



Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
<http://scientiairanica.sharif.edu>



Estimating the parameters of mixed shifted negative binomial distributions via an EM algorithm

M. Varmazyar^a, R. Akhavan-Tabatabaei^b, N. Salmasi^a, and M. Modarres^{a,*}

a. *Department of Industrial Engineering, Sharif University of Technology, Tehran, P.O. Box 11155-8639, Iran.*

b. *School of Management, Sabanci University, Istanbul, Turkey.*

Received 24 September 2017; received in revised form 12 November 2017; accepted 2 January 2018

KEYWORDS

Parameter estimation;
 Discrete Phase-Type
 (DPH) distributions;
 Expectation-
 Maximization (EM)
 algorithm;
 Mixed shifted negative
 binomial distributions.

Abstract. Discrete Phase-Type (DPH) distributions have one property that is not shared by Continuous Phase-Type (CPH) distributions, i.e., representing a deterministic value as a DPH random variable. This property distinguishes the application of DPH in stochastic modeling of real-life problems, such as stochastic scheduling, in which service time random variables should be compared with a deadline that is usually a constant value. In this paper, we consider a restricted class of DPH distributions, called Mixed Shifted Negative Binomial (MSNB), and show its flexibility in producing a wide range of variances as well as its adequacy in fitting fat-tailed distributions. These properties render MSNB applicable to represent data on certain types of service time. Therefore, we adapt an Expectation-Maximization (EM) algorithm to estimate the parameters of MSNB distributions that accurately fit trace data. To present the applicability of the proposed algorithm, we use it to fit real operating room times and a set of benchmark traces generated from continuous distributions as case studies. Finally, we illustrate the efficiency of the proposed algorithm by comparing its results with those of two existing algorithms in the literature. We conclude that our proposed algorithm outperforms other DPH algorithms in fitting trace data and distributions.

© 2019 Sharif University of Technology. All rights reserved.

1. Introduction

Phase-type (PH) distributions, introduced by Neuts [1,2], are a family of discrete and continuous probability distributions constructed by mixtures of geometric or exponential phases. Special properties and characteristics of PH distributions make them attractive for approximating a variety of random variables and modeling real-world stochastic arrival or

service times [3]. The common property to Discrete PH (DPH) and Continuous PH (CPH) distributions is that the convolution, minimum, maximum, and convex mixtures of PH random variables yield new PH random variables. However, representing deterministic values by a PH random variable (deterministic values property) is restricted to the DPH distributions [2].

With regards to the property of deterministic values, shifting (adding a constant to) the DPH random variables results in a new DPH structure [2]. Using this property, we can derive the distribution function of the convolution or maximum/minimum of the DPH random variables with constant values. One important application of this derivation is in stochastic scheduling problems. In such problems, random processing times should be combined with constant durations (e.g., set up times) or the maximum/minimum of random times

*. *Corresponding author. Tel.: +98 21 6616 5719;
 Fax: +98 21 66022702
 E-mail addresses: varmaziyar@ie.sharif.edu (M. Varmazyar); akhavan@sabanciuniv.edu (R. Akhavan-Tabatabaei); nsalmasi@sharif.edu (N. Salmasi); modarres@sharif.edu (M. Modarres)*

should be compared with constant deadline values in order to compute the objective function.

Shifted Negative Binomial (SNB) distribution is a subclass of DPH distributions and the discrete analogous to the Erlang distribution in the CPH class. The Mixed Shifted Negative Binomial (MSNB) distribution is a mixture of independent SNB distributions and a subclass of DPH. The MSNB family is the discrete equivalent of Hyper-Erlang Distribution (HErD) in the CPH family. Similar to the HErD, which can approximate any distribution on \mathbb{R} [4,5], the MSNB subclass models distributions that are defined on \mathbb{N} . Moreover, the Coefficient of Variation (CoV) of MSNB distribution can be tuned to less than, equal to, or greater than unity. Also, it provides good approximations to fat-tailed distributions. These properties render MSNB desirable for fitting data on high variance service times or when the tail of service time distribution is heavier than the exponential distribution [6]. Therefore, in this research, our goal is to estimate the parameters of the MSNB distribution in order to fit empirical service time data with such characteristics.

In the previous research, some approaches to fitting the parameters of a general DPH distribution have been provided. Horváth and Telek [7] presented a tool called Phfit, which estimated the parameters of Acyclic DPH (ADPH) and Acyclic CPH (ACP) distributions to minimize a distance measure by using a non-linear optimization method. The purpose of their algorithm was optimization by iterative linearization to numerically compute the partial derivatives.

The first detailed study on DPH and ADPH distributions and their fitting methods by Maximum Likelihood (ML) estimation was conducted by Bobbio et al. [8]. They also proved several properties of ADPH distributions and showed that the ADPH distributions had a unique minimal representation, named the canonical form.

Callut and Dupont [9] studied the Expectation-Maximization (EM) algorithm to fit the general DPH distributions. Their algorithm was an adapted version of the EM algorithm proposed by Asmussen et al. [10], which applied to continuous PH distributions. Three different methods, namely an EM algorithm, a Gibbs sampler algorithm, and a Quasi-Newton method, were applied for maximum likelihood estimation of general DPH distributions by Bladt et al. [11].

New results on the canonical representation of DPH with 2 and 3 phases (DPH (2) and DPH (3)) as well as Discrete Markov Arrival Processes (DMAP) with 2 phases (DMAP (2)) were presented by Meszáros et al. [12]. They presented explicit formulae to match parameters applying canonical forms (DPH (2), DPH (3), and DMAP (2)) and gave moments and correlation bounds. They showed the efficiency of fitting procedures with numerical examples. The Canonical

Representation of DPH (CRDPH) distributions with 3 phases (CRDPH (3)) was investigated by Horváth et al. [13]. They demonstrated that the problem of CRDPH (3) was far more complicated than the one of CPH distribution with 3 phases. They also needed to define 8 different subclasses of DPH distribution with 3 phases for their canonical representation.

In this research, we present an EM algorithm to estimate the parameters of MSNB distributions. The advantages of the EM algorithm over other alternatives such as non-linear programming have been explored by Springer and Urban [14]. The most significant property of the EM algorithm is guaranteeing the increase in the likelihood at each iteration. The other reason is that the EM algorithm needs neither the analytic expression nor the gradient of the log-likelihood function, and it does not even require being differentiable. It also estimates the parameters of the distribution from a given data trace when the data has some missing values or is incomplete. The time complexity of the EM algorithm is linear, only depending on the number of SNB branches, and independent of the number of states. The number of SNB branches might be remarkably lower than the number of states in most cases. Therefore, the fitting algorithm by EM algorithm is rather stable because of the specific structure of the ADPH distribution, which provides a reliable and fast convergence of the EM algorithm.

In Section 2, we define the MSNB distribution by PH representation and prove some of its properties for approximating fat-tailed trace data. In Section 3, we propose a specialized EM algorithm to fit the parameters of the MSNB distribution and use this modified EM algorithm to fit continuous distributions. In Section 4, we showcase the applicability of the proposed algorithm to fit real-world operating room service time data as well as a set of benchmark traces generated from conventional distributions. We compare the accuracy of our results with two other algorithms designed by Thümmel et al. [4] and Bladt et al. [11]. In Section 5, we provide the conclusions and show directions for future research.

2. Mixed shifted negative binomial distribution and its properties

The DPH distributions are constructed by a system of one or a group of inter-related geometric distributions occurring in sequence or phases (see Appendix A for detailed description and properties).

Shifted Geometric (SG) distribution is another nonequivalent definition of the geometric distribution, which describes the number of failures before the first success (as opposed to the number of trials until success) in an infinite sequence of independent Bernoulli trials. Shifted Negative Binomial (SNB) distribution is

the convolution of a number of SG random variables, defined as the number of failures before reaching a fixed number of successes in a sequence of Bernoulli trials. In Appendix B, we present the definition and properties of SG and SNB distributions in PH representation. In the rest of this section, we present the definition and properties of the Mixed Shifted Negative Binomial (MSNB) distribution.

A mixed shifted negative binomial distribution ($X \sim \text{MSNB}(m, n_i, p_i, \alpha_i)$) is considered as a mixture of m mutually independent SNB random variables, weighted by the probability vector $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_m]$, in which $\alpha_i \geq 0$ and the vector α is stochastic, i.e., $\sum_{i=1}^m \alpha_i = 1$. Let n_i define the number of phases of the i th SNB distribution. Then, the MSNB probability mass function is given by Eq. (1) [29]:

$$\Pr(X = x) = \sum_{i=1}^m \alpha_i \binom{x + n_i - 1}{n_i - 1} (1 - p_i)^x p_i^{n_i},$$

for $x = 0, 1, \dots$ (1)

The state space includes one absorbing state and $\sum_{i=1}^m n_i$ transient states. The DPH representation of the MSNB distribution can be described by Eq. (2) and is illustrated in Figure 1.

$$\pi_{\text{MSNB}} = (\pi_{\text{MSNB}_1}, \pi_{\text{MSNB}_2}, \dots, \pi_{\text{MSNB}_m}),$$

$$\pi_{\text{MSNB}_i} = (\pi_1^i, \pi_2^i, \dots, \pi_{n_i}^i),$$

$$\pi_j^i = \alpha_i \binom{n_i}{j-1} (1 - p_i)^{n_i - (j-1)} p_i^{j-1},$$

$$\text{for } j = 1, \dots, n_i, \quad i = 1, \dots, m,$$

$$\mathbf{T}_{\text{MNB}} = \begin{pmatrix} \mathbf{T}_1 & 0 & \dots & 0 \\ 0 & \mathbf{T}_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \mathbf{T}_m \end{pmatrix}, \quad (2)$$

where \mathbf{T}_i is a matrix shown in Eq. (3):

$$\mathbf{T}_i = \begin{pmatrix} 1-p_i & p_i & 0 & \dots & 0 & 0 \\ 0 & 1-p_i & p_i & \dots & 0 & 0 \\ \dots & \dots & \dots & \ddots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1-p_i & p_i \\ 0 & 0 & 0 & \dots & 0 & 1-p_i \end{pmatrix},$$

$i = 1, \dots, m. \quad (3)$

The k th factorial moment of the MSNB distribution is calculated by Eq. (4):

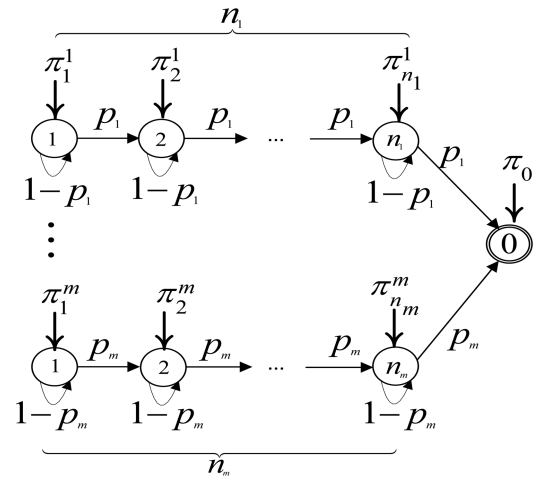


Figure 1. The DPH representation of MSNB (m, n_i, p_i, α_i) .

$$f_k = E[X(X-1)\dots(X-k+1)]$$

$$= \sum_{i=1}^m \alpha_i \frac{\Gamma(n_i + k)}{\Gamma(n_i)} \frac{(1 - p_i)^k}{p_i^k},$$

for $k = 1, 2, \dots$ (4)

Let \mathcal{H} be a set of all MSNB distributions with n states, i.e.:

$$\mathcal{H} = \left\{ \Pr(X) : \Pr(X = x) = \sum_{i=1}^m \alpha_i \binom{x + n_i - 1}{n_i - 1} (1 - p_i)^x p_i^{n_i}, \right.$$

$$x = 0, 1, \dots; \quad \sum_{i=1}^m \alpha_i = 1, \quad \sum_{i=1}^m n_i = n \left. \right\}. \quad (5)$$

Note that set \mathcal{H} contains all MSNB distributions including at most n states and the MSNB with less than n states can be acquired by setting some α_i values to zero. We present the following theorem to show the versatility of the MSNB in approximating general distributions on \mathbb{N} .

Theorem 1. The set \mathcal{H} has the following properties:

1. Let \mathcal{F} denote the set of all discrete distributions with finite support on \mathbb{N} . Then, \mathcal{H} is a dense set in \mathcal{F} , i.e., any distribution on \mathbb{N} can be approximated by MSNB distribution;
2. Let h denote an MSNB distribution and be out of the set \mathcal{H} , with $n \geq 2$. The parameters of h (MSNB distribution) can be tuned such that the Coefficient of Variation (CoV) takes an arbitrary value less

than, equal to, or greater than unity. It can also be tuned such that its CoV takes value as large as or as small as desired.

Proof. (1) is proven by weak convergence rule as shown by Verbelen [15]. We concentrate on the proof of (2). The first and the second moments of MSNB distribution are given by Eqs. (1) and (4):

$$E[X] = \sum_{i=1}^m \alpha_i \frac{n_i q_i}{p_i}, (q_i = 1 - p_i),$$

$$E[X^2] = \sum_{i=1}^m \alpha_i \frac{n_i q_i (1 + n_i q_i)}{p_i^2}. \quad (6)$$

We calculate the coefficient of variation by Eq. (7):

$$\text{CoV}^2 = \frac{E[X^2] - E[X]^2}{E[X]^2} = \frac{\sum_{i=1}^m \alpha_i \frac{n_i q_i (1 + n_i q_i)}{p_i^2}}{\left[\sum_{i=1}^m \alpha_i \frac{n_i q_i}{p_i} \right]^2} - 1. \quad (7)$$

Let us choose $m = 2$ and define:

$$h(\alpha) = \frac{\alpha \left[\frac{n_1 q_1 (1 + n_1 q_1)}{p_1^2} \right] + (1 - \alpha) \left[\frac{n_2 q_2 (1 + n_2 q_2)}{p_2^2} \right]}{\left[\alpha \frac{n_1 q_1}{p_1} + (1 - \alpha) \frac{n_2 q_2}{p_2} \right]^2} - 1. \quad (8)$$

It is clear that $h(0) = \frac{1}{n_2 q_2}$ and $h(1) = \frac{1}{n_1 q_1}$. Since $h(\alpha)$ is continuous in $[0, 1]$ and we assume that $n_1 q_1 \leq n_2 q_2$, $h(\alpha)$ can take any value in $[\frac{1}{n_2 q_2}, 1]$. If we select $n_1 q_1 = 1$ and assume that $n_2 q_2$ varies in the interval $[1, \infty)$, then we can demonstrate that $h(\alpha)$ can take any value in $(0, 1]$. Then, the coefficient of variation for MSNB distribution can be tuned to take an arbitrary value in $(0, 1]$.

Next, we demonstrate that $h(\alpha)$ can also take an arbitrary value in $[1, \infty)$. We consider $n_1 q_1 = n_2 q_2 = 1$; then:

$$g(\alpha, p_1, p_2) = \frac{2 \left[\frac{\alpha}{p_1^2} \right] + 2 \left[\frac{(1-\alpha)}{p_2^2} \right]}{\left[\frac{\alpha}{p_1} + \frac{(1-\alpha)}{p_2} \right]^2} - 1$$

$$= \frac{2\alpha + 2(1-\alpha) \left[\frac{p_1}{p_2} \right]^2}{\left[\alpha + (1-\alpha) \frac{p_1}{p_2} \right]^2} - 1,$$

$$\rightarrow \frac{2}{\alpha} - 1 \quad \left(\frac{p_1}{p_2} \rightarrow 0 \right), \quad (9)$$

from which we reach the conclusion that when we select α and $\frac{p_1}{p_2}$ sufficiently small, $g(\alpha, p_1, p_2)$ can take value as large as desired. In addition, $g(0.5, p, p) = 1$. Since $g(\alpha, p_1, p_2)$ is continuous in $[0, 1] \times [0, 1] \times [0, 1]$,

$g(\alpha, p_1, p_2)$ can take any value in $[1, \infty)$. Thus, the coefficient of variation can also be tuned to take value in $[1, \infty)$. This ends the proof of (2). ■

Notably, Theorem 1 states that any probability mass function with its domain in the natural numbers can be approximated arbitrarily closely by the proper selection of MSNB parameters. Then, for every point of a general probability mass function $f \in \mathcal{F}$, choosing a sequence of MSNB distributions with m SNB branches with each one having scale parameter, p , is possible.

Next, we show that the MSNB distribution can also approximate the fat-tailed distributions properly. Let $F(x)$ be a cumulative MSNB distribution function given in Eq. (1), and $\bar{F}(x) = 1 - F(x)$ be a complementary cumulative MSNB distribution function calculated by:

$$\bar{F}(x) = \sum_{i=1}^m \alpha_i \sum_{k=x+1}^{\infty} \binom{k + n_i - 1}{n_i - 1} (1 - p_i)^k p_i^{n_i}. \quad (10)$$

The distribution is called fat-tailed if Complementary Cumulative Distribution Function (CCDF) ($\bar{F}(x)$) is in the order of $\frac{1}{x^r}$, where x is sufficiently large for $r > 0$. Intuitively, a fat-tailed distribution has a “fat tail” compared to the exponential distribution. Based on the definition of fat-tailed distribution, we study the property of $\bar{F}(x)$. In order to determine whether a distribution is a fat-tailed distribution, we observe the probability mass functions on the log-linear graphs. We consider MSNB distribution with MSNB ($m = 4$, $n_i = [2, 3, 4, 5]$, $p_i = [0.2, 0.3, 0.5, 0.7]$, and $\alpha_i = [0.1, 0.2, 0.3, 0.4]$) as an example and compare it with the exponential distribution as illustrated in Figure 2. In fact, the MSNB distribution does provide the fat-tailed property in the time range of interest.

The infinite variance is another specification of the fat-tailed distribution. We show that the MSNB

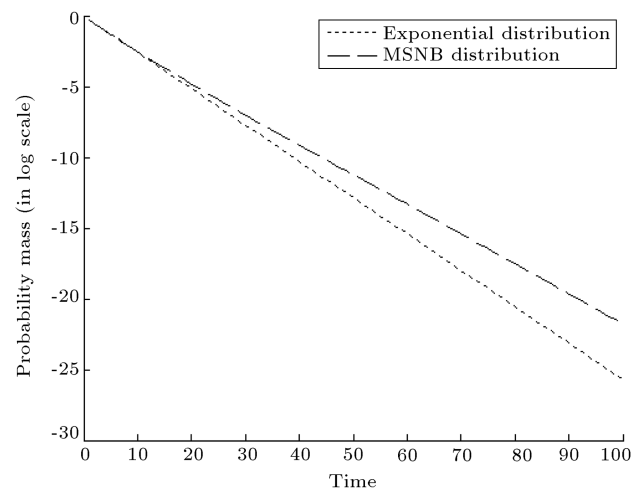


Figure 2. Comparing the fat-tailed property for MSNB and exponential distributions.

distributions can be tuned to take a finite bounded expectation with a sufficiently large variance. In Eq. (6), we show that we can select α_i and p_i appropriately; then, $E[x]$ is finite while $E[x^2]$ is sufficiently large. For example, when we select $m = 2$, $r > 1$, $\alpha_1 = \frac{1}{\sqrt{r}}$, $p_1 = \frac{1}{\sqrt{r}}$, and $p_2 = 1$:

$$\begin{aligned} E[x] &= n_1 \left(1 - \frac{1}{\sqrt{r}}\right) < n_1, \\ E[x^2] &= n_1 \sqrt{r} \left(1 - \frac{1}{\sqrt{r}}\right) \left(1 + n_1 \left(1 - \frac{1}{\sqrt{r}}\right)\right) \\ &\geq n_1 \sqrt{r} \rightarrow \infty \quad (r \rightarrow \infty). \end{aligned} \quad (11)$$

In this section, we demonstrated the PH representation of MSNB distribution and its properties such as approximating any distribution on \mathbb{N} , wide range of CoV, and fat-tailed property.

In the next section, we introduce an EM algorithm to estimate the parameters of the MSNB distribution by fitting high variance or fat-tailed service time data.

3. An EM algorithm to fit mixed shifted negative binomial distributions

The EM algorithm is an iterative approach to derive maximum likelihood for estimating the parameters of stochastic models, which are dependent on unobserved latent variables. Each iteration of EM algorithm consists of two steps: an expectation (E) step and a maximization (M) step. In the E step, a function is created for the expectation of the log-likelihood and evaluated using the current estimate of the parameters. In the M step, the parameters are computed while maximizing the expected log-likelihood determined in the E step. These estimated parameters are then applied to find the distribution of the latent variables in the next E step.

In the rest of this section, we first explain fitting a mixture-density with the EM algorithm and present its application to the MSNB distributions. Then, we discuss the implementation of the EM algorithm over the weighted discrete sample to fit continuous distributions by the MSNB distribution.

3.1. MSNB parameter estimation via an EM algorithm

One of the most common issues related to EM algorithm is the mixture-density parameter estimation method [16,17]. The probabilistic model of this method is assumed as follows:

$$\Pr(x|\Theta) = \sum_{i=1}^m \alpha_i \Pr_i(x|\theta_i), \quad (12)$$

where the parameters are $\Theta = (\alpha_1, \dots, \alpha_m, \theta_1, \dots, \theta_m)$, in which $\sum_{i=1}^m \alpha_i = 1$ and each \Pr_i is a probability mass function parameterized by θ_i . In other words, m -component probability mass functions are mixed using m mixing coefficients: α_i , $i = 1, \dots, m$. Generally, θ_i is a vector of parameters for each probability mass function, \Pr_i , while it is a single value in the proposed EM algorithm.

Let $\mathcal{X} = \{x_1, \dots, x_N\}$ be an incomplete dataset and $\mathcal{Y} = \{y_j\}_{j=1}^N$ be the existence of unobserved data items, where the values inform which component function “generates” each data item of \mathcal{X} . Therefore, assume that $y_j \in \{1, \dots, m\}$, for $j = 1, \dots, N$, and $y_j = i$ if the j th sample (x_j) is generated by the i th mixture component. If the value of \mathcal{Y} is known, the likelihood expression can be calculated by Eq. (13):

$$\begin{aligned} \log L(\Theta|\mathcal{X}, \mathcal{Y}) &= \log(\Pr(\mathcal{X}, \mathcal{Y}|\Theta)) \\ &= \sum_{j=1}^N \log(\Pr(x_j|y_j) \Pr(y_j)) \\ &= \sum_{j=1}^N \log(\alpha_{y_j} \Pr_{y_j}(x_j|\theta_{y_j})). \end{aligned} \quad (13)$$

The dilemma in dealing with Eq. (13) is that the values of y_j are usually unknown. If y_j is considered as random values drawn from a random variable \mathcal{Y} , the derivation of expression for the probability mass function of unobserved data, shown by $q(y)$, is possible. At first, the parameters for probability mass function are guessed, i.e., the parameters $\hat{\Theta} = (\hat{\alpha}_1, \dots, \hat{\alpha}_m, \hat{\theta}_1, \dots, \hat{\theta}_m)$ are guessed as proper parameters for likelihood $L(\hat{\Theta}|\mathcal{X}, \mathcal{Y})$. Given $\hat{\Theta}$, the probability $\Pr_i(x_j|\hat{\theta}_i)$ can be easily computed for each i and j . Moreover, the mixing parameters, denoted by α_i , are considered as prior probabilities of each mixture component or the probability of selecting the i th mixture component. Thus, the probability mass function of the unobserved data calculated by the observed data \mathcal{X} and the estimates $\hat{\Theta}$ are computed by applying Bayes’s rule in Eqs. (14) and (15).

$$\begin{aligned} q(y_j|x_j, \hat{\Theta}) &= \frac{q(y_j|\hat{\Theta}) \Pr(x_j|y_j, \hat{\Theta})}{\Pr(x_j|\hat{\Theta})} \\ &= \frac{\hat{\alpha}_{y_j} \cdot \Pr_{y_j}(x_j|\hat{\theta}_{y_j})}{\sum_{i=1}^m \hat{\alpha}_i \cdot \Pr_i(x_j|\hat{\theta}_i)}, \end{aligned} \quad (14)$$

$$q(\mathbf{y}|\mathcal{X}, \hat{\Theta}) = \prod_{j=1}^N q(y_j|x_j, \hat{\Theta}), \quad (15)$$

where $\mathbf{y} = \{y_1, \dots, y_N\}$ is a sample of the unobserved data independently drawn from \mathcal{Y} . The expectation value of the complete-data log-likelihood is given by Eq. (16) by considering the unknown random variable \mathcal{Y} , the observed data \mathcal{X} , and the current parameter estimates $\hat{\Theta}$:

$$Q(\Theta, \hat{\Theta}) = E \left[\log \mathcal{L}(\Theta | \mathcal{X}, \mathbf{y}) | \mathcal{X}, \hat{\Theta} \right] \\ = \sum_{\mathbf{y} \in \mathcal{Y}} \sum_{i=1}^N \log(\alpha_{y_i} p_{y_i}(x_i | \theta_{y_i})) \prod_{j=1}^N q(y_j | x_j, \hat{\Theta}). \quad (16)$$

Integrating Eqs. (13) and (15) into Eq. (16) produces Eq. (17) according to Bilmes [18].

$$Q(\Theta, \hat{\Theta}) = \sum_{l=1}^m \sum_{i=1}^N \log(\alpha_l) \cdot q(l | x_i, \hat{\Theta}) \\ + \sum_{l=1}^m \sum_{i=1}^N \log(p_l(x_i | \theta_l)) \cdot q(l | x_i, \hat{\Theta}). \quad (17)$$

The computation of the expectation in Eq. (17) includes the E step of the EM algorithm. Generally, the main problem in calculating this expectation is to find an expression for the distribution of the unobserved data, although the distribution of the unobserved data can be easily calculated by Eqs. (14) and (15). The purpose of M step in EM algorithm is to maximize the expectation determined in the E step by considering Θ . To maximize Eq. (17), we can independently maximize the term including α_i (first sum in Eq. (17)) and the term including θ_i (second sum in Eq. (17)), because both terms are not relevant. According to Bilmes [15] and McLachlan and Krishnan [16], a Lagrange multiplier can be used to obtain the expression for α_i , resulting in:

$$\alpha_i = \frac{1}{N} \sum_{j=1}^N q(i | x_j, \hat{\Theta}). \quad (18)$$

In order to estimate the parameters of MSNB distributions, we apply the proposed EM algorithm. The i th mixture component of MSNB distributions is an SNB distribution with a fixed number of phases shown in Eq. (19).

$$\Pr_i(x_j | p_i) = \binom{x_j + n_i - 1}{n_i - 1} (1 - p_i)^{x_j} p_i^{n_i}. \quad (19)$$

The mixture distribution is defined by the vector of Θ . The parameters α_i and p_i , $i = 1, \dots, m$, are determined according to Eqs. (18) and (21). The value of p_i can be calculated by Eq. (21) and integrating Eq. (19) into Eq. (20) and applying logarithm-rules:

$$\sum_{j=1}^N q(i | x_j, \hat{\Theta}) \frac{\partial}{\partial p_i} \log(\Pr_i(x_j | p_i)) = 0, \quad (20)$$

$$p_i = \frac{n_i \sum_{j=1}^N q(i | x_j, \hat{\Theta})}{\sum_{j=1}^N (x_j + n_i) q(i | x_j, \hat{\Theta})}. \quad (21)$$

Note that the value of p_i is always no more than one and the condition of probability is satisfied by this equation.

3.1.1. Implementation and time complexity of the EM algorithm

The EM algorithm starts with estimating the initial parameters $\Theta = (\alpha_1, \dots, \alpha_m, p_1, \dots, p_m)$ and iterates between the E step and the M step by using three Eqs. (14), (18), and (21) to reach the stopping criteria. The flowchart of this algorithm is presented in Figure 3. Each iteration is guaranteed to increase the

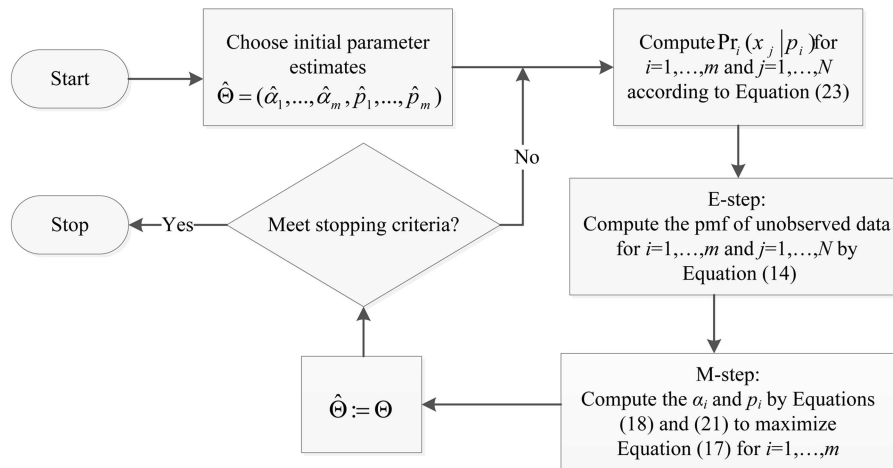


Figure 3. Flowchart of the EM algorithm tailored to mixed shifted negative binomial distribution.

log-likelihood value and the algorithm is guaranteed to converge to a local maximum of the likelihood function [19]. To check whether convergence is achieved, the relative difference of the log-likelihood values of successive iterations is computed and the algorithm stops when the computed difference is less than a predefined value of ε , e.g., $\varepsilon = 10^{-6}$.

By a given number of branches (m) and a number of phases of each branch (n_i), the EM algorithm determines the best setting of parameter vectors α and \mathbf{p} . However, in order to find the “best” distribution, all MSNB distributions have to be considered as candidates. Thümmel et al. [4] presented the discrete parameter setting for Hyper-Erlang Distribution (HErD) and used the recursive formula to consider all possible settings in an algorithmic fashion. In this paper, Thümmel’s recursive formula [4] is applied to find the best MSNB distribution according to Eq. (22).

$$\varphi_i(n, j) = \sum_{r=j}^{\lfloor n/i \rfloor} \varphi_{i-1}(n-r, r),$$

where:

$$\varphi_1(n, j) = \begin{cases} 0, & \text{if } j > n \\ 1, & \text{if } j \leq n \end{cases} \quad (22)$$

Direct computation of the SNB distribution (Eq. (19)) can reveal numerical difficulties because large factorials must be computed for a high number of phases (e.g., $n > 50$). As a solution to this difficulty, the logarithmic form is proposed and shown in Eq. (23):

$$\begin{aligned} \Pr_i(x_j | p_i) \\ = e^{n_i \ln(p_i) + x_j \ln(1-p_i) + \ln(n_i + x_j - 1)! - \ln(x_j)! - \ln(n_i - 1)!}. \end{aligned} \quad (23)$$

Also, the logarithms of the factorial values are computed by Eq. (24) before the EM algorithm starts:

$$\ln r! = \sum_{i=1}^r \ln i. \quad (24)$$

The time complexity of the proposed EM algorithm depends on its two main steps (E and M). The complexity of the E step is $O(m.N)$ in computing the numerator and denominator of the unobserved data based on Eq. (14). The complexity of the M step is also $O(m.N)$. The overall complexity of each iteration in the EM algorithm is $O(m.N)$.

3.2. Fitting MSNB distributions to continuous distributions

In this section, we apply the proposed algorithm to fit the data generated from continuous distributions. To approximate a continuous distribution by MSNB, a discretization method should be considered. The following steps are required to this end.

Step 1: Apply the discretization method and generate discrete samples;

Step 2: Implement the EM algorithm by using MSNB over the discrete sample provided in Step 1.

3.2.1. Discretization method

Bobbio et al. [8] introduced a discretization method based on the Cumulative Distribution Function (CDF). In this method, finite (ordered) set $S = \{x_1, x_2, x_3, \dots\}$ (with $x_1 < x_2 < x_3 < \dots$) is defined by multiplying an integer with discretization interval (δ), i.e., $x_i = i\delta$. Then, a probability mass is assigned to each element of S by Eq. (25):

$$\begin{aligned} w_i &= F_X\left(\frac{x_i + x_{i+1}}{2}\right) - F_X\left(\frac{x_{i-1} + x_i}{2}\right), \\ i > 1 \quad \text{and} \quad w_1 &= F_X\left(\frac{x_1 + x_2}{2}\right), \end{aligned} \quad (25)$$

where $F_X(x)$ is the cdf of x and w_i is the probability associated with x_i .

Dougherty et al. [20] proposed the “equal with interval binning” method for the case of using the generated traces instead of probability distribution function. It involves sorting the value bounded by x_{\min} and x_{\max} , and dividing the range of values by k equally sized bins by Eq. (26), in which k is a parameter given by the user:

$$\delta = \left\lceil \frac{x_{\max} - x_{\min}}{k} \right\rceil. \quad (26)$$

The bin thresholds are computed by $x_{\min} + i\delta$, where $i = 1, \dots, k$. The probability mass (w_i) is defined by Eq. (27):

$$w_i = \frac{m_i}{\sum_{j=1}^k m_j}, \quad \text{for } i = 1, \dots, k. \quad (27)$$

In this equation, m_i is the number of values that fall into each of the bins.

3.2.2. Implementation of the EM algorithm over the weighted discrete sample

In Section 3.1, the proposed EM algorithm is based on the incomplete dataset, $\mathcal{X} = \{x_1, \dots, x_N\}$, with weight one. When the discretization method is considered as an approximation of continuous distributions, the incomplete dataset, $\mathcal{X} = \{x_1, \dots, x_N\}$, with weight one is substituted by incomplete dataset, $\mathcal{X}^* = \{x_1^*, \dots, x_{N^*}^*\}$, with weight w_i (Eq. (25) or (27)). The modified set includes N^* elements with different values. Thus, Eqs. (18) and (21) can be efficiently calculated by Eqs. (28) and (29), respectively:

$$\alpha_i = \frac{1}{\sum_{j=1}^{N^*} w_j} \sum \lim_{j \rightarrow 1}^{N^*} w_j \cdot q\left(i \mid x_j^*, \hat{\Theta}\right), \quad (28)$$

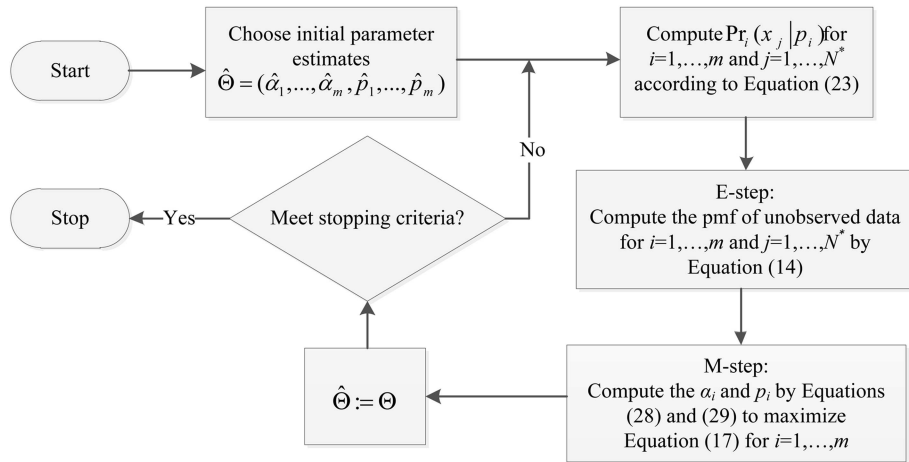


Figure 4. Flowchart of the modified EM algorithm tailored to mixed shifted negative binomial distribution.

$$p_i = \frac{n_i \sum_{j=1}^N w_j \cdot q(i | x_j^*, \hat{\Theta})}{\sum_{j=1}^N w_j \cdot (x_j^* + n_i) \cdot q(i | x_j^*, \hat{\Theta})}. \quad (29)$$

The flowchart of the modified EM algorithm is shown in Figure 4. Similar to the proposed algorithm in Section 3, the modified EM algorithm based on Eq. (22) determines the best setting of the parameter vectors α and p by a given number of branches (m) and number of phases of each branch (n_i). The computational complexity of the E and M steps is $O(m.N^*)$. Thus, the overall time complexity of one iteration for the EM algorithm is $O(m.N^*)$.

4. Results

In this section, we apply the proposed EM algorithm to fit real-world operating room service times as well as six synthetically generated trace datasets from continuous distributions. The algorithm is coded in MATLAB 2014a [21] and performed on a personal computer with Intel(R) Core(TM) i3-2120, running at 3.30 GHz with 8 GB RAM.

4.1. Fitting MSNB distributions to real operating room data

To study an example with real data, the proposed EM algorithm is applied to fit three different datasets of Start Anesthesia to Start Operation (SASO), Operating Room (OR), and Post-Anesthesia Care Unit (PACU) times. SASO time is defined as the duration of time from starting anesthesia procedure until starting the operation. OR time is the time of operation and PACU time is the recovery time of patients from anesthesia after the operation. The three datasets are collected from orthopedic operating theatre in the Scottish NHS hospital from 1998 to 1999, and are used in other studies [22].

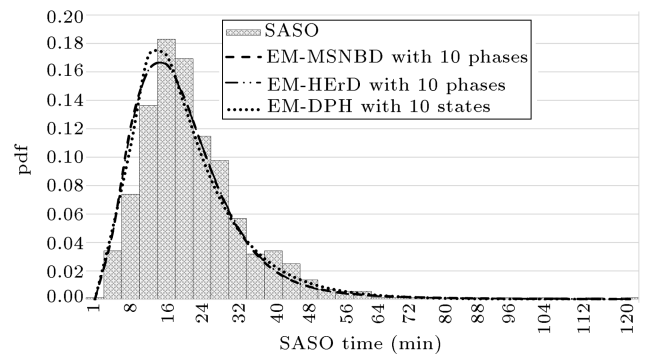


Figure 5. Original dataset of SASO and the approximating distributions.

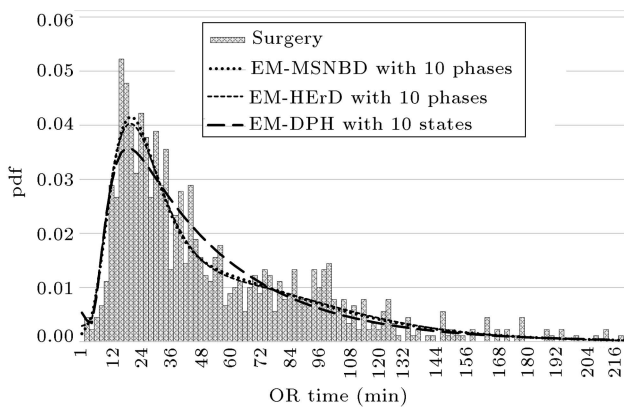
We compare the fitting quality of the proposed EM algorithm (EM-MSNBD) with the quality of an EM algorithm designed for general DPH distributions (EM-DPH) presented by Bladt et al. [11] and the EM algorithm for the continuous HERD (EM-HERD) developed by Thümmel et al. [4]. This algorithm is one of the best-performing algorithms in fitting CPH [23]. The fitting quality is measured by four different criteria, i.e., the first three moments and the chi-square statistic ($\chi^2 = \frac{(\text{trace data} - \text{result of algorithm})^2}{\text{trace data}}$). The stopping criterion for all algorithms is convergence with $\varepsilon = 10^{-6}$.

In Figures 5 to 7, the empirical distributions of SASO, OR, and PACU traces as well as the probability mass and density functions for the fitted MSNB, general DPH, and HERD with 10 states are illustrated. The curve fitting of MSNB shows that the proposed algorithm fits adequately to the fat-tailed data traces and confirms the fat-tailed property of MSNB distributions discussed in Section 2.

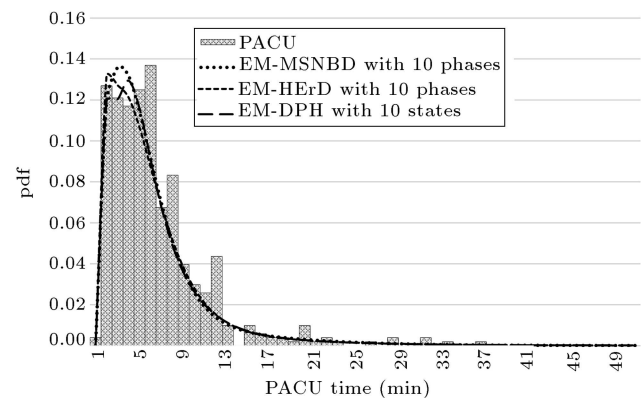
The results of fitting quality are presented in Table 1. Relative errors of the first three moments and the chi-square measure are represented in Table 2. According to Table 2, the relative errors of the first

Table 1. Quality indices for fitted MSNBD, DPH distribution, and HErD for three datasets.

		Trace	EM-MSNBD	EM-DPH	EM-HErD
SASO	First moment	19.75	19.75	19.75	19.75
	Second moment	557.56	560.11	538.59	556.99
	Third moment	22600.79	23487.23	19186.93	22935.57
	CPU time (sec)		4.18	190.05	4.60
	Number of phases		2,2,6	—	2,2,2,4
	α_i		0.04; 0.34; 0.62	—	0.06; 0.04; 0.05; 0.85
	p_i or λ_i		0.04; 0.10; 0.24	—	0.05; 0.10; 0.48; 0.21
OR	First moment	51.44	51.44	51.44	51.44
	Second moment	4321.42	4372.55	4373.14	4331.21
	Third moment	495266.84	515595.72	542221.03	497585.35
	CPU time (sec)		2.81	1002.07	3.46
	Number of phases		1,3,6	—	2,4,4
	α_i		0.00; 0.57; 0.43	—	0.00; 0.47; 0.53
	p_i or λ_i		0.33; 0.04; 0.21	—	1.48; 0.05; 0.16
PACU	First moment	9.49	9.49	9.49	9.49
	Second moment	1559.53	1575.99	1946.51	1606.13
	Third moment	684423.87	854044.08	1445876.41	893073.27
	CPU time (sec)		2.12	378.70	2.29
	Number of phases		1,3,6	—	1,2,3,4
	α_i		0.02; 0.13; 0.85	—	0.02; 0.11; 0.75; 0.12
	p_i or λ_i		0.00; 0.20; 0.57	—	0.00; 0.15; 0.58; 2.82

**Figure 6.** Original dataset of OR and the approximating distributions .

moment for all algorithms in each dataset are zero. The relative errors of the second and third moments and the chi-square measure for EM-MSNBD and EM-HErD algorithms are nearly equal in all datasets, while these measures are relatively high for the EM-DPH algorithm. Therefore, we conclude that the fitting quality of our proposed EM-MSNBD algorithm is as adequate as the fitting quality of EM-HErD.

**Figure 7.** Original dataset of PACU and the approximating distributions.

A single-factor (algorithm effect) experiment with three levels (number of algorithms) is performed in order to statistically compare the performance of the three mentioned algorithms [24]. The hypothesis to be tested is as follows:

$$H_0 : \text{EM-MSNBD} = \text{EM-DPH} = \text{EM-HErD},$$

$$H_1 : \text{Otherwise.} \quad (30)$$

Table 2. Relative errors of first three moments and chi-square measure for three datasets.

		EM-MSNBD	EM-DPH	EM-HErD
SASO	First moment	0.00%	0.00%	0.00%
	Second moment	0.46%	3.40%	0.10%
	Third moment	3.92%	15.11%	1.48%
	Chi-square	0.08	0.20	0.08
OR	First moment	0.00%	0.00%	0.00%
	Second moment	1.18%	1.19%	0.23%
	Third moment	4.10%	9.48 %	0.46%
	Chi-square	0.19	0.22	0.20
PACU	First moment	0.00%	0.00%	0.00%
	Second moment	1.05%	24.81%	2.98%
	Third moment	24.78%	111.25%	30.48%
	Chi-square	0.65	0.63	0.65

Table 3. The ANOVA table for real OR problem.

Source	DF	Sum of squares	Mean square	F-value	Pr > F
Dependent variable: Y					
Model	7	0.85237222	0.12176746	3.73	0.0298
Error	10	0.32612222	0.03261222		
Corrected total	17	1.17849444			
	R-square	Coeff Var	Root MSE	Objective mean	
	0.723272	138.3231	0.180589	0.130556	
Source	DF	Type I SS	Mean square	F-value	Pr > F
Method	2	0.19214444	0.09607222	2.95	0.0987
Block	5	0.66022778	0.13204556	4.05	0.0286

The single factor experiment is conducted as follows:

$$Y_{ab} = \mu + \text{method}_a + \text{block}_b + \varepsilon_{ab}, \quad (31)$$

with the following characteristics:

Y_{ab}	The response variable (relative errors);
μ	The overall mean;
method_a	The algorithm (effect factor), $a = 1, 2, 3$;
block_b	The second and third moments of each dataset, $b = 1, \dots, 6$;
ε_{ab}	The error term.

The three algorithms are considered as the effect factor and the second and third moments of each dataset are considered as blocks. The experimental design is coded in the Statistical Analysis System (SAS) software package, release 9.1 [25], with a significance

level of 10%. The ANOVA tables are provided in Table 3. Based on the results of ANOVA, there is a statistically significant difference among the algorithms (the p -values are less than 0.1).

In order to find the algorithm with the best performance, we also perform the Tukey test (Table 4). The results show that there is a significant difference among the EM-DPH and the two other algorithms, and the EM-MSNBD and EM-HErD algorithms are not statistically different. Throughout all experiments, the EM-MSNBD and EM-HErD algorithms outperform EM-DPH in terms of fitting quality and CPU time requirements.

4.2. Fitting MSNB distributions to continuous distributions

Thümmel et al. [4] generated six groups of samples by applying partial peak function of Weibull (1.0, 5.0), value function of uniform (0.5, 1.5), a fat-

Table 4. Tukey test for real OR problem; differences of least squares means.

Algorithms comparison*	Difference between means	90% Confidence limits	
2-1	0.2183	[0.0294, 0.4073]	***
2-3	0.2200	[0.0310, 0.4090]	***
1-2	-0.2183	[-0.4073, -0.0294]	***
1-3	0.0017	[-0.1873, 0.1906]	
3-2	-0.2200	[-0.4090, -0.0310]	***
3-1	-0.0017	[-0.1906, 0.1873]	

*1 is the EM-MSNBD algorithm; 2 is the EM-DPH algorithm; and 3 is the EM-HERD algorithm.

tailed distribution function of Weibull (1.0, 0.5), step function with shifted exponent, heavy-tailed distribution function of Pareto-II (1.5, 2.0), and multi-peak function with Matrix Exponential, which are all representative theoretical distributions. Each sample size is 10^4 . The six groups of samples are chosen for the following reasons. Weibull distributions are often applied in the interpretation of experimental data in engineering such as reliability, queuing, transmission, etc. [26]. Uniform and shifted exponential distributions are difficult to closely approximate with a PH distribution [26]. Pareto-II distribution is an example of a heavy-tailed distribution, which is not monotonically decreasing [4]. The matrix exponential distribution has a matrix exponential representation, while it is not PH [26].

In this part, we test the three algorithms, namely the modified EM algorithm proposed in Section 3.2 (M-EM-MSNBD), the EM-DPH, and the EM-HERD, using the six groups of samples used by Thümmel et al. [4]. The fitting effects of the six groups of samples with three algorithms are illustrated in Figure 8 and the detailed records of fitting quality measures are presented in Tables 5 and 6.

In order to compare our proposed algorithm with the two other algorithms, we draw conclusions based on Table 6 for each algorithm. The results of relative errors of the moments and the chi-square measure show that the M-EM-MSNBD algorithm can be considerably more efficient than the other two algorithms in fitting the fat-tailed distributions such as Weibull (1.0, 0.5), as seen in Figure 8(a), and in the matrix exponential distributions, as shown in Figure 8(f). The results indicate that EM-DPH algorithm yields the lowest relative errors of moments and chi-square measure in symmetric distributions such as Weibull (1.0, 5.0) and uniform (0.5, 1.5) distributions as illustrated in Figure 8(b) and (c).

The results of EM-HERD algorithm show that the fitting quality of the EM-HERD in Pareto-II (1.5, 2.0) and shifted exponent distributions (Figure 8(d) and (e)) is as good as the fitting quality for EM-MSNBD. Based on six groups of samples, we firstly conclude that the results of relative errors of moments and the chi-square measure for the proposed algorithm and EM-HERD are nearly close to each other and different from the third one. Secondly, our proposed algorithm fits adequately to the fat-tailed distribution.

To statistically compare the performance of the three mentioned algorithms, we perform a single-factor (algorithm effect) experiment with three levels (number of algorithms), as in Section 4.1 and Eqs. (30) and (31). The first, second, and third moments of each function are considered as blocks.

The experimental design is coded in SAS, release 9.1 [25], with a significance level of 10%. The ANOVA tables and the results of Tukey test are provided in Tables 7 and 8. These results show that there is a significant difference among the EM-DPH and the two other algorithms, and the M-EM-MSNBD and EM-HERD algorithms are not statistically different. Throughout all experiments, M-EM-MSNBD and EM-HERD algorithms perform better than EM-DPH in terms of CPU time requirement, as seen in Table 5. Also, in the majority of cases, M-EM-MSNBD and EM-HERD algorithms outperform EM-DPH in terms of fitting quality, as seen in Table 6.

5. Conclusions and suggestions for future research

Discrete phase-type distributions have several advantages over their continuous equivalents. In this research, we considered the problem of fitting a restricted class of DPH distributions, namely Mixed Shifted Negative Binomial (MSNB) distributions, to trace data. We proved some properties of MSNB distribution to fit trace data and developed a fitting algorithm based on Expectation Maximization (EM) to estimate its parameters.

We measured the effectiveness of this algorithm by comparing it with two existing methods in the literature, EM-HERD (for continuous data) and EM-DPH, using the operating room data and six benchmark traces. To evaluate their goodness of fit, we considered the first three moments and the chi-square (χ^2) measure and devised a single-factor (algorithm effect) experiment with three levels (number of algorithms).

The results of the experiment showed that (M-) EM-MSNBD and EM-HERD algorithms were not statistically different, but they both outperformed the EM-DPH algorithm. Given that the EM-HERD algo-

Table 5. Quality indices for fitted MSNBD, HErD, and DPH for synthetically generated traces.

		Trace	M-EM-MSNBD	EM-DPH	EM-HErD
Weibull (1.0, 0.5)	First moment	1.99	2.01	1.94	1.99
	Second moment	25.61	25.81	17.65	23.78
	Third moment	995.13	1163.30	276.20	638.03
	CPU time (sec)		10.99	404.52	83.36
	Number of phases		2,2,2,2,2	—	1,1,1,1,1,1,1,1,1,1
	α_i		0.00;0.05;0.17;0.27;0.51	—	0.07; 0.27; 0.16; 0.11; 0.09; 0.08; 0.08; 0.09; 0.04; 0.01;
	p_i or λ_i		0.00; 0.01; 0.04; 0.12; 0.47	—	0.09; 0.31; 0.89; 1.69; 2.24; 5.68; 8.69; 54.02; 642.27; 29520.07
Matrix exponential	First moment	1.06	1.06	1.06	1.06
	Second moment	2.12	2.15	1.90	2.16
	Third moment	6.57	7.16	4.59	7.24
	CPU time (sec)		5.15	378.70	23.99
	Number of phases		1,4,5	—	1,4,5
	α_i		0.01, 0.38, 0.61	—	0.06, 0.38, 0.56
	p_i or λ_i		0.03, 0.18, 0.49	—	0.48, 2.23, 10.82
Weibull (1.0, 5.0)	First moment	0.92	0.92	0.92	0.92
	Second moment	0.89	0.96	0.92	0.93
	Third moment	0.90	1.12	1.01	1.03
	CPU time (sec)		1.99	33.97	8.61
	Number of phases		10	—	10
	α_i		1	—	1
	p_i or λ_i		0.26	—	10.85
Uniform (0.5, 1.5)	First moment	0.99	0.99	0.99	0.99
	Second moment	1.08	1.12	1.08	1.09
	Third moment	1.24	1.41	1.29	1.30
	CPU time (sec)		2.77	16.78	9.78
	Number of phases		10	—	10
	α_i		1	—	1
	p_i or λ_i		0.25	—	10.04
Pareto-II (1.5, 2.0)	First moment	4.34	3.80	3.29	4.33
	Second moment	1057.61	88.78	24.69	340.67
	Third moment	1768568.21	5974.58	300.25	114715.85
	CPU time (sec)		3.55	429.82	23.69
	Number of phases		3,3,3	—	1,2,3,4
	α_i		0.02; 0.20; 0.78	—	0.01; 0.11; 0.36; 0.52
	p_i or λ_i		0.01; 0.04; 0.16	—	0.01; 0.18; 0.96; 3.34
Shifted exponential	First moment	1.51	1.51	1.51	1.51
	Second moment	3.57	3.61	3.29	3.61
	Third moment	11.55	11.81	9.31	11.79
	CPU time (sec)		11.58	107.41	46.83
	Number of phases		3,3,4	—	1,3,6
	α_i		0.10, 0.15, 0.75	—	0.25,0.27,0.48
	p_i or λ_i		0.10, 0.54, 0.20	—	2.38, 1.22, 3.89

Table 6. Relative errors of the first three moments and chi-square measure for synthetically generated Traces.

		M-EM-MSNBD	EM-DPH	EM-HErD
Weibull (1.0, 0.5)	First moment	0.91%	2.15%	0.00%
	Second moment	0.79%	31.06%	7.12%
	Third moment	16.89%	72.24%	35.88%
	Chi-square	0.31	0.70	0.61
Matrix exponential	First moment	0.00%	0.00%	0.00%
	Second moment	1.39%	10.01%	1.93%
	Third moment	8.91%	30.17%	10.20%
	Chi-square	2.77	4.24	1.39
Weibull (1.0, 5.0)	First moment	0.00%	0.00%	0.00%
	Second moment	8.02%	3.54%	4.55%
	Third moment	24.96%	11.52%	14.43%
	Chi-square	0.3	0.15	0.18
Uniform (0.5, 1.5)	First moment	0.00%	0.00%	0.00%
	Second moment	4.49%	1.08%	1.40%
	Third moment	14.09%	4.31%	5.18%
	Chi-square	0.28	0.29	0.28
Pareto-II (1.5, 2.0)	First moment	12.33%	24.11%	0.00%
	Second moment	91.60%	97.66	67.78%
	Third moment	99.66%	99.98%	93.51%
	Chi-square	0.04	0.05	0.01
Shifted exponential	First moment	0.11%	0.11%	0.00%
	Second moment	0.92%	8.05%	0.76%
	Third moment	2.18%	19.37%	2.07%
	Chi-square	0.07	0.12	0.02

Table 7. The ANOVA table for continuous distributions.

Source	DF	Sum of squares	Mean square	F-value	Pr > F
Dependent variable: Y					
Model	19	42867.01607	2256.15874	26.79	< .0001
Error	33	2779.51994	84.22788		
Corrected total	52	45646.53601			
	R-square	Coeff Var	Root MSE	Objective mean	
	0.939108	51.01674	9.177575	17.98934	
Source	DF	Type I SS	Mean square	F-value	Pr > F
Method	2	1069.98712	534.99356	6.35	0.0046
Block	17	41797.02895	2458.64876	29.19	< .0001

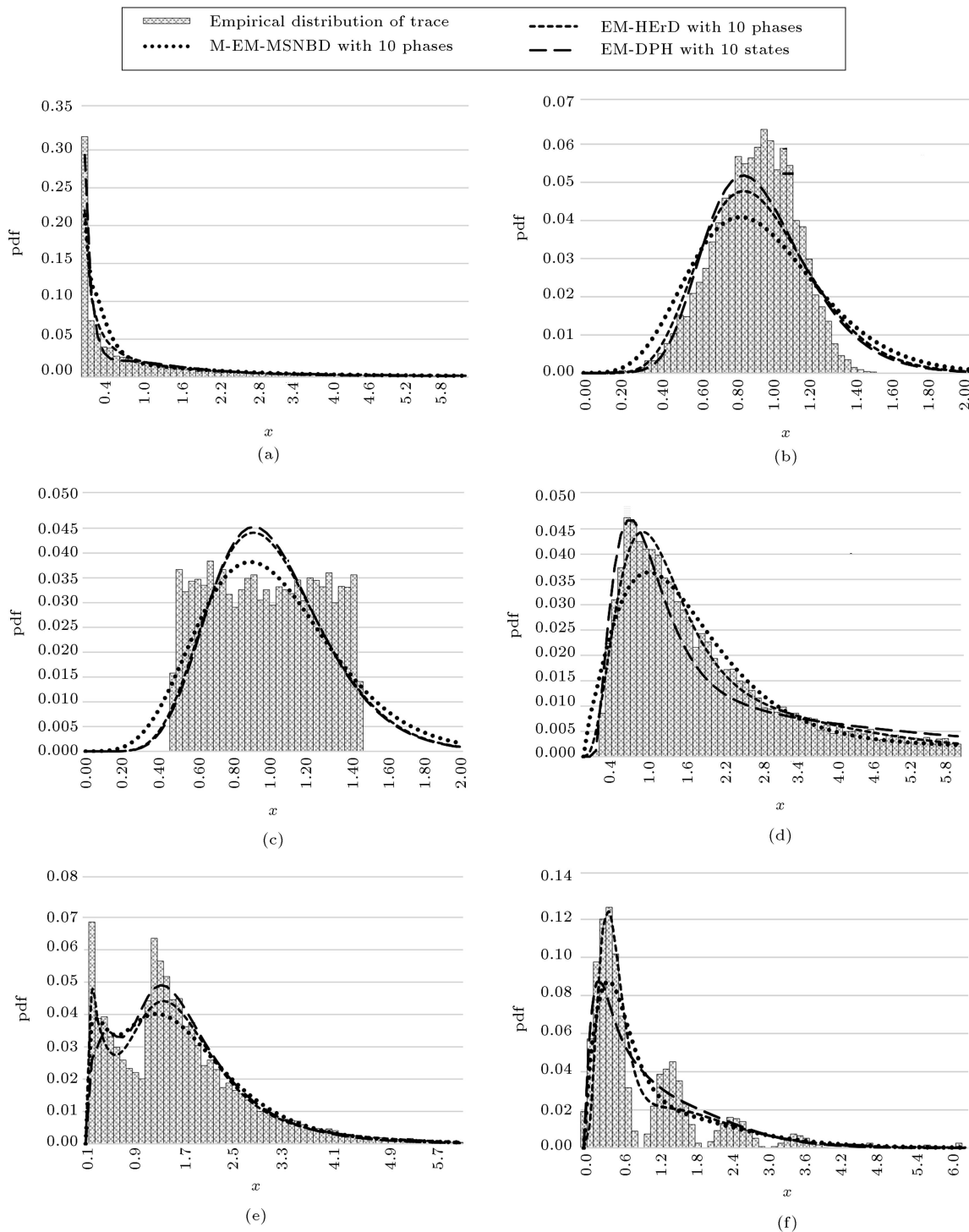


Figure 8. Densities of fitted MSNB, HErD, and DPH for synthetically generated traces: (a) Distribution of Weibull (1.0, 0.5), (b) distribution of Weibull (1.0, 5.0), (c) distribution of uniform (0.5, 1.5), (d) distribution of Pareto-II (1.5, 2.0), (e) distribution of shifted exponential, and (f) distribution of matrix exponential.

rithm is known as one of the most appropriate ones to fit continuous PH distributions, we consider our proposed (M-)EM-MSNBD algorithm as equivalently appropriate to fit discrete PH.

For further research, applying the result of the proposed EM algorithm in operating room scheduling

and developing a suitable method to solve this kind of problem are suggested.

Acknowledgment

The authors thankfully acknowledge the sharing of

Table 8. Tukey test for continuous distributions; differences of least squares means.

Algorithms comparison*	Difference between means	90% confidence limits	
2-1	7.892	[1.483 14.301]	***
2-3	10.707	[4.298 17.116]	***
1-2	-7.892	[-14.301 -1.483]	***
1-3	2.815	[-3.502 9.132]	
3-2	-10.707	[-17.116 -4.298]	***
3-1	-2.815	[-9.132 3.502]	

*1 is the M-EM-MSNBD algorithm; 2 is the EM-DPH algorithm; and 3 is the EM-HERD algorithm.

information of Scottish NHS Hospital from 1998 to 1999 by Professor John Bowers.

References

- Neuts, M.F. "Computational uses of the method of phases in the theory of queues", *Computers & Mathematics with Applications*, **1**(2), pp. 151-166 (1975).
- Neuts, M.F., *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Johns Hopkins University, Baltimore (1981).
- Fackrell, M. "Modelling healthcare systems with phase-type distributions", *Health Care Management Science*, **12**, pp. 11-26 (2009).
- Thümmler, A., Buchholz, P., and Telek, M. "A novel approach for phase-type fitting with the EM algorithm", *IEEE Transactions on Dependable and Secure Computing*, **3**(3), pp. 245-258 (2006).
- Hu, L., Jiang, Y., Zhu, J., and Chen, Y. "Hybrid of the scatter search, improved adaptive genetic, and expectation maximization algorithms for phase-type distribution fitting", *Applied Mathematics and Computation*, **219**(10), pp. 5495-5515 (2013).
- Anbazzhagan, N., *Stochastic Processes and Models in Operations Research*, Hershey, Pennsylvania 701 E. Chocolate Avenue, Hershey, PA 17033, USA: IGI Global (2016).
- Horváth, A. and Telek, M. "Phfit: A general phase-type fitting tool", *Computer Performance Evaluation: Modelling Techniques and Tools*, pp. 82-91 (2002).
- Bobbio, A., Horváth, A., Scarpa, M., and Telek, M. "Acyclic discrete phase type distributions: properties and a parameter estimation algorithm", *Performance Evaluation*, **54**(1), pp. 1-32 (2003).
- Callut, J. and Dupont, P. "Sequence discrimination using phase-type distributions", *Machine Learning: ECML 2006*, pp. 78-89 (2006).
- Asmussen, S., Nerman, O., and Olsson, M. "Fitting phase-type distributions via the EM algorithm", *Scandinavian Journal of Statistics*, **23**(4), pp. 419-441 (1996).
- Bladt, M., Esparza, L.J.R., Nielsen, B.F., et al. "Fisher information and statistical inference for phase-type distributions", *Journal of Applied Probability*, **48**, pp. 277-293 (2011).
- Meszáros, A., Papp, J., and Telek, M. "Fitting traffic traces with discrete canonical phase type distributions and Markov arrival processes", *International Journal of Applied Mathematics and Computer Science*, **24**(3), pp. 453-470 (2014).
- Horváth, I., Papp, J., and Telek, M. "On the canonical representation of order 3 discrete phase type distributions", *Electronic Notes in Theoretical Computer Science*, **318**, pp. 143-158 (2015).
- Springer, T. and Urban, K. "Comparison of the EM algorithm and alternatives", *Numerical Algorithms*, **67**(2), pp. 335-364 (2014).
- Verbelen, R., *Phase-Type Distributions & Mixtures of Erlangs*, University of Leuven (2013).
- O'Hagan, A., Murphy, T.B., and Gormley, I.C. "Computational aspects of fitting mixture models via the expectation-maximization algorithm", *Computational Statistics & Data Analysis*, **56**(12), pp. 3843-3864 (2012).
- Xu, J. and Ma, J. "Fitting finite mixture models using iterative Monte Carlo classification", *Communications in Statistics-Theory and Methods*, **46**(13), pp. 6684-6693 (2017).
- Bilmes, J.A. "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models", *International Computer Science Institute*, **4**(510), p. 126 (1998).
- McLachlan, G. and Krishnan, T., *The EM Algorithm and Extensions*, **382**, John Wiley & Sons (2007).
- Dougherty, J., Kohavi, R., and Sahami, M. "Supervised and unsupervised discretization of continuous feature", In *Proceedings of 12th International Conference of Machine Learning*, pp. 194-202 (1995).
- MATLAB Release 2014a, MathWorks, Inc. (2015).
- Bowers, J. and Mould, G. "Managing uncertainty in orthopaedic trauma theatres", *European Journal of Operational Research*, **154**(3), pp. 599-608 (2004).
- Perez, J. and Riano, G. "Benchmarking of fitting algorithms for continuous phase-type distributions", Working Paper, COPA Universidad de los Andes, pp. 1-20 (2007).
- Chiarandini, M., Basso, D., and Stützle, T. "Statistical methods for the comparison of stochastic optimizers", In *MIC2005: The Sixth Metaheuristics International Conference*, pp. 189-196 (2005).
- SAS Release, 9.1, SAS Institute Inc. (2003).
- Bobbio, A. and Telek, M. "A benchmark for PH estimation algorithms: results for Acyclic-PH", *Stochastic Models*, **10**(3), pp. 661-677 (1994).

27. Latouche, G. and Ramaswami, V., *Introduction to Matrix Analytic Methods in Stochastic Modeling*, **5**, Siam (1999).
28. Alfa, A., *Applied Discrete-time Queues*, Springer-Verlag New York (2016).
29. Varmazyar, M., Akhavan-Tabatabaei, R., Salmasi, N., and Modarres, M. "Classification and properties of acyclic discrete phase-type distributions based on geometric and shifted geometric distributions", *Journal of Industrial Engineering International* (Nov 2018).

Appendix A. Discrete phase-type distribution and its properties

Definition and notation

DPH distributions have been introduced and formalized by Neuts [1,2], 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 $\{X(n)\}_{n \geq 0}$ denotes 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 $Z \sim \text{PH}_d(\boldsymbol{\pi}, \mathbf{T})$. The one-step transition probability matrix of the corresponding DTMC can be partitioned as:

$$\mathbf{P} = \begin{bmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}, \quad (\text{A.1})$$

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 $T_{ij} \geq 0$, $t_i \geq 0 \forall i, j \in S$, and $\mathbf{T}\mathbf{1} + \mathbf{t} = \mathbf{1}$, where $\mathbf{1}$ is the column vector of 1s of the appropriate dimension n . The probability distribution of the initial 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 $Z \sim \text{PH}_d(\boldsymbol{\pi}, \mathbf{T})$ is calculated by:

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

The probability mass function is:

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

and the factorial moments are:

$$f_k = E[x(x-1)\cdots(x-k+1)] = k!\boldsymbol{\pi}(\mathbf{I}-\mathbf{T})^{-k}\mathbf{T}^{k-1}\mathbf{1} \quad \text{for } k = 1, 2, \dots \quad (\text{A.4})$$

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 the main contributing factors to the popularity of these distributions in stochastic modeling. In particular, it is shown that the DPH distributions are closed under convolution, finite mixtures, minimum, maximum, and shifted and deterministic time.

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

The basic properties of DPH distribution are presented as follows:

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

$$\boldsymbol{\pi} = (\boldsymbol{\pi}^{(1)}, \pi_0^{(1)}\boldsymbol{\pi}^{(2)}),$$

and:

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}^{(1)} & \mathbf{t}^{(1)}\boldsymbol{\pi}^{(2)} \\ \mathbf{0} & \mathbf{T}^{(2)} \end{pmatrix}. \quad (\text{A.5})$$

Proof. See Latouche and Ramaswami [27], Theorem 2.6.1.

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

$$\boldsymbol{\pi} = (\alpha\boldsymbol{\pi}^{(1)}, (1 - \alpha)\boldsymbol{\pi}^{(2)}),$$

and:

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{(2)} \end{pmatrix}. \quad (\text{A.6})$$

Proof. See Latouche and Ramaswami [27], Theorem 2.6.2.

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

$$\boldsymbol{\pi} = \boldsymbol{\pi}^{(1)} \otimes \boldsymbol{\pi}^{(2)},$$

and:

$$\mathbf{T} = \mathbf{T}^{(1)} \otimes \mathbf{T}^{(2)}, \quad (\text{A.7})$$

where \otimes is the Kronecker product.

Proof. See Latouche and Ramaswami [27], Theorem 2.6.4.

- (iv) Maximum of PH_d : The maximum $Z = \max(Z_1, Z_2) \sim \text{PH}_d(\boldsymbol{\pi}, \mathbf{T})$ has a DPH distribution of order $n = n_1.n_2 + n_1 + n_2 + 1$ with the representation shown in Box A.I.

$$\pi = (\pi^{(1)} \otimes \pi^{(2)}, \pi^{(1)} \pi_0^{(2)}, \pi_0^{(1)} \pi^{(2)}, 0), \text{ and } \mathbf{T} = \begin{pmatrix} \mathbf{T}^{(1)} \otimes \mathbf{T}^{(2)} & \mathbf{T}^{(1)} \otimes \mathbf{t}^{(2)} & \mathbf{t}^{(1)} \otimes \mathbf{T}^{(2)} & \mathbf{t}^{(1)} \otimes \mathbf{t}^{(2)} \\ 0 & \mathbf{T}^{(1)} & 0 & 0 \\ 0 & 0 & \mathbf{T}^{(2)} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (\text{A.8})$$

Box A.I

Proof. See Alfa [28], p. 40.

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

$$\pi = \pi^{(1)} \left(\mathbf{T}^{(1)} \right)^r,$$

and:

$$\mathbf{T} = \mathbf{T}^{(1)}. \quad (\text{A.9})$$

Proof. See Neuts [2], p. 47.

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

$$\pi = (\overbrace{1, 0, \dots, 0}^r),$$

and:

$$\mathbf{T} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}. \quad (\text{A.10})$$

Proof. See Neuts [2], p. 47.

Appendix B. Shifted geometric distribution and shifted negative binomial distribution

Shifted geometric distribution ($X \sim \text{SG}(p)$, with $p \in (0, 1)$) is another nonequivalent definition of geometric distribution, 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 $\Pr(X = x) = (1 - p)^x p$, for $x = 0, 1, 2, \dots$. The DPH representation of the shifted geometric distribution is given by Eq. (B.1) and presented in Figure B.1.

$$\pi_{\text{SG}} = [1 - p], \quad \mathbf{T}_{\text{SG}} = [1 - p], \quad \mathbf{t}_{\text{SG}} = [p]. \quad (\text{B.1})$$

Shifted negative binomial distribution ($X \sim \text{SNB}(n, p)$) 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 $\text{SG}(p)$ -distributed; thus:

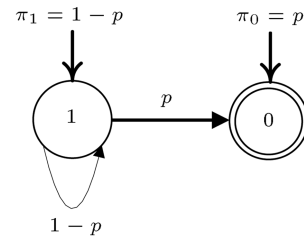


Figure B.1. The DPH representation of $\text{SG}(p)$.

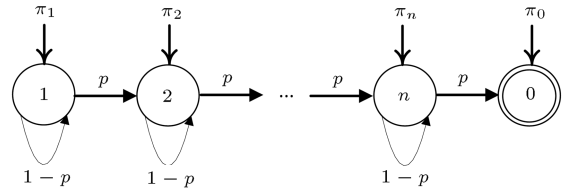


Figure B.2. The DPH representation of $\text{SNB}(n, p)$.

$$\Pr(X = x) = \binom{x+n-1}{n-1} (1-p)^x p^n, \text{ for}$$

$$x = 0, 1, \dots$$

The DPH and diagrammatic representation of shifted negative binomial distribution are presented in Eq. (B.2) and Figure B.2, respectively.

$$\pi_{\text{SNB}} = (\pi_1, \pi_2, \dots, \pi_n),$$

$$\pi_j = \binom{n}{j-1} (1-p)^{n-(j-1)} p^{j-1},$$

$$\mathbf{T}_{\text{SNB}} = \begin{pmatrix} 1-p & p & 0 & 0 & 0 & 0 \\ 0 & 1-p & p & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p & p \\ 0 & 0 & 0 & 0 & 0 & 1-p \end{pmatrix},$$

$$\mathbf{t}_{\text{SNB}} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ p \end{pmatrix}. \quad (\text{B.2})$$

For more information about geometric and shifted geometric distributions, please see Varmazyar et al. [29].

Biographies

Mohsen Varmazyar is a PhD candidate in Industrial Engineering at Sharif University of Technology (SUT) in IRAN. He holds MSc in Industrial Engineering from SUT. His research interests focus on applied operations research, scheduling, and stochastic processes.

Raha Akhavan-Tabatabaei is Associate Professor of Operations Management and co-director of Masters in Business Analytics at Sabanci School of Management. Prior to this position, she was Associate Professor of Industrial Engineering and the founding director of Masters in Analytics at Universidad de los Andes in Bogota, Colombia. Before that, she worked as senior industrial engineer at Intel Corporation in Arizona, USA. She has received her PhD and MSc in Industrial Engineering and Operations Research from North Carolina State University, and her BSc from Sharif University of Technology. Her research is focused on stochastic modeling and data-driven decision making

with applications in healthcare, logistics, revenue management, and reliability.

Nasser Salmasi received his PhD in the area of Industrial Engineering from Oregon State University, USA. He has been working in the Department of Industrial Engineering at Sharif University of Technology, Tehran, Iran, as an Associate Professor since nine years ago. Since October 2015, he has joined Corning Incorporated in USA as an operations research analyst. His primary areas of research interests are applied operations research, scheduling, and simulation.

Mohammad Modarres is a Professor in the Department of Industrial Engineering, Sharif University of Technology, Iran. He received his PhD in Systems Engineering and Operations Research from University of California, Los Angeles (UCLA), in 1975. His research interests are operations research, revenue management, and robust optimization.