is hyper-exponential distribution introduced by Botta and Harris (1986). An ML estimation procedure for hyper-exponential distribution has been described by Harris and Sykes (1984). A new technique for fitting long-tailed data sets is proposed by Riska et al. (2004). This technique fits data sets with non-monotone densities into a mixture of Erlang and hyper-exponential distributions, and data sets with completely monotone densities into hyper-exponential distributions. Their method partitions the data set in a divide-and-conquer fashion and uses the EM algorithm to fit the data of each partition into a hyper-exponential distribution.

Sadre and Haverkort (2008) focus on the EM-based fitting of heavy-tailed distributed data to hyper-exponential distributions. They present a data aggregation algorithm which accelerates the fitting by several orders of magnitude.

The primary attempt to define the subclass of ADPH is given by Bobbio et al. (2003). They propose three canonical forms to introduce the subclasses of ADPH and present the ML estimation algorithm for one of them. Callut and Dupont (2006) mention some example of ADPH such as negative binomial, the mixture of negative binomials and the discrete Coxian distribution. They also present an EM algorithm considered as an adaptation to discrete distributions of the work of Asmussen et al. (1996), which handles CPH distributions.

Table 1 summarizes the majority of studies performed on fitting algorithms and smaller-sized representation approaches related to APH distributions. Based on the reviewed literature, listed in Table 1, ACPH distribution and its subclasses have been extensively studied. Conversely, the ADPH distribution and its subclasses (ADPH family) have received very little attention, and most studies presented in the literature are just concentrated on general ADPH.

Moreover, the ADPH family takes advantage of the canonical form, minimal representation as well as simplification of computation.

The present paper is concentrated on the definition, properties, characteristics, and PH representations of ADPH family with finite state space. In this research, a distinct classification is developed for the subclasses of ADPH distributions based on two different definitions of the geometric distribution. The advantage of ADPH classifications is in applying correct representation of each class and preventing miscalculation of probability mass function (pmf), by computing the ADPH family based on geometric and shifted geometric distributions. For example, Esparza et al. (2010) define pmf of Shifted negative binomial distribution while use PH representation of negative binomial distribution. In addition, all the subclasses of ADPH analogues to ACPH are introduced and the properties, characteristics and PH representations related to each subclass are calculated and proven.

The rest of the paper is organized as follows: Sect. 2 describes the basic definitions, notation, and properties of DPH distributions. Section 3 introduces the definition of

Table 1 A review of APH studies based on the fitting algorithm and smaller-sized representation

| Distribution | Fitting algorithm | Smaller-sized representation | |
|--------------|--|--|--|
| ACPH | Hyper-exponential | Whitt (1982); Harris and Sykes (1984); Botta and Harris (1986); Johnson and Taaffe (1991); Feldmann and Whitt (1997); Khayari et al. (2003); Riska et al. (2004); Dufresne (2007); Singh and Dattatreya (2007); Sadre and Haverkort (2008); Yu et al. (2012); Reinecke et al. (2013) | Assaf et al. (1982); Augustin and Büscher (1982); Dehon and Latouche (1982); David and Larry (1987); Maier (1991); Harris et al. (1992); Maier and O’Cinneide (1992); Commault and Chemla (1993); O’Cinneide (1993); Commault and Chemla (1996); Chauveau et al. (1996); Mocanu and Commault (1999); O’Cinneide (1999); Commault et al. (2002); Commault (2003); Commault and Mocanu (2003); M. W. Fackrell (2003); Bobbio et al. (2004); He and Zhang (2005), (2006a, b); Horváth and Telek (2007b); He and Zhang (2007); Telek and Horváth (2007); Bodrog et al. (2008); Éltető and Vadera (2008); Pulungan and Hermanns (2008b); Horváth and Telek (2009); Fackrell et al. (2010); Pulungan and Hermanns (2013); Jain and Bhagat (2014); Horváth and Telek (2015) |
| | Hyper-erlang | Sauer and Chandy (1975); Bux and Herzog (1977); Singh et al. (1977); Johnson and Taaffe (1989), (1990a, b); Schmickler (1992); Johnson (1993); Malhotra and Reibman (1993); Wang et al. (2005), (2006); Thümmeler et al. (2006); Panchenko and Thümmeler (2007); Wang et al. (2008); Lee and Lin (2010); Kim and Thomas (2011); Horváth (2013); Hu et al. (2013); Gong (2014) | |
| | General ACPH & Coxian | Marie (1980); Parr and Schucany (1980); Altioik (1985); Van Der Heijden (1988); de Liefvoort (1990); Bobbio and Cumani (1992); Faddy (1993); Bobbio and Telek (1994); Faddy (1994), (1998); Faddy and McClean (1999); Horvath and Telek (2000); Vanden Bosch et al. (2000); Faddy (2002); A. Horváth and Telek (2002); Telek and Heindl (2002); Osogami and Harchol-Balter (2003a, b); Pérez-Ocón and Ruiz-Castro (2003); Bobbio et al. (2005); Osogami and Harchol-Balter (2006); Horváth and Telek (2007a, b); Buchholz and Kriege (2009); Buchholz et al. (2010); Marshall and Zenga (2012) | |
| ADPH | General ADPH | Horváth and Telek (2002); Telek and Heindl (2002); Isensee and Horton (2005); Mészáros et al. (2014); Akar (2015) | Telek (2000); Bobbio et al. (2004); Dayar (2005); Mészáros and Telek (2013); Papp and Telek (2013); Mészáros et al. (2014); Horváth et al. (2015) |
| | Mixtures of binomial/negative binomial/geometric distributions | Adan et al. (1995); Bobbio et al. (2003) | |

ADPH distribution and presents two different representations of ADPH family based on two different definitions of the geometric distribution. The subclasses of ADPH distribution compared with the subclasses of ACPH distribution and some properties are proven. Finally, concluding remarks are given in Sect. 4.

Discrete phase-type distribution and their properties

The following subsections summarize the definition and main properties of DPH family of distributions.

Definition and notation

DPH distributions have been introduced and formalized by Neuts (1981) which are defined as the distribution of time until absorption in a discrete-state discrete-time Markov chain (DTMC) with n transient states, and one absorbing state. More precisely, assume that $\{X(n)\}_{n \geq 0}$ denote the DTMC with finite state space $S = \{0, 1, 2, \dots, n\}$, where the absorbing state is numbered 0 and the transient states are numbered $1, 2, \dots, n$. DPH distribution is defined by $Z = \inf\{i \in \mathbb{N} : X_i = 0\}$ with representation $(\boldsymbol{\pi}, \mathbf{T})$, and is shown by $Z \sim PH_d(\boldsymbol{\pi}, \mathbf{T})$. The one-step transition probability matrix of the corresponding DTMC can be partitioned as

$$\mathbf{P} = \begin{bmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}, \tag{1}$$

where \mathbf{T} is a square matrix of dimension n , \mathbf{t} is a column vector and $\mathbf{0}$ is a row vector of dimension n . Since \mathbf{P} is a transition probability matrix, we have that $T_{ij} \geq 0$ and $t_i \geq 0 \forall i, j \in S$ and $\mathbf{T}\mathbf{1} + \mathbf{t} = \mathbf{1}$ where $\mathbf{1}$ is the column vector ones of the appropriate dimension n . The initial probability for transient and absorbing states is denoted with the row vector $(\boldsymbol{\pi}, \pi_0)$ and $\pi_0 = 1 - \boldsymbol{\pi}\mathbf{1}$.

The cumulative distribution function of the DPH distribution $Z \sim PH_d(\boldsymbol{\pi}, \mathbf{T})$ is calculated by

$$F_Z(x) = P(Z \leq x) = \mathbf{1} - \boldsymbol{\pi}\mathbf{T}^x\mathbf{1} \quad \text{for } x = 0, 1, 2, \dots \tag{2}$$

the probability mass function is

$$\begin{aligned} P_Z(x) &= \Pr(Z = x) = \boldsymbol{\pi}\mathbf{T}^{x-1}\mathbf{t} \quad \text{for } x = 1, 2, \dots \\ P_Z(0) &= \Pr(Z = 0) = \pi_0 \end{aligned} \tag{3}$$

and the factorial moment is

$$\begin{aligned} f_k &= E[X(X - 1) \dots (X - k + 1)] \\ &= k!\boldsymbol{\pi}(\mathbf{I} - \mathbf{T})^{-k}\mathbf{T}^{k-1}\mathbf{1} \quad \text{for } k = 1, 2, \dots \end{aligned} \tag{4}$$

Closure properties

One of the appealing features of PH distributions is that the class is closed under a number of operations. The closure properties are a main contributing factor to the popularity of these distributions in stochastic modeling. The DPH distributions inherit many properties from the CPH distributions (Maier 1991), and both of them are closed under addition, finite mixtures, and finite order statistics (Esparza et al. 2010). However, one of the most interesting properties of the DPH distributions is that they can represent in an exact way a number of distributions with finite support.

Assume that $Z_i \sim PH_d(\boldsymbol{\pi}^{(i)}, \mathbf{T}^{(i)})$ for $i = 1, 2$ are two independent DPH distributed random variables of order n_i .

- (1) Convolution of PH_d : the sum $Z = Z_1 + Z_2 \sim PH_d(\boldsymbol{\pi}, \mathbf{T})$ has a DPH distribution of order $n = n_1 + n_2$ with representation

$$\boldsymbol{\pi} = \left(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}_0^{(1)} \boldsymbol{\pi}^{(2)} \right) \quad \text{and} \quad \mathbf{T} = \begin{pmatrix} \mathbf{T}^{(1)} & \mathbf{t}^{(1)}\boldsymbol{\pi}^{(2)} \\ \mathbf{0} & \mathbf{T}^{(2)} \end{pmatrix} \tag{5}$$

Proof See Latouche and Ramaswami (1999), Theorem 2.6.1.

- (2) Mixture of PH_d : the convex mixture sum $Z = \alpha Z_1 + (1 - \alpha)Z_2 \sim PH_d(\boldsymbol{\pi}, \mathbf{T})$ has a DPH distribution of order $n = n_1 + n_2$ with representation

$$\boldsymbol{\pi} = (\alpha\boldsymbol{\pi}^{(1)}, (1 - \alpha)\boldsymbol{\pi}^{(2)}) \quad \text{and} \quad \mathbf{T} = \begin{pmatrix} \mathbf{T}^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{(2)} \end{pmatrix} \tag{6}$$

Proof See Latouche and Ramaswami (1999), Theorem 2.6.2.

- (3) Minimum of PH_d : The minimum $Z = \min(Z_1, Z_2) \sim PH_d(\boldsymbol{\pi}, \mathbf{T})$ has a DPH distribution of order $n = n_1 \cdot n_2$ with representation

$$\boldsymbol{\pi} = \boldsymbol{\pi}^{(1)} \otimes \boldsymbol{\pi}^{(2)} \quad \text{and} \quad \mathbf{T} = \mathbf{T}^{(1)} \otimes \mathbf{T}^{(2)} \tag{7}$$

where \otimes is the Kronecker product.

Proof See Latouche and Ramaswami (1999), Theorem 2.6.4.

- (4) Maximum of PH_d : The maximum $Z = \max(Z_1, Z_2) \sim PH_d(\boldsymbol{\pi}, \mathbf{T})$ has a DPH distribution of order $n = n_1 \cdot n_2 + n_1 + n_2 + 1$ with representation

$$\begin{aligned} \boldsymbol{\pi} &= \left(\boldsymbol{\pi}^{(1)} \otimes \boldsymbol{\pi}^{(2)}, \boldsymbol{\pi}^{(1)}\boldsymbol{\pi}_0^{(2)}, \boldsymbol{\pi}_0^{(1)}\boldsymbol{\pi}^{(2)}, 0 \right) \quad \text{and} \\ \mathbf{T} &= \begin{pmatrix} \mathbf{T}^{(1)} \otimes \mathbf{T}^{(2)} & \mathbf{T}^{(1)} \otimes \mathbf{t}^{(2)} & \mathbf{t}^{(1)} \otimes \mathbf{T}^{(2)} & \mathbf{t}^{(1)} \otimes \mathbf{t}^{(2)} \\ \mathbf{0} & \mathbf{T}^{(1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}^{(2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \end{aligned} \tag{8}$$

Proof See Alfa (2016), p. 40.

- (5) Shift of PH_d : The shifted $Z = \max(Z_1 - r, 0) \sim PH_d(\boldsymbol{\pi}, \mathbf{T})$ where $r \in \mathbb{N}$ has a DPH distribution of order $n = n_1$ with representation

$$\boldsymbol{\pi} = \boldsymbol{\pi}^{(1)}(\mathbf{T}^{(1)})^r \text{ and } \mathbf{T} = \mathbf{T}^{(1)} \tag{9}$$

Proof See Neuts (1981), p.47.

- (6) Deterministic time: The constant number $Z = r \sim PH_d(\boldsymbol{\pi}, \mathbf{T})$ where $r \in \mathbb{N}$ has a DPH distribution of order $n = r$ with representation

$$\boldsymbol{\pi} = \overbrace{(1, 0, \dots, 0)}^r \text{ and } \mathbf{T} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \tag{10}$$

Proof See Neuts (1981), p. 47.

Acyclic discrete phase-type distributions and their subclasses

The DPH is defined as an acyclic DPH (ADPH) if its states can be ordered in such a way that matrix \mathbf{T} is an upper triangular matrix (Bobbio et al. 2003). Based on this definition, matrix representation $(\boldsymbol{\pi}, \mathbf{T})$ has $\frac{n^2+n}{2}$ parameters for the upper triangular matrix (\mathbf{T}) and $n - 1$ free parameters for the initial probability vector $\boldsymbol{\pi}$. Same as ACPH, ADPH distributions can be divided into various subclasses depending on the structure of \mathbf{T} and $\boldsymbol{\pi}$ which are shown in Table 2. The continuous analogous of these ADPH distributions is also illustrated in this table. The simplest DPH distribution is the geometric distribution that is defined by two ways, the geometric distribution and shifted geometric distribution (Kroese et al. 2013). In the following, we will give an overview of ADPH distributions based on geometric distribution and shifted geometric distribution and show some properties and characteristics related to them.

Table 2 Subclasses of ADPH and ACPH distribution

| ADPH | ACPH |
|--|--------------------------------|
| Geometric distribution | Exponential distribution |
| Negative binomial distribution | Erlang distribution |
| Generalized negative binomial distribution | Hypo-exponential distribution |
| Mixed geometric distribution | Hyper-exponential distribution |
| Mixed negative binomial distribution | Hyper-Erlang distribution |
| Discrete Coxian distribution | Coxian distribution |

Subclasses of ADPH distributions based on geometric distribution

The DPH distributions are created by a system of one or more inter-related geometric distributions occurring in sequence or phases. The geometric distribution ($X \sim G(p)$, with $p \in (0, 1)$) is used to describe the time of first success in an infinite sequence of independent Bernoulli trials with success probability p . Then, X is the number of Bernoulli trials needed to get first success and its probability mass function is $\Pr(X = x) = (1 - p)^{x-1}p$, for $x = 1, 2, \dots$. The DPH representation of geometric distribution (Kroese et al. 2013) is given by Eq. (11) and shown in Fig. 1. In all figures related to DPH representation, the absorbing state is numbered by 0 and the transient states are numbered by $1, 2, \dots, n$.

$$\boldsymbol{\pi}_G = [1], \quad \mathbf{T}_G = [1 - p], \quad \mathbf{t}_G = [p] \tag{11}$$

The mean and variance of geometric distribution are $E[X] = \frac{1}{p}$ and $Var[X] = \frac{1-p}{p^2}$, respectively.

Negative binomial distribution ($X \sim NB(n, p)$) is defined as a number of Bernoulli trials needed before the n th success and introduced as the sum of n independent random variables $G(p)$ -distributed, so $\Pr(X = x) = \binom{x-1}{n-1} (1-p)^{x-n} p^n$, for $x = n, n + 1, \dots$. Based on the definition of negative binomial distribution and using Eq. (5), the DPH representation of negative binomial distribution is given by Eq. (12) and illustrated in Fig. 2.

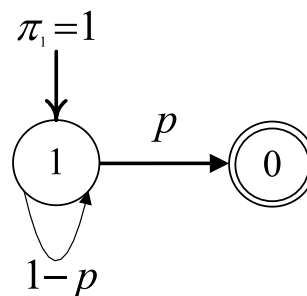


Fig. 1 The DPH representation of $G(p)$

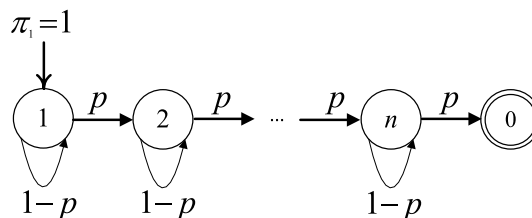


Fig. 2 The DPH representation of $NB(n,p)$

$$\pi_{NB} = (1, 0, \dots, 0), \quad \mathbf{T}_{NB} = \begin{pmatrix} 1-p & p & 0 & 0 & 0 & 0 \\ 0 & 1-p & p & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p & p \\ 0 & 0 & 0 & 0 & 0 & 1-p \end{pmatrix},$$

$$\mathbf{t}_{NB} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ p \end{pmatrix} \tag{12}$$

The mean and variance of negative binomial distribution are $E[X] = \frac{n}{p}$ and $Var[X] = \frac{n(1-p)}{p^2}$, respectively.

The generalized negative binomial distribution ($X \sim GNB(n, p_i)$) is considered as the next subclass of ADPH which is the general case of negative binomial distribution. Consider a set of different geometric distributions where their success probabilities p_1, p_2, \dots, p_n are not necessarily identical. The *GNB* distribution is introduced as the sum of n independent random variables of geometric distributions with distinct parameters. The probability mass function is given by Eq. (13).

$$\Pr(X = x) = \sum_{i=1}^n \left(\prod_{\substack{j=1 \\ i \neq j}}^n \frac{p_j}{p_j - p_i} \right) (1 - p_i)^{x-1} p_i, \tag{13}$$

for $x = n, n + 1, \dots \quad p_i \neq p_j$

Derivation of GNB pmf

Let $X_{G_1}, X_{G_2}, \dots, X_{G_n}$ be independent geometric random variables where their probability mass function is $\Pr(X = x) = (1 - p)^{x-1}p$, for $x = 1, 2, \dots$. We assume that the probability mass function of $S_n = X_{GNB} = \sum_{i=1}^n X_{G_i}$ is calculated by Eq. (13). The proof of this equation follows by induction on n based on Sen and Balakrishnan (1999). Equation (13) is trivially true for $n = 1$, where by definition, $\prod_{\substack{j=1 \\ i \neq j}}^n \frac{p_j}{p_j - p_i} \equiv 1$. Now suppose that the same equality holds

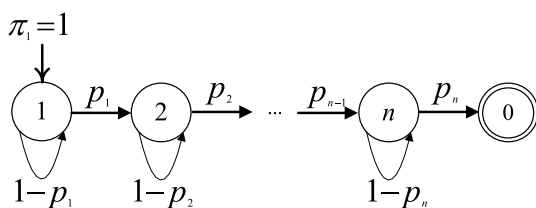


Fig. 3 The DPH representation of $GNB(n, p_i)$

when $n = m$, we shall show that the equation holds for $n = m + 1$. Noting that $S_{m+1} = S_m + X_{G_{m+1}}$, we have

$$\begin{aligned} \Pr(S_{m+1} = k) &= \sum_{x=m}^{k-1} \Pr(S_m = x) * \Pr(X_{G_{m+1}} = k - x) \\ &= \sum_{x=m}^{k-1} p_{m+1} (1 - p_{m+1})^{k-x-1} \sum_{i=1}^m \left(\prod_{\substack{j=1 \\ i \neq j}}^m \frac{p_j}{p_j - p_i} \right) (1 - p_i)^{x-1} p_i \\ &= p_{m+1} (1 - p_{m+1})^{k-1} \sum_{i=1}^m \frac{p_i}{(1 - p_i)} \left(\prod_{\substack{j=1 \\ i \neq j}}^m \frac{p_j}{p_j - p_i} \right) \sum_{x=m}^{k-1} \left(\frac{1 - p_i}{1 - p_{m+1}} \right)^x \end{aligned}$$

Using the geometric sum formula $\sum_{x=n_1}^{n_2} a^x = \frac{a^{n_2+1} - a^{n_1}}{a - 1}$, $a \neq 1$, and some simplifications, the above equation reduces to the following:

$$\begin{aligned} &= \sum_{i=1}^m \left(\prod_{\substack{j=1 \\ i \neq j}}^{m+1} \frac{p_j}{p_j - p_i} \right) p_i (1 - p_i)^{k-1} \\ &\quad - (1 - p_{m+1})^{k-m} \sum_{i=1}^m \left(\prod_{\substack{j=1 \\ i \neq j}}^{m+1} \frac{p_j}{p_j - p_i} \right) p_i (1 - p_i)^{m-1} \\ &= \sum_{i=1}^{m+1} \left(\prod_{\substack{j=1 \\ i \neq j}}^{m+1} \frac{p_j}{p_j - p_i} \right) p_i (1 - p_i)^{k-1} \\ &\quad - (1 - p_{m+1})^{k-m} \sum_{i=1}^{m+1} \left(\prod_{\substack{j=1 \\ i \neq j}}^{m+1} \frac{p_j}{p_j - p_i} \right) p_i (1 - p_i)^{m-1} \end{aligned}$$

By adding and subtracting the $(m+1)^{st}$ term of the first sum to the entire expression, we get

Due to the finite sum of Lagrange polynomials $Q(x) = \sum_{i=1}^n Q(p_i) \prod_{\substack{j=1 \\ j \neq i}}^n \frac{p_j - x}{p_j - p_i}$ (Yang et al. 2005), the second

term on the right-hand side of above equation is equal to zero because $(1 - p_{m+1})^{k-m} Q(0) \equiv 0$ where $Q(x) = x(1 - x)^{m-1}$. Therefore, $\Pr(S_{m+1} = k) = \sum_{i=1}^{m+1} \left(\prod_{\substack{j=1 \\ i \neq j}}^{m+1} \frac{p_j}{p_j - p_i} \right) p_i (1 - p_i)^{k-1}$

and the proof is completed. □

By Eq. (5), the convolution of different geometric distributions can be represented as a DPH distribution with Eq. (14) and the graphical representation is demonstrated in Fig. 3.

$$\begin{aligned} \boldsymbol{\pi}_{GNB} &= (1, 0, \dots, 0), \\ \mathbf{T}_{GNB} &= \begin{pmatrix} 1-p_1 & p_1 & 0 & \dots & 0 & 0 \\ 0 & 1-p_2 & p_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1-p_{n-1} & p_{n-1} \\ 0 & 0 & 0 & \dots & 0 & 1-p_n \end{pmatrix}, \\ \mathbf{t}_{GNB} &= \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ p_n \end{pmatrix} \end{aligned} \tag{14}$$

The mean of the generalized negative binomial distributed random variable is calculated as $E[X] = \sum_{i=1}^n \frac{1}{p_i}$ and the variance as $Var[X] = \sum_{i=1}^n \frac{1-p_i}{p_i^2}$. For instance, we assume that

$X \sim GNB(3, p_1 = 0.2, p_2 = 0.4, p_3 = 0.6)$, the pmf of X is $Pr(X = x) = \sum_{i=1}^3 (\prod_{j \neq i} \frac{p_j}{p_j - p_i})(1 - p_i)^{x-n_i} p_i$, for $x = 3, 4, \dots$

$p_i \neq p_j$. The DPH and diagrammatic representation of X are shown in Fig. 4. The mean and variance of X are $E[X] = 9.1667$ and $Var[X] = 24.8611$, respectively.

The mixed geometric distribution ($X \sim MG(n, p_i, \pi_i)$) is a convex mixture of n geometric distributions. The probability mass function is $Pr(X = x) = \sum_{i=1}^n \pi_i (1 - p_i)^{x-1} p_i$, for $x = 1, 2, \dots$ where $\pi_i > 0$ for all phases i and $\sum_{i=1}^n \pi_i = 1$. The DPH representation of the mixed geometric distribution which is calculated by Eq. (6) is given by Eq. (15). Diagrammatic representation of mixed geometric distribution is presented in Fig. 5.

$$\begin{aligned} \boldsymbol{\pi}_{MG} &= (\boldsymbol{\pi}_{MG_1}, \boldsymbol{\pi}_{MG_2}, \dots, \boldsymbol{\pi}_{MG_n}) = (\pi_1, \pi_2, \dots, \pi_n), \\ \mathbf{T}_{MG} &= \begin{pmatrix} 1-p_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1-p_2 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1-p_{n-1} & 0 \\ 0 & 0 & 0 & \dots & 0 & 1-p_n \end{pmatrix}, \\ \mathbf{t}_{MG} &= \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n-1} \\ p_n \end{pmatrix} \end{aligned} \tag{15}$$

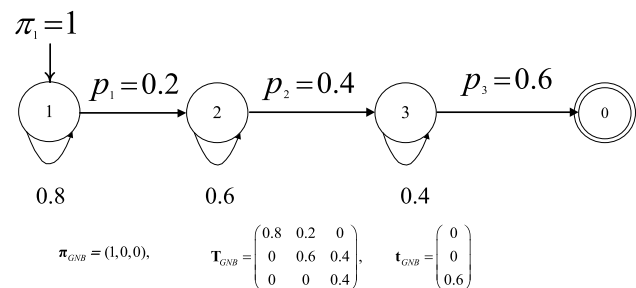


Fig. 4 The DPH representation of $GNB(3, p_1 = 0.2, p_2 = 0.4, p_3 = 0.6)$

The k th factorial moment can be obtained as

$$f_k = E[X(X - 1) \dots (X - k + 1)] = \sum_{i=1}^n \pi_i k! \frac{(1 - p_i)^{k-1}}{p_i^k} \tag{16}$$

for $k = 1, 2, \dots$

Thus, the first moment is obtained by $E[X] = \sum_{i=1}^n \frac{\pi_i}{p_i}$ and its variance is given by $Var[X] = \sum_{i=1}^n \pi_i \frac{(2-p_i)}{p_i^2} - (\sum_{i=1}^n \frac{\pi_i}{p_i})^2$.

A mixed negative binomial distribution ($X \sim MNB(m, n_i, p_i, \pi_i)$) is considered as a mixture of m mutually independent negative binomial distribution weighted with the initial probabilities $\pi_1, \pi_2, \dots, \pi_n$, where $\pi_i \geq 0$ and the vector $\boldsymbol{\pi}$ is stochastic, i.e., $\sum_{i=1}^n \pi_i = 1$. Let n_i denote the number of phases of the i th negative binomial distribution.

Then the probability mass function is $Pr(X = x) = \sum_{i=1}^m \pi_i \binom{x-1}{n_i-1} (1 - p_i)^{x-n_i} p_i^{n_i}$, for $x = \min_{j \in \{1, \dots, m\}} \{n_j\}, \min_{j \in \{1, \dots, m\}} \{n_j\} + 1, \dots$. The state space includes of $\sum_{i=1}^m n_i$ transient and one absorbing state. For $m = 1$, a single negative binomial distribution is formed and the case that $n_i = 1$ for all $1 \leq i \leq m$ represent a mixed geometric distribution. In order to calculate the DPH representation, the Eqs. (6) and (14) are applied that can be described by

$$\begin{aligned} \boldsymbol{\pi}_{MNB} &= (\boldsymbol{\pi}_{MNB_1}, \boldsymbol{\pi}_{MNB_2}, \dots, \boldsymbol{\pi}_{MNB_m}) \\ &= (\overbrace{\pi_1, 0, \dots, 0}^{n_1}, \overbrace{\pi_2, 0, \dots, 0}^{n_2}, \dots, \overbrace{\pi_m, 0, \dots, 0}^{n_m}), \\ \mathbf{T}_{MNB} &= \begin{pmatrix} \mathbf{T}_1 & 0 & \dots & 0 \\ 0 & \mathbf{T}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{T}_m \end{pmatrix}, \end{aligned} \tag{17}$$

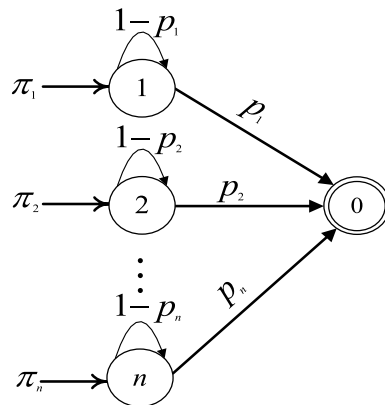


Fig. 5 The DPH representation of $MG(n, p_i, \pi_i)$

where \mathbf{T}_i is calculated based on Eq. (14). Diagrammatic representation of mixed geometric distribution is shown in Fig. 6.

Mixtures of general negative binomial and mixed geometric distribution are considered as discrete Coxian distributions ($X \sim DCo(n_i, p_i, g_i, \pi_i)$). The initial probability vector is given by $\boldsymbol{\pi} = (1, 0, \dots, 0)$. It means that the process starts from phase one and then traverses through the n successive phases with different success probabilities p_i . From phase i transition into the next phase $i + 1^{st}$ can occur with probability g_i or the absorbing state is reached with the complementary probability $1 - g_i$. The DPH representation of the discrete Coxian distribution is given by Eq. (18) and illustrated in Fig. 7.

$$\boldsymbol{\pi}_{DC} = (1, 0, \dots, 0),$$

$$\mathbf{T}_{DC} = \begin{pmatrix} 1-p_1 & g_1 p_1 & 0 & 0 & 0 & 0 \\ 0 & 1-p_2 & g_2 p_2 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1-p_{n-1} & g_{n-1} p_{n-1} \\ 0 & 0 & 0 & 0 & 0 & 1-p_n \end{pmatrix}, \quad (18)$$

$$\mathbf{t}_{DC} = \begin{pmatrix} (1-g_1)p_1 \\ (1-g_2)p_2 \\ \vdots \\ (1-g_{n-1})p_{n-1} \\ p_n \end{pmatrix}$$

Subclasses of ADPH distributions based on shifted geometric distribution

Shifted geometric distribution ($Y \sim SG(p)$, with $p \in (0, 1)$) is another, nonequivalent, definition of the geometric distribution ($X \sim G(p)$) which describes the number of failures before the first success in an infinite sequence of independent Bernoulli trials. The shifted geometric distribution is completely characterized by its success probability p and the probability mass function is $\Pr(Y = y) = (1 - p)^y p$, for $y = 0, 1, 2, \dots$. The DPH representation of shifted geometric distribution is given by Eq. (19) and presented in Fig. 8.

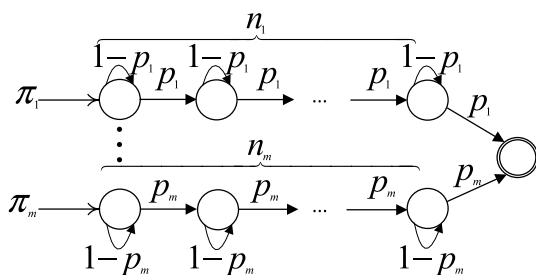


Fig. 6 The DPH representation of $MNB(m, n_i, p_i, \pi_i)$

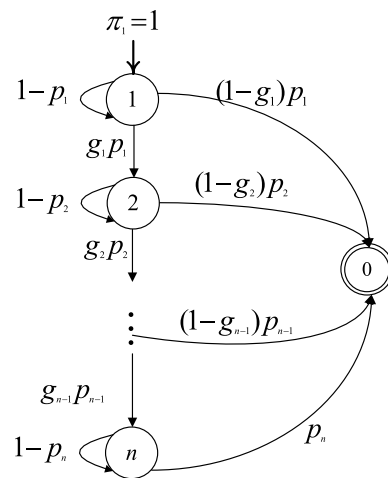


Fig. 7 The DPH representation of $DCo(n_i, p_i, g_i, \pi_i)$

$$\boldsymbol{\pi}_{SG} = [1 - p], \quad \mathbf{T}_{SG} = [1 - p], \quad \mathbf{t}_{SG} = [p] \quad (19)$$

Derivation of SG representation

Based on the definition of geometric and shifted geometric distribution, there is $Y = X - 1$. It means that the geometric distribution is shifted by one unit. Therefore, by using Eq. (9), we can calculate the parameters of shifted geometric distribution as following

$$\mathbf{T}_{SG} = \mathbf{T}_G = [1 - p],$$

$$\boldsymbol{\pi}_{SG} = \boldsymbol{\pi}_G \mathbf{T}_{SG} = [1 - p].$$

□

The mean and variance of shifted geometric distribution are $E[Y] = \frac{1-p}{p}$ and $\text{Var}[Y] = \frac{1-p}{p^2}$, respectively.

Shifted negative binomial distribution ($Y \sim SNB(n, p)$) is described as the number of failures before the n th success in a Bernoulli process and defined as the sum of n independent random variables $SG(p)$ -distributed, so $\Pr(Y = y) = \binom{y+n-1}{n-1} (1-p)^y p^n$, for $y = 0, 1, \dots$. The

DPH and diagrammatic representation of shifted negative

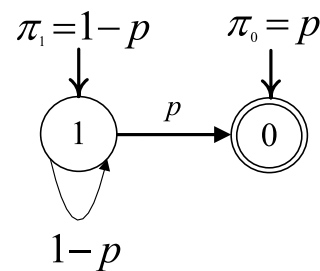


Fig. 8 The DPH representation of $SG(p)$

binomial distribution are presented in Eq. (20) and Fig. 9, respectively.

$$\begin{aligned} \boldsymbol{\pi}_{\text{SNB}} &= (\pi_1, \pi_2, \dots, \pi_n), \pi_j = \binom{n}{j-1} (1-p)^{n-(j-1)} p^{j-1}, \\ \mathbf{T}_{\text{SNB}} &= \begin{pmatrix} 1-p & p & 0 & 0 & 0 & 0 \\ 0 & 1-p & p & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p & p \\ 0 & 0 & 0 & 0 & 0 & 1-p \end{pmatrix}, \\ \mathbf{t}_{\text{SNB}} &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ p \end{pmatrix} \end{aligned} \tag{20}$$

$$\begin{aligned} \boldsymbol{\pi}_{\text{SNB}} &= (\pi_1, \pi_2, \dots, \pi_n) = (1, 0, \dots, 0) \mathbf{T}^n \text{ and} \\ \pi_j &= \binom{n}{j-1} (1-p)^{n-(j-1)} p^{j-1} \text{ for } j = 1, \dots, n \end{aligned} \tag{21}$$

In order to reach Eq. (21), we must determine the value of \mathbf{T}^n and prove following Equation.

$$(\mathbf{T}_{\text{SNB}}^n)_{ij} = \begin{cases} \binom{n}{j-i} (1-p)^{n-(j-i)} p^{j-i} & i \leq j, j-i \leq n \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

where $(\mathbf{T}^n)_{ij}$ is the entry in the i th row and the j th column of a matrix \mathbf{T}^n . To prove Eq. (22), induction on n is applied. Equation (22) is clearly true for $n = 1$. Now suppose that the same equality holds when $n = m$, we shall show that the equation is hold for $n = m + 1$.

$$\begin{aligned} (\mathbf{T}_{\text{SNB}}^{m+1})_{ij} &= \sum_{k=1}^m (\mathbf{T}_{\text{SNB}}^m)_{ik} (\mathbf{T}_{\text{SNB}}^1)_{kj} \quad i \leq k \leq j, k-i \leq m, j-k \leq 1 \\ &= \sum_{k=1}^m \binom{m}{k-i} (1-p)^{m-(k-i)} p^{k-i} \binom{1}{j-k} (1-p)^{1-(j-k)} p^{j-k} \\ &= (1-p)^{m+1-(j-i)} p^{j-i} \sum_{k=1}^m \binom{m}{k-i} \binom{1}{j-k} \quad i \leq k \leq j \leq m, k-i \leq m, j-k \leq 1 \\ &= (1-p)^{m+1-(j-i)} p^{j-i} \sum_{k=i}^j \binom{m}{k-i} \binom{1}{j-k} \stackrel{l=k-i}{=} (1-p)^{m+1-(j-i)} p^{j-i} \sum_{l=0}^{j-i} \binom{m}{l} \binom{1}{j-i-l} \end{aligned}$$

Derivation of SNB representation

By definition of negative binomial and shifted negative binomial distribution, the relation between $X \sim \text{NB}(n, p)$ and $Y \sim \text{SNB}(n, p)$ is $Y = X - n$. This implies that negative binomial is shifted by n unite to constitute the shifted negative binomial. Therefore, by using Eq. (9), the matrix \mathbf{T}_{SNB} is equal to the matrix \mathbf{T}_{NB} in Eq. (12) and the vector $\boldsymbol{\pi}_{\text{SNB}}$ is calculated by Eq. (21).

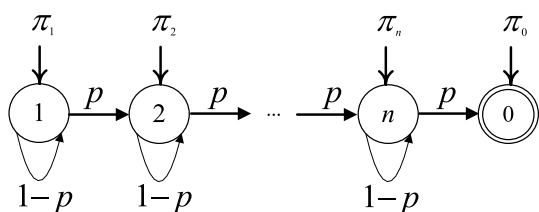


Fig. 9 The DPH representation of $\text{SNB}(n, p)$

Due to the $\binom{a+b}{r} = \sum_{i=0}^r \binom{a}{i} \binom{b}{r-i}$ (Ross 2014), the equation is written as following,

$$\begin{aligned} (\mathbf{T}_{\text{SNB}}^{m+1})_{ij} &= (1-p)^{m+1-(j-i)} p^{j-i} \sum_{l=0}^{j-i} \binom{m}{l} \binom{1}{j-i-l} \\ &= \binom{m+1}{j-i} (1-p)^{m+1-(j-i)} p^{j-i} \end{aligned}$$

So Eq. (22) is true for $n = m + 1$. As a result, $\boldsymbol{\pi}_{\text{SNB}} = \boldsymbol{\pi}_{\text{NB}} \mathbf{T}_{\text{SNB}}^n = (1, 0, \dots, 0) \mathbf{T}_{\text{SNB}}^n$ is the first row of matrix \mathbf{T}^n . It means that $i = 1$ and when the $i = 1$ substitutes in Eqs. (22), the Eqs. (21) is proved. \square

Conceptual interpretation of the initial probability vector (π_j) is the discrete probability distribution of the number of failures before the $(n-j+1)$ st success. In other words, the j th initial probability (π_j) states the probability of the j failures before the $(n-j+1)$ th success.

The factorial moment of the shifted negative binomial distribution is given by Eqs. (23),

$$f_k = E[Y(Y - 1) \dots (Y - k + 1)] = \frac{\Gamma(n + k) (1 - p)^k}{\Gamma(n) p^k} \text{ for } k = 1, 2, \dots \tag{23}$$

where $\Gamma(\cdot)$ is the gamma function defined by:

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx, \quad t > 0 \tag{24}$$

The mean and variance of shifted negative binomial distribution are $E[Y] = \frac{n(1-p)}{p}$ and $Var[Y] = \frac{n(1-p)}{p^2}$, respectively.

The generalized shifted negative binomial distribution ($Y \sim GSNB(n, p_i)$) is considered as a general case of shifted negative binomial distribution. Its probability mass function is given by Eq. (25).

$$Pr(Y = y) = \sum_{i=1}^n \left(\prod_{\substack{j=1 \\ i \neq j}}^n \frac{p_j}{p_j - p_i} \right) (1 - p_i)^{y+n-1} p_i, \tag{25}$$

for $y = 0, 1, \dots$ $p_i \neq p_j$

Derivation of GSNB pmf

Let $Y_{SG_1}, Y_{SG_2}, \dots, Y_{SG_n}$ be independent shifted geometric random variables where their probability mass function is $Pr(Y = y) = (1 - p)^y p$, for $y = 0, 1, 2, \dots$. The probability mass function (pmf) of generalized shifted negative binomial distribution is calculated by $S_n = Y_{GSNB} = \sum_{i=1}^n Y_{SG_i}$ and Eq. (25). Same as the proof of pmf for generalized negative binomial distribution, the proof of Eq. (25) follows by induction on n . Equation (25) is clearly true for $n = 1$, where by definition, $\prod_{\substack{j=1 \\ i \neq j}}^n \frac{p_j}{p_j - p_i} \equiv 1$. We assume Eq. (25) holds for

$n = m$ and proceed to establish it for $n = m + 1$. Noting that $S_{m+1} = S_m + Y_{SG_{m+1}}$, we have

$$\begin{aligned} Pr(S_{m+1} = k) &= \sum_{y=0}^k Pr(S_m = y) * Pr(Y_{SG_{m+1}} = k - y) \\ &= \sum_{y=0}^k p_{m+1} (1 - p_{m+1})^{k-y} \sum_{i=1}^m \left(\prod_{\substack{j=1 \\ i \neq j}}^m \frac{p_j}{p_j - p_i} \right) (1 - p_i)^{y+m-1} p_i \\ &= p_{m+1} (1 - p_{m+1})^k \sum_{i=1}^m p_i (1 - p_i)^{m-1} \left(\prod_{\substack{j=1 \\ i \neq j}}^m \frac{p_j}{p_j - p_i} \right) \sum_{y=0}^k \left(\frac{1 - p_i}{1 - p_{m+1}} \right)^y \end{aligned}$$

Using the geometric sum formula, and some simplifications, the above equation reduces to the following:

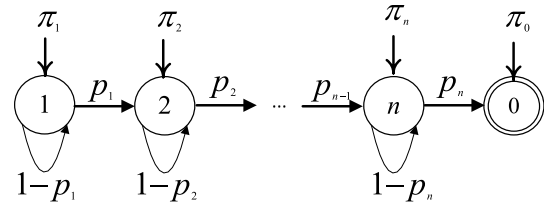


Fig. 10 The DPH representation of $GSNB(n, p_i)$

$$\begin{aligned} &= \sum_{i=1}^m \left(\prod_{\substack{j=1 \\ i \neq j}}^{m+1} \frac{p_j}{p_j - p_i} \right) p_i (1 - p_i)^{k+m} \\ &\quad - (1 - p_{m+1})^{k+1} \sum_{i=1}^m \left(\prod_{\substack{j=1 \\ i \neq j}}^{m+1} \frac{p_j}{p_j - p_i} \right) p_i (1 - p_i)^{m-1} \end{aligned}$$

By adding and subtracting the $(m+1)$ th term of the first sum to the entire expression, we get

$$\begin{aligned} &= \sum_{i=1}^{m+1} \left(\prod_{\substack{j=1 \\ i \neq j}}^{m+1} \frac{p_j}{p_j - p_i} \right) p_i (1 - p_i)^{k+m} \\ &\quad - (1 - p_{m+1})^{k+1} \sum_{i=1}^{m+1} \left(\prod_{\substack{j=1 \\ i \neq j}}^{m+1} \frac{p_j}{p_j - p_i} \right) p_i (1 - p_i)^{m-1} \end{aligned}$$

Due to the finite sum of Lagrange polynomials, the second term on the right-hand side of above equation is equal to zero. Therefore, the proof is completed. \square

The DPH representation of generalized shifted negative binomial distribution is calculated by Eq. (9) and given by

$$\begin{aligned} \boldsymbol{\pi}_{GSNB} &= \boldsymbol{\pi}_{GNB} \mathbf{T}_{GNB}^n = (\pi_1, \pi_2, \dots, \pi_n), \\ \mathbf{T}_{GSNB} &= \mathbf{T}_{GNB} = \begin{pmatrix} 1 - p_1 & p_1 & 0 & \dots & 0 & 0 \\ 0 & 1 - p_2 & p_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 - p_{n-1} & p_{n-1} \\ 0 & 0 & 0 & \dots & 0 & 1 - p_n \end{pmatrix}, \\ \mathbf{t}_{GSNB} &= \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ p_n \end{pmatrix} \end{aligned} \tag{26}$$

The mean of the generalized shifted negative binomial distributed random variable is calculated as $E[Y] = \sum_{i=1}^n \frac{1-p_i}{p_i}$ and the variance as $\text{Var}[Y] = \sum_{i=1}^n \frac{1-p_i}{p_i^2}$. Figure 10 shows the DPH representation of generalized shifted negative binomial distribution.

The mixed shifted geometric distribution ($Y \sim \text{MSG}(n, p_i, \pi_i)$) is a convex mixture of n shifted geometric distributions. The probability mass function is $\text{Pr}(Y = y) = \sum_{i=1}^n \pi_i(1 - p_i)^y p_i$, for $y = 0, 1, 2, \dots$ where $\pi_i > 0$ for all phases i and $\sum_{i=1}^n \pi_i = 1$. This distribution is also the mixed geometric distribution with shifted by one unit ($Y = X - 1$), and its DPH representation and factorial moment are given by Eqs. (27) and (28) and shown in Fig. 11.

$$\begin{aligned} \boldsymbol{\pi}_{\text{MSG}} &= (\pi_1, \pi_2, \dots, \pi_n) = \boldsymbol{\pi}_{\text{MG}} \mathbf{T}_{\text{MG}} = (\boldsymbol{\pi}_{\text{MG}_1}(1 - p_1), \\ &\boldsymbol{\pi}_{\text{MG}_2}(1 - p_2), \dots, \boldsymbol{\pi}_{\text{MG}_n}(1 - p_n)), \end{aligned} \tag{27}$$

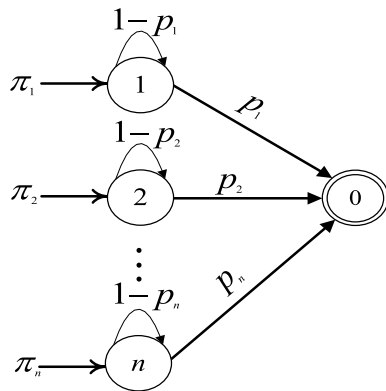
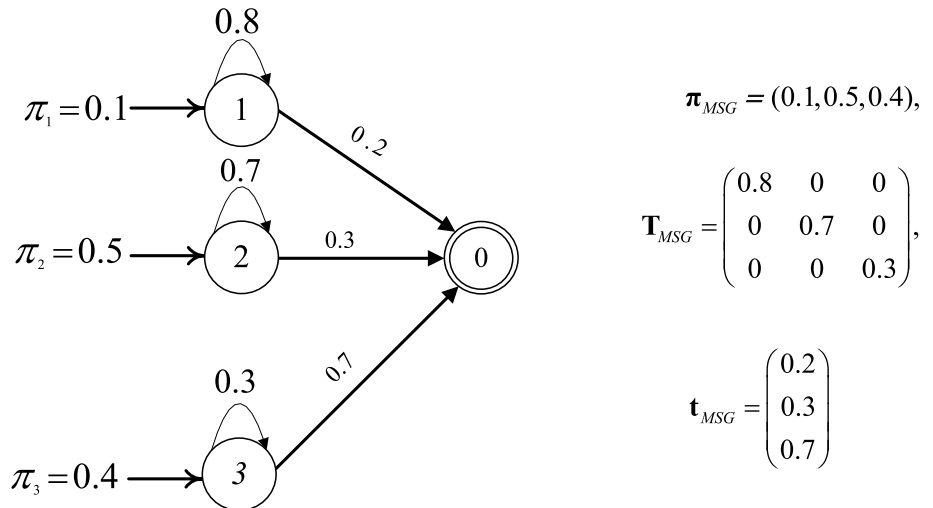


Fig. 11 The DPH representation of $\text{MSG}(n, p_i, \pi_i)$

Fig. 12 The DPH representation of $\text{MSG}(3, p_1=0.2, p_2=0.3, p_3=0.7, \pi_1=0.1, \pi_2=0.5, \pi_3=0.4)$



$$\mathbf{T}_{\text{MSG}} = \begin{pmatrix} 1-p_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1-p_2 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1-p_{n-1} & 0 \\ 0 & 0 & 0 & \dots & 0 & 1-p_n \end{pmatrix}, \quad \mathbf{t}_{\text{MSG}} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n-1} \\ p_n \end{pmatrix}$$

$$f_k = E[Y(Y-1)\dots(Y-k+1)] = \sum_{i=1}^n \pi_i k! \left(\frac{1-p_i}{p_i}\right)^k \tag{28}$$

for $k = 1, 2, \dots$

The first moment is obtained as $E[Y] = \sum_{i=1}^n \frac{\pi_i(1-p_i)}{p_i}$, and its variance is given by $\text{Var}[Y] = \sum_{i=1}^n \pi_i \frac{(1-p_i)(2-p_i)}{p_i^2} - (\sum_{i=1}^n \frac{\pi_i(1-p_i)}{p_i})^2$. For instance, we assume that $Y \sim \text{MSG}(3, p_1 = 0.2, p_2 = 0.3, p_3 = 0.7, \pi_1 = 0.1, \pi_2 = 0.5, \pi_3 = 0.4)$, the pmf of Y is $\text{Pr}(Y = y) = 0.1 * (0.8)^y * 0.2 + 0.5 * (0.7)^y * 0.3 + 0.4 * (0.3)^y * 0.7$, for $y = 0, 1, 2, \dots$. The DPH and diagrammatic representation of Y are shown in Fig. 12. The mean and variance of Y are $E[Y] = 1.7381$ and $\text{Var}[Y] = 7.5085$, respectively.

The mixed shifted negative binomial distribution ($Y \sim \text{MSNB}(m, n_i, p_i, \alpha_i)$) is considered a mixture of m mutually independent shifted negative binomial distributions weighted with the probabilities $\alpha_1, \alpha_2, \dots, \alpha_m$, where $\alpha_i \geq 0$ and the vector α is stochastic, i.e., $\sum_{i=1}^m \alpha_i = 1$. Let n_i denote the number of phases of the i th shifted negative binomial distribution. Then the probability mass function is $\text{Pr}(Y = y) = \sum_{i=1}^m \alpha_i \binom{y+n_i-1}{n_i-1} (1-p_i)^y p_i^{n_i}$, for $y = 0, 1, \dots$. The state space includes $\sum_{i=1}^m n_i$ transient and one absorbing state. The DPH representation of the mixed shifted negative binomial distribution can be described by Eq. (29).

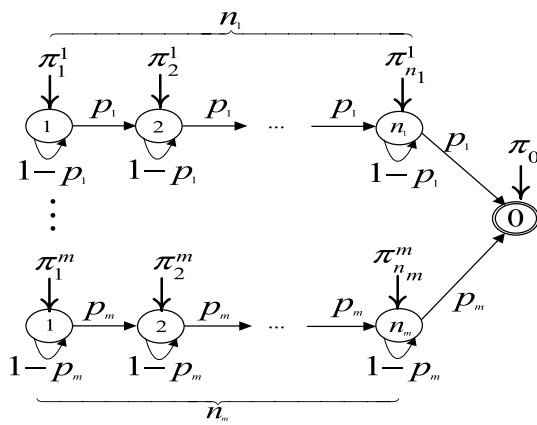


Fig. 13 The DPH representation of $MSNB(m, n_i, p_i, \pi_i)$

$$\begin{aligned} \boldsymbol{\pi}_{MSNB} &= (\boldsymbol{\pi}_{MSNB_1}, \boldsymbol{\pi}_{MSNB_2}, \dots, \boldsymbol{\pi}_{MSNB_m}), \\ \boldsymbol{\pi}_{MSNB_i} &= (\pi_1^i, \pi_2^i, \dots, \pi_{n_i}^i), \\ \pi_j^i &= \alpha_i \binom{n_i}{j-1} (1-p_i)^{n_i-(j-1)} p_i^{j-1} \end{aligned} \tag{29}$$

for $j = 1, \dots, n_i, i = 1, \dots, m$

$$\mathbf{T}_{MSNB} = \begin{pmatrix} \mathbf{T}_1 & 0 & \dots & 0 \\ 0 & \mathbf{T}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathbf{T}_m \end{pmatrix},$$

where \mathbf{T}_i is calculated based on Eq. (14).

Derivation of MSNB representation

We define initial probabilities as $\boldsymbol{\pi}_{MSNB} = (\boldsymbol{\pi}_{MSNB_1}, \boldsymbol{\pi}_{MSNB_2}, \dots, \boldsymbol{\pi}_{MSNB_m})$ and $\boldsymbol{\pi}_{MSNB_i} = (\pi_1^i, \pi_2^i, \dots, \pi_{n_i}^i)$ for $i = 1, \dots, m$. Each $\boldsymbol{\pi}_{MSNB_i}$ is initial probability of transient state for shifted negative binomial distribution and by using Eq. (9) can reach to $\boldsymbol{\pi}_{MSNB_i} = \alpha_i(1, 0, \dots, 0)\mathbf{T}_i^{n_i}$ for $i = 1, \dots, m$. Based on Eq. (21), $(1, 0, \dots, 0)\mathbf{T}_i^{n_i} = \binom{n_i}{j-1} (1-p_i)^{n_i-(j-1)} p_i^{j-1}$. Then, we can calculate each π_j^i for $j = 1, \dots, n_i, i = 1, \dots, m$ by Eq. (29). □

A factorial moment of the mixed shifted negative binomial distribution is calculated by Eq. (30), and its diagrammatic representation is illustrated in Fig. 13.

$$\begin{aligned} f_k &= E[Y(Y-1) \dots (Y-k+1)] \\ &= \sum_{i=1}^m \alpha_i \frac{\Gamma(n_i+k)}{\Gamma(n_i)} \frac{(1-p_i)^k}{p_i^k} \end{aligned} \tag{30}$$

for $k = 1, 2, \dots$

Conclusions and suggestions for future research

In this paper, we presented the definition, properties, characteristics and PH representations of acyclic discrete phase-type (ADPH) distributions and their subclasses (ADPH family). The simplest ADPH distribution is the geometric distribution defined by either of the two discrete probability distributions, the geometric or the shifted geometric distribution. Based on the two definitions of the geometric distribution, we proposed a distinct classification for the ADPH subclasses and introduced their definitions. The advantage of our proposed classifications is in applying precise representations of each subclass and preventing miscalculation of the probability mass function, by computing the ADPH family based on geometric and shifted geometric distributions. To this end, we developed the PH representation for each subclass and proved them by using the closure properties of ADPH, especially “shifted DPH.” In addition, all the subclasses of ADPH analogous to ACPH are considered and their properties and characteristics are discussed.

For further research, applying the proposed classification in real stochastic modeling and developing fitting algorithms based on the ADPH subclasses are suggested.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Adan I, van Eenige M, Resing J (1995) Fitting discrete distributions on the first two moments. *Probab Eng Inf Sci* 9:623–632

Akar N (2015) Fitting matrix geometric distributions by model reduction. *Stoch Model* 31:292–315

Alfa A (2016) *Applied discrete-time queues*. Springer, New York

Altioek T (1985) On the phase-type approximations of general distributions. *IIE Trans* 17:110–116

Asmussen S, Nerman O, Olsson M (1996) Fitting phase-type distributions via the EM algorithm. *Scand J Stat* 23:419–441

Assaf D, Levikson B et al (1982) Closure of phase type distributions under operations arising in reliability theory. *Ann Probab* 10:265–269

Augustin R, Büscher K-J (1982) Characteristics of the COX-distribution. *ACM Sigmetrics Perform Eval Rev* 12:22–32

Bobbio A, Cumani A (1992) ML estimation of the parameters of a PH distribution in triangular canonical form. *Comput Perform Eval* 22:33–46

Bobbio A, Telek M (1994) A benchmark for PH estimation algorithms: results for Acyclic-PH. *Stoch Model* 10:661–677

Bobbio A, Horváth A, Scarpa M, Telek M (2003) Acyclic discrete phase type distributions: properties and a parameter estimation algorithm. *Perform Eval* 54:1–32

- Bobbio A, Horváth A, Telek M (2004) The scale factor: a new degree of freedom in phase-type approximation. *Perform Eval* 56:121–144
- Bobbio A, Horváth A, Telek M (2005) Matching three moments with minimal acyclic phase type distributions. *Stoch Model* 21:303–326
- Bodrog L, Horváth A, Telek M (2008) Moment characterization of matrix exponential and Markovian arrival processes. *Ann Oper Res* 160:51–68
- Botta RF, Harris CM (1986) Approximation with generalized hyper-exponential distributions: weak convergence results. *Queueing Syst* 1:169–190
- Buchholz P, Kriege J (2009) A heuristic approach for fitting MAPs to moments and joint moments. In: Sixth international conference on the quantitative evaluation of systems, 2009. QEST'09, pp 53–62
- Buchholz P, Kemper P, Kriege J (2010) Multi-class Markovian arrival processes and their parameter fitting. *Perform Eval* 67:1092–1106
- Bux W, Herzog U (1977) The phase concept: approximation of measured data and performance analysis. *Computer Performance*. Amsterdam, North-Holland, pp 23–38
- Callut J, Dupont P (2006) Sequence discrimination using phase-type distributions. *Mach Learn ECML 2006*:78–89
- Chauveau D, Martin CF, van Rooij ACM, Ruymgaart FH (1996) Discrete signed mixtures of exponentials. *Stoch Model* 12:245–263
- Commault C (2003) Linear positive systems and phase-type representations. *Positive systems*. Springer, New York, pp 281–288
- Commault C, Chemla J-P (1993) On dual and minimal phase-type representations. *Stoch Model* 9:421–434
- Commault C, Chemla J-P (1996) An invariant of representations of phase-type distributions and some applications. *J Appl Probab* 33(2):368–381
- Commault C, Mocanu S (2003) Phase-type distributions and representations: some results and open problems for system theory. *Int J Control* 76:566–580
- Commault C, Mocanu S et al (2002) A generic property of phase-type representations. *J Appl Probab* 39:775–785
- Cox DR (1955) A use of complex probabilities in the theory of stochastic processes. In: *Mathematical proceedings of the Cambridge Philosophical Society*, pp 313–319
- Cumani A (1982) On the canonical representation of homogeneous Markov processes modelling failure-time distributions. *Microelectron Reliab* 22:583–602
- David A, Larry S (1987) The least variable phase type distribution is Erlang. *Stoch Model* 3:467–473
- Dayar T (2005) On moments of discrete phase-type distributions. *Formal techniques for computer systems and business processes*. Springer, Berlin, pp 51–63
- De Liefvoort A (1990) The moment problem for continuous distributions. Unpubl Tech report, Univ Missouri, WP-CM-1990-02, Kansas City
- Dehon M, Latouche G (1982) A geometric interpretation of the relations between the exponential and generalized Erlang distributions. *Adv Appl Probab* 14:885–897
- Dufresne D (2007) Fitting combinations of exponentials to probability distributions. *Appl Stoch Model Bus Ind* 23:23–48
- Éltető T, Vadera P (2008) Finding upper-triangular representations for phase-type distributions with 3 distinct real poles. *Ann Oper Res* 160:139–172
- Esparza LJR, Nielsen BF, Bladt M (2010) Maximum likelihood estimation of phase-type distributions. Technical University of Denmark Danmarks Tekniske Universitet, Department of Applied Mathematics and Computer Science Institut for Matematik og Computer Science
- Fackrell MW (2003) Characterization of matrix-exponential distributions. The University of Adelaide
- Fackrell M, He Q-M, Taylor P et al (2010) The algebraic degree of phase-type distributions. *J Appl Probab* 47:611–629
- Faddy MJ (1993) A structured compartmental model for drug kinetics. *Biometrics* 49:243–248
- Faddy MJ (1994) Examples of fitting structured phase-type distributions. *Appl Stoch Model data Anal* 10:247–255
- Faddy MJ (1998) On inferring the number of phases in a Coxian phase-type distribution. *Stoch Model* 14:407–417
- Faddy MJ (2002) Penalized maximum likelihood estimation of the parameters in a Coxian phase-type distribution. *Matrix-analytic methods: theory and Applications*. World Sci, Singapore, pp 107–114
- Faddy MJ, McClean SI (1999) Analysing data on lengths of stay of hospital patients using phase-type distributions. *Appl Stoch Model Bus Ind* 15:311–317
- Feldmann A, Whitt W (1997) Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. In: *INFOCOM'97. Sixteenth annual joint conference of the IEEE computer and communications societies. Driving the information revolution, Proceedings IEEE*, pp 1096–1104
- Gong L (2014) Erlang-based methods in modeling losses in insurance and applications. University of Toronto
- Harris CM, Sykes EA (1984) *Likelihood estimation for generalized mixed exponential distributions*. Clarendon Press, Oxford
- Harris CM, Marchal WG, Botta RF (1992) A note on generalized hyperexponential distributions. *Commun Stat Stoch Model* 8:179–191
- He Q-M, Zhang H (2005) A note on unicyclic representations of phase type distributions. *Stoch Model* 21:465–483
- He Q-M, Zhang H (2006a) PH-invariant polytopes and Coxian representations of phase type distributions. *Stoch Model* 22:383–409
- He Q-M, Zhang H (2006b) Spectral polynomial algorithms for computing bi-diagonal representations for phase type distributions and matrix-exponential distributions. *Stoch Model* 22:289–317
- He Q-M, Zhang H (2007) Coxian approximations of matrix-exponential distributions. *Calcolo* 44:235–264
- He Q-M, Zhang H (2008) An algorithm for computing minimal Coxian representations. *Inform J Comput* 20:179–190
- He Q-M, Zhang H, Xue J (2011) Algorithms for coxianization of phase-type generators. *Inform J Comput* 23:153–164
- Horváth G (2013) *Moment matching-based distribution fitting with generalized hyper-erlang distributions. Analytical and stochastic modeling techniques and applications*. Springer, Berlin, pp 232–246
- Horvath A, Telek M (2000) Approximating heavy tailed behaviour with phase type distributions. In: *3rd International conference on matrix-analytic methods in stochastic models, MAM3, (Leuven, Belgium), Citeseer*, pp 391–400
- Horváth A, Telek M (2002) Phfit: a general phase-type fitting tool. In: *Proceedings of the computer performance evaluation, modelling techniques and tools*, pp 82–91
- Horváth A, Telek M (2007a) Matching more than three moments with acyclic phase type distributions. *Stoch Model* 23:167–194
- Horváth G, Telek M (2007b) A canonical representation of order 3 phase type distributions. In: *European performance engineering workshop*. Springer, Berlin, Heidelberg, pp 48–62
- Horváth G, Telek M (2009) On the canonical representation of phase type distributions. *Perform Eval* 66:396–409
- Horváth I, Telek M (2015) A constructive proof of the phase-type characterization theorem. *Stoch Model* 31:316–350
- Horváth I, Papp J, Telek M (2015) On the canonical representation of order 3 discrete phase type distributions. *Electron Notes Theor Comput Sci* 318:143–158
- Hu L, Jiang Y, Zhu J, Chen Y (2013) Hybrid of the scatter search, improved adaptive genetic, and expectation maximization

- algorithms for phase-type distribution fitting. *Appl Math Comput* 219:5495–5515
- Isensee C, Horton G (2005) Approximation of discrete phase-type distributions. In: *Proceedings of the 38th annual symposium on simulation*, pp 99–106
- Jain M, Bhagat A (2014) Unreliable bulk retrial queues with delayed repairs and modified vacation policy. *J Ind Eng Int* 10:63
- Johnson MA (1993) Selecting parameters of phase distributions: combining nonlinear programming, heuristics, and Erlang distributions. *ORSA J Comput* 5:69–83
- Johnson MA, Taaffe MR (1989) Matching moments to phase distributions: mixtures of Erlang distributions of common order. *Stoch Model* 5:711–743
- Johnson MA, Taaffe MR (1990a) Matching moments to phase distributions: nonlinear programming approaches. *Stoch Model* 6:259–281
- Johnson MA, Taaffe MR (1990b) Matching moments to phase distributions: density function shapes. *Stoch Model* 6:283–306
- Johnson MA, Taaffe MR (1991) An investigation of phase-distribution moment-matching algorithms for use in queueing models. *Queueing Syst* 8:129–147
- Khayari REA, Sadre R, Haverkort BR (2003) Fitting world-wide web request traces with the EM-algorithm. *Perform Eval* 52:175–191
- Kim K, Thomas N (2011) A fitting method with generalized Erlang distributions. *Simul Model Pract Theory* 19:1507–1517
- Kroese DP, Taimre T, Botev ZI (2013) *Handbook of Monte Carlo methods*. Wiley, New York
- Latouche G, Ramaswami V (1999) *Introduction to matrix analytic methods in stochastic modeling*. SIAM, Philadelphia
- Lee SCK, Lin XS (2010) Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North Am Actuar J* 14:107–130
- Maier RS (1991) The algebraic construction of phase-type distributions. *Stoch Model* 7:573–602
- Maier RS, O’Cinneide CA (1992) A closure characterisation of phase-type distributions. *J Appl Probab* 29:92–103
- Malhotra M, Reibman A (1993) Selecting and implementing phase approximations for semi-Markov models. *Stoch Model* 9:473–506
- Marie R (1980) Calculating equilibrium probabilities for $\lambda(n)/Ck/1/N$ queues. *Sigmetrics Perform Eval Rev* 9:117–125. <https://doi.org/10.1145/1009375.806155>
- Marshall AH, Zenga M (2012) Experimenting with the Coxian phase-type distribution to uncover suitable fits. *Methodol Comput Appl Probab* 14:71–86
- Mészáros A, Telek M (2013) Canonical representation of discrete order 2 MAP and RAP. In: *European workshop on performance engineering*, pp 89–103
- Mészáros A, Papp J, Telek M (2014) Fitting traffic traces with discrete canonical phase type distributions and Markov arrival processes. *Int J Appl Math Comput Sci* 24:453–470
- Mocanu Ş, Commault C (1999) Sparse representations of phase-type distributions. *Stoch Model* 15:759–778
- Neuts MF (1975) Computational uses of the method of phases in the theory of queues. *Comput Math with Appl* 1:151–166
- Neuts MF (1981) *Matrix-geometric solutions in stochastic models: an algorithmic approach*. Johns Hopkins University, Baltimore
- O’Cinneide CA (1989) On non-uniqueness of representations of phase-type distributions. *Commun Stat Stoch Model* 5:247–259
- O’Cinneide CA (1991) Phase-type distributions and invariant polytopes. *Adv Appl Probab* 23:515–535
- O’Cinneide CA (1993) Triangular order of triangular phase-type distributions. *Stoch Model* 9:507–529
- O’Cinneide CA (1999) Phase-type distributions: open problems and a few properties. *Stoch Model* 15:731–757
- Osogami T, Harchol-Balter M (2003a) A closed-form solution for mapping general distributions to minimal PH distributions. Springer, Berlin
- Osogami T, Harchol-Balter M (2003b) Necessary and sufficient conditions for representing general distributions by Coxians. In: *International conference on modelling techniques and tools for computer performance evaluation*. Springer, Berlin, Heidelberg, pp 182–199
- Osogami T, Harchol-Balter M (2006) Closed form solutions for mapping general distributions to quasi-minimal PH distributions. *Perform Eval* 63:524–552
- Panchenko A, Thümmler A (2007) Efficient phase-type fitting with aggregated traffic traces. *Perform Eval* 64:629–645
- Papp J, Telek M (2013) Canonical representation of discrete phase type distributions of order 2 and 3. In: *Proceedings of UK performance evaluation workshop, UKPEW*
- Parr WC, Schucany WR (1980) Minimum distance and robust estimation. *J Am Stat Assoc* 75:616–624
- Pérez-Ocón R, Ruiz-Castro JE (2003) A multiple-absorbent markov process in survival studies: application to breast cancer. *Biom J* 45:783–797
- Pulungan R, Hermanns H (2008a) Effective minimization of acyclic phase-type representations. In: *International conference on analytical and stochastic modeling techniques and applications*, pp 128–143
- Pulungan R, Hermanns H (2008b) The minimal representation of the maximum of Erlang distributions. In: *2008 14th GI/ITG conference on measuring, modelling and evaluation of computer and communication systems (MMB)*, pp 1–15
- Pulungan R, Hermanns H (2013) A construction and minimization service for continuous probability distributions. *Int J Softw Tools Technol Transf* 17:77–90
- Reinecke P, Krauß T, Wolter K (2013) Phase-type fitting using HyperStar. In: *European Workshop on Performance Engineering*. Springer, Berlin, Heidelberg, pp 164–175
- Riska A, Diev V, Smiri E (2004) An EM-based technique for approximating long-tailed data sets with PH distributions. *Perform Eval* 55:147–164
- Ross S (2014) *A first course in probability*. Pearson Education Inc, Upper Saddle River
- Sadre R, Haverkort BR (2008) Fitting heavy-tailed HTTP traces with the new stratified EM-algorithm. In: *4th international telecommunication networking workshop on QoS in multiservice IP networks, 2008, IT-NEWS 2008*, pp 254–261
- Sauer CH, Chandy KM (1975) Approximate analysis of central server models. *IBM J Res Dev* 19:301–313
- Schmickler L (1992) Meda: mixed erlang distributions as phase-type representations of empirical distribution functions. *Commun Stat Stoch Model* 8:131–156. <https://doi.org/10.1080/15326349208807217>
- Sen A, Balakrishnan N (1999) Convolution of geometrics and a reliability problem. *Stat Probab Lett* 43:421–426
- Singh LN, Dattatreya GR (2007) Estimation of the hyperexponential density with applications in sensor networks. *Int J Distrib Sens Netw* 3:311–330
- Singh C, Billinton R, Lee SY (1977) The method of stages for non-Markov models. *IEEE Trans Reliab* 26:135–137
- Slud EV, Suntorncost J (2014) Parametric survival densities from phase-type models. *Lifetime Data Anal* 20:459–480
- Telek M (2000) The minimal coefficient of variation of discrete phase type distributions. In: *3rd international conference on matrix-analytic methods in stochastic models, MAM3*, (Leuven, Belgium). Notable Publications Inc, pp 391–400
- Telek M, Heindl A (2002) Matching moments for acyclic discrete and continuous phase-type distributions of second order. *Int J Simul Syst Sci Technol* 3:47–57
- Telek M, Horváth G (2007) A minimal representation of Markov arrival processes and a moments matching method. *Perform Eval* 64:1153–1168

- Thümmler A, Buchholz P, Telek M (2006) A novel approach for phase-type fitting with the EM algorithm. *IEEE Trans Dependable Secur Comput* 3:245–258
- Van Der Heijden MC (1988) On the three-moment approximation of a general distribution by a Coxian distribution. *Probab Eng Inf Sci* 2:257–261
- Vanden Bosch PM, Dietz DC, Pohl EA (2000) Moment matching using a family of phase-type distributions. *Stoch Model* 16:391–398
- Wang J, Zhou H, Xu F, Li L (2005) Hyper-Erlang based model for network traffic approximation. *Parallel and distributed processing and applications*. Springer, Berlin, pp 1012–1023
- Wang J, Zhou H, Zhou M, Li L (2006) A general model for long-tailed network traffic approximation. *J Supercomput* 38:155–172
- Wang J, Liu J, She C (2008) Segment-based adaptive hyper-Erlang model for long-tailed network traffic approximation. *J Supercomput* 45:296–312
- Whitt W (1982) Approximating a point process by a renewal process, I: two basic methods. *Oper Res* 30:125–147
- Yang WY, Cao W, Chung T-S, Morris J (2005) *Applied numerical methods using MATLAB*. Wiley, Hoboken
- Yu K, Huang M-L, Brill PH (2012) An algorithm for fitting heavy-tailed distributions via generalized hyperexponentials. *Inform J Comput* 24:42–52