Methodol Comput Appl Probab
DOI 10.1007/s11009-015-9442-y

Parameter Estimation of Discrete Multivariate Phase-Type Distributions

Qi-Ming He · Jiandong Ren

Received: 20 July 2014 / Revised: 10 January 2015 / Accepted: 17 February 2015

Abstract This paper considers parameter estimation of a class of discrete multi-variate phase-type distributions (DMPH). Such discrete phase-type distributions are based on discrete Markov chains with marked transitions introduced by He and Neuts (Stoch Process Appl 74(1):37–52, 1998) and is a generalization of the discrete univariate phase–type distributions. Properties of the DMPHs are presented. An EM-algorithm is developed for estimating the parameters for DMPHs. A number of numerical examples are provided to address some interesting parameter selection issues and to show possible applications of DMPHs.

Keywords Risk analysis · Parameter estimation · PH-distributions

Mathematics Subject Classification (2010) Primary: 60-08 · Secondary: 60J22

1 Introduction

Continuous multivariate phase-type distribution was introduced in Assaf et al. (1984) and Kulkarni (1989) and have found applications in many areas. For example, Cai and Li (2005) applied it to model different types of claims in insurance risk; Herbertsson (2011) used it to model default contagion in credit risk. He and Ren (2014) introduced discrete multivariate phase-type (DMPH) distribution that is based on the marked Markovian arrival processes

Q.-M. He (✉)
Department of Management Sciences, University of Waterloo,
Waterloo, ON Canada
e-mail: q7he@uwaterloo.ca

J. Ren
Department of Statistical and Actuarial Sciences, University of Western Ontario,
London, ON Canada

Published online: 03 March 2015

introduced and studied in He and Neuts (1998). The DMPH distribution bears resemblance to the continuous MPH-distribution introduced in Kulkarni (1989), in the sense that the "rewards" associated with a transition in the Markov chain J defined in Kulkarni (1989) are distributed to different type of event batches.

This paper focuses on the parameter estimation methods for the DMPH introduced in He and Ren (2014). The problem is related to the parameter estimation in Markovian arrival process and phase–type distribution and the latter has been studied in the literature for a long time. For example, Asmussen et al. (1996) introduced an expectation–maximization (EM) algorithm for fitting phase–type distributions; Olsson (1996) extended the methods to the case where observations are censored. Breuer (2003) introduced an EM-algorithm for fitting parameters of a Markovian arrival process with batch arrivals. Buchholz et al. (2010) presented a method of fitting a Markovian arrival process with marked events according to moments and joint moments. Hassan Zadeh and Bilodeau (2013) used bivariate phase–type distributions to fit insurance loss data.

The main contribution of the current paper is that we present an EM-algorithm for estimating parameters of the DMPH distribution, when the exact values of the data is known or when the data is censored. Issues related to parameter selection and parameters estimation for some well-known special cases are addressed numerically.

The reminder of the paper is organized as follows. In Section 2, the DMPHs are defined and some basic properties are given. Section 3 provides a number of quantities and their properties that are utilized in the EM-algorithm developed in Section 4. Numerical examples are presented in Section 5 to explore a few issues related to parameter selection and to show the application of DMPHs. Section 6 concludes the paper.

2 Preliminary: Definition and Basic Properties

The DMPH is based on a discrete Markov chains with marked transitions proposed by He and Neuts (1998). First, we introduce a standard matrix-representation for the DMPH. Let $\{J(t)\}_{t=0,1,\cdots}$ be a discrete time Markov chain with finite state space $\{0, 1, \cdots, m\}$ and transition probability matrix

$$\begin{pmatrix} 1 & 0 \\ \mathbf{b}^0 & B \end{pmatrix}, \tag{1}$$

where $\mathbf{b}^0$ is an $m \times 1$ vector and B is an $m \times m$ matrix. The state 0 is assumed to be absorbing, and other states are transient. A state transition between the transient states may be accompanied by the occurrence of a batch of events (to be called an event batch). Each event in the batch may belong to one of K categories of events of interest (i.e, $\{1, 2, ..., K\}$). Let C^0 be the collection of all possible event batches. An element in C^0 can be denoted as $\mathbf{h} = (h_1, h_2, \cdots, h_K)$, where h_k represents the number of type k events in $\mathbf{h}$. According to the possible occurrence of the event batches, transition matrix B can be decomposed into $\{B_0, B_{\mathbf{h}}, \mathbf{h} \in C^0\}$ (i.e., $B = B_0 + \sum_{\mathbf{h}:\, \mathbf{h} \in C^0} B_{\mathbf{h}}$), where B_0 is a sub-stochastic matrix giving transition probabilities between transient states without any accompanying event, and $B_{\mathbf{h}}$, for $\mathbf{h} \in C^0$, is a sub-stochastic matrix giving transition probabilities between transient states with event batch $\mathbf{h}$. The vector $\mathbf{b}^0$ contains the transition probabilities from transient states to the absorbing state 0. Note that, by definition, $\mathbf{b}^0 = \mathbf{e} - (B_0 + \sum_{\mathbf{h} \in C^0} B_{\mathbf{h}})\mathbf{e}$, where $\mathbf{e}$ is a column vector of ones. We assumed that the transition from a transient state to the absorbing state is not accompanied by an occurrence of an event batch. However, as discussed in Note 1, this is not a big constraint.

Define $X_{\mathbf{h}}$ to be the number of type $\mathbf{h}$ event batches occurred before the underlying Markov chain is absorbed into state 0. Then $\mathbf{X} = \{X_{\mathbf{h}}, \mathbf{h} \in C^0\}$ forms a vector of non–negative discrete random variables. For $1 \le k \le K$, let Y_k be the total number of type k events occurred before the absorption into state 0, then Y_k can be expressed in $\{X_{\mathbf{h}}, \mathbf{h} \in C^0\}$ as $Y_k = \sum_{\mathbf{h} \in C^0} h_k X_{\mathbf{h}}$. The vector of random variables $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_K)$ is also an object of interest.

Because of its obvious relationship with phase–type random variables, we say that the vectors $\mathbf{X}$ and $\mathbf{Y}$ follows discrete multivaiate phase–type (DMPH) distributions. In He and Ren (2014), the order of the events within each event batch $\mathbf{h}$ is considered, which can be useful for some applications. For simplicity, in this paper, we ignore the order of individual events in a batch and then count the total number of events of each type in $\mathbf{h}$. All results obtained in this paper can be extended to the case in which the order of events in each batch is considered, though.

Assume that $J(0)$ has a distribution $(\beta_0, \boldsymbol{\beta})$. Then the distribution of the DMPH $\mathbf{X}$ (or $\mathbf{Y}$) is well-defined. Without loss of generality, we assume that $\beta_0 = 0$. A standard matrix-representation of the DMPH is given by $\{\boldsymbol{\beta}, B_0, B_{\mathbf{h}}, \mathbf{h} \in C^0, \mathbf{b}^0\}$. As mentioned, the objective of this paper is to estimate the matrix-representation of the DMPH, based on a sample of $\mathbf{Y}$ or $\mathbf{X}$.

Note 1 If the transitions from transient states to the absorbing state 0 may also trigger events, the vector $\mathbf{b}^0$ is decomposed into $\{\mathbf{b}^0_0,\ \mathbf{b}^0_{\mathbf{h}}, \mathbf{h} \in C^0\}$, where column vector $\mathbf{b}^0_{\mathbf{h}}$ contains the transition probabilities from transient states to the absorbing state with event batch $\mathbf{h}$. Then the DMPH has a matrix-representation $\{\boldsymbol{\beta}, B_0, B_{\mathbf{h}}, \mathbf{b}^0_{\mathbf{h}}, \mathbf{h} \in C^0, \mathbf{b}^0_0\}$. We can modify the underlying Markov chain to transfer the matrix-representation of the DMPH into the standard form. For transitions governed by $\{\mathbf{b}^0_{\mathbf{h}}, \mathbf{h} \in C^0\}$, we introduce a fictitious state (e.g., $m + 1$) that the underlying Markov chain will only go from it to the absorbing state. Then the matrix-representation of DMPH $\mathbf{X}$ has an equivalent matrix-representation of the standard form given as follows:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{\beta}, 0),\ \hat{B}_0 = \begin{pmatrix} B_0 & 0 \\ 0 & 0 \end{pmatrix},\ \hat{B}_{\mathbf{h}} = \begin{pmatrix} B_{\mathbf{h}} & \mathbf{b}^0_{\mathbf{h}} \\ 0 & 0 \end{pmatrix},\ \hat{\mathbf{b}}^0 = \begin{pmatrix} \mathbf{b}^0_0 \\ 1 \end{pmatrix}. \tag{2}$$

Next, we present formulas for joint distributions and moments of random vectors $\mathbf{X}$ and $\mathbf{Y}$. We start with the analysis of $\mathbf{X}$. Let $\mathbf{x} = \{x_{\mathbf{h}}, \mathbf{h} \in C^0\}$, a vector with nonnegative integer elements. Define

$$\begin{aligned} &p_{\mathbf{X},i}(\mathbf{x}) = \mathbb{P}\{X_{\mathbf{h}} = x_{\mathbf{h}}, \mathbf{h} \in C^0 | J(0) = i\},\ \text{for } i = 1, 2, ..., m; \\ &\mu_{X_{\mathbf{h}},i} = \mathbb{E}[X_{\mathbf{h}} | J(0) = i],\ \text{for } i = 1, 2, ..., m. \end{aligned} \tag{3}$$

Let $\mathbf{p_X}(\mathbf{x}) = (p_{\mathbf{X},1}(\mathbf{x}), p_{\mathbf{X},2}(\mathbf{x}), \cdots, p_{\mathbf{X},m}(\mathbf{x}))$, and $\boldsymbol{\mu}_{X_{\mathbf{h}}} = (\mu_{X_{\mathbf{h}},1}, \mu_{X_{\mathbf{h}},2}, \cdots, \mu_{X_{\mathbf{h}},m})$, In He and Ren (2014), it has been shown that

$$\begin{aligned} &\mathbf{p_X}(0) = (I - B_0)^{-1}\mathbf{b}^0; \\ &\mathbf{p_X}(\mathbf{x}) = \sum_{\mathbf{h} \in C^0:\ x_{\mathbf{h}} > 0} (I - B_0)^{-1} B_{\mathbf{h}} \mathbf{p_X}(\{x_{\mathbf{h}} - 1, x_{\mathbf{u}}, \mathbf{u} \ne \mathbf{h}, \mathbf{u} \in C^0\}),\ \text{if } \mathbf{x} \ne 0. \end{aligned} \tag{4}$$

The first moments and joint second moments of random variables $\{X_{\mathbf{h}},\ \mathbf{h} \in C^0\}$ are given by

$$\begin{aligned} &\boldsymbol{\mu}_{X_{\mathbf{h}}} = (I - B)^{-1} B_{\mathbf{h}} \mathbf{e}; \\ &\{\mathbb{E}[X_{\mathbf{h}}(X_{\mathbf{h}} - 1) | J(0) = i]\}_{i=1\cdots,m} = 2(I - B)^{-1} B_{\mathbf{h}} \boldsymbol{\mu}_{X_{\mathbf{h}}}; \\ &\{\mathbb{E}[X_{\mathbf{h}} X_{\mathbf{u}} | J(0) = i]\}_{i=1\cdots,m} = (I - B)^{-1}(B_{\mathbf{h}} \boldsymbol{\mu}_{X_{\mathbf{u}}} + B_{\mathbf{u}} \boldsymbol{\mu}_{X_{\mathbf{h}}}), \end{aligned} \tag{5}$$

for $\mathbf{h}$ and $\mathbf{u}$ in C^0.

For the random vector $\mathbf{Y}$, let $\mathbf{y} = (y_1, y_2, \cdots, y_K)$, and

$$p_{\mathbf{Y},i}(\mathbf{y}) = \mathbb{P}\{Y_k = y_k, k = 1, \cdots, K | J(0) = i\}, \text{ for } i = 1, 2, ..., m. \tag{6}$$

Let $\mathbf{p}_{\mathbf{Y}}(\mathbf{y}) = (p_{\mathbf{Y},1}(\mathbf{y}), p_{\mathbf{Y},2}(\mathbf{y}), \cdots, p_{\mathbf{Y},m}(\mathbf{y}))'$. The following results have been shown in He and Ren (2014). The distribution function of $\mathbf{Y}$ is given by

$$\begin{aligned} &\mathbf{p}_{\mathbf{Y}}(0) = (I - B_0)^{-1}\mathbf{b}^0; \\ &\mathbf{p}_{\mathbf{Y}}(\mathbf{y}) = \textstyle\sum_{\mathbf{h}\in C^0:\, \mathbf{u}\le\mathbf{y}} (I - B_0)^{-1} B_{\mathbf{h}\mathbf{p}_{\mathbf{Y}}}(\mathbf{y} - \mathbf{h}), \text{ if } \mathbf{y} \neq 0. \end{aligned} \tag{7}$$

Note that, for $\mathbf{y} \neq 0$, if $\{\mathbf{h} \in C^0 : \mathbf{h} \le \mathbf{y}\} = \phi$ (i.e., an empty set), then $\mathbf{p}_{\mathbf{Y}}(\mathbf{y}) = 0$.

Recall that $Y_k = \sum_{\mathbf{h}\in C^0} h_k X_{\mathbf{h}}$. Then the moments of $\mathbf{Y}$ can be obtained from those of $\{X_{\mathbf{h}}, \mathbf{h} \in C^0\}$ as follows: for $i = 1, \cdots, m$,

$$\begin{aligned} &\mathbb{E}[Y_k | J(0) = i] = \textstyle\sum_{\mathbf{h}\in C^0} h_k \mathbb{E}[X_{\mathbf{h}} | J(0) = i], \text{ for } k = 1, \cdots, K; \\ &\mathbb{E}[Y_{k_1} Y_{k_2} | J(0) = i] = \textstyle\sum_{\mathbf{h},\mathbf{u}\in C^0} u_{k_1} h_{k_2} \mathbb{E}[X_{\mathbf{u}} X_{\mathbf{h}} | J(0) = i], \text{ for } k_1, k_2 = 1, \cdots, K. \end{aligned} \tag{8}$$

As discussed in He and Ren (2014), DMPH allows positive or negative correlations between the pairs of random variables. In the rest of this section, we present some special cases of DMPHs.

Example 2.1 Multi-variate geometric distribution. The simplest form of the DMPH distribution is a multi-variate geometric distribution, which is a special case of the bivariate negative binomial distribution studied in Subrahmaniam and Subrahmaniam (1973). Let $K = 2$, $C^0 = \{(1, 0), (0, 1), (1, 1)\}$, $m = 1$, $\boldsymbol{\beta} = 1$, $B_{(1,0)} = p_{(1,0)}$, $B_{(0,1)} = p_{(0,1)}$, $B_{(1,1)} = p_{(1,1)}$, $B_0 = 0$, and $\mathbf{b}^0 = 1 - p_{(1,0)} - p_{(0,1)} - p_{(1,1)}$. Then the random vector $(X_{(1,0)}, X_{(0,1)}, X_{(1,1)})$ follows a tri-variate geometric distribution. It is easy to see that a *multi-variate negative binomial distribution* can be easily constructed by including more transient states.

Example 2.2 This example considers a special case of the general model, where we assume that $B_0 = w_0 B$, and $B_{\mathbf{h}} = w_{\mathbf{h}} B$ with $w_0 + \sum_{\mathbf{h}\in C^0} w_{\mathbf{h}} = 1$. This example greatly reduces the number of parameters in the general model, while it still covers some interesting cases.

- Let $N = X + 1$ follow a discrete phase–type distribution with representation $(\boldsymbol{\beta}, B)$. Conditioning on $X = n$, $X_{(1,0)}$ follows a binomial distribution with parameters (n, ρ) and $X_{(0,1)}$ follows a binomial distribution with parameters $(n, 1 - \rho)$. Then it is easy to show that (see for example Ren 2010) the distribution of $(X_{(1,0)}, X_{(0,1)})$ is a special case of the DMPH distribution with representation $\{K = 2, C^0 = \{(1, 0), (0, 1)\}, \boldsymbol{\beta}, B_0 = 0, B_{(1,0)} = \rho B, B_{(0,1)} = (1 - \rho)B, \mathbf{b}^0 = (I - B)\mathbf{e}\}$.
- Let Θ be a continuous phase–type random variable having PH-representation $(\boldsymbol{\beta}, T)$, where $\boldsymbol{\beta}$ is a stochastic vector and T is a PH-generator.. Assume that conditioning on $\Theta = \theta$, the number of type 1 events $X_{(1,0,0)}$ follows a Poisson distribution with mean $\lambda_1\theta$; the number of type 2 events $X_{(0,1,0)}$ follows a Poisson distribution with mean $\lambda_2\theta$; and the number of type 3 claim batch $X_{(0,0,1)}$ follows a Poisson distribution with mean $\lambda_3\theta$. Then it can be shown that $(X_{(1,0,0)}, X_{(0,1,0)}, X_{(0,0,1)})$ has a DMPH with representation $\{K = 3, C^0 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}, \boldsymbol{\beta}, B_0 = 0, B_{(1,0,0)} = \lambda_1 B/\lambda, B_{(0,1,0)} = \lambda_2 B/\lambda, B_{(0,0,1)} = \lambda_3 B/\lambda, \mathbf{b}^0 = (I - B)\mathbf{e}\}$, where $\lambda = \lambda_1 + \lambda_2 + \lambda_3$ and $B = (I - T/\lambda)^{-1}$. Note that B is substochastic since i) $T/\lambda - I$ is a PH-generator, which implies that $B = -(T/\lambda - I)^{-1}$ is nonnegative; and ii) $\mathbf{e} = B\mathbf{e} + B(-T\mathbf{e})/\lambda \ge B\mathbf{e}$.

3 Conditional Expectations of DMPH Distributions

For the development of the EM-algorithm to be presented in Section 4, we introduce a few random variables and find their conditional expectations. Let $Y_k(t)$ be the total number of type k events occurred in $(0, t]$ and $\mathbf{Y}(t) = (Y_1(t), Y_2(t), ..., Y_K(t))$. Define $\tau = \min\{t : J(t) = 0\}$. It is easy to see that $\mathbf{Y} = \mathbf{Y}(\tau)$. Let $\mathbf{h}(t)$ be the event batch accompanying the transition from period t to $t + 1$; $\mathbf{h}(t) = 0$ if there is no event occuring in period t. For $i, j = 1, \cdots, m$, define

- $D_i = I(J(0) = i)$, where $I(\cdot)$ is the indicator function;
- $N_i = \sum_{t=0}^{\tau-1} I(J(t) = i)$: the total number of times $J(t)$ enters state i before the absorption into state 0;
- $N_{(i,j),0} = \sum_{t=0}^{\tau-1} I(J(t) = i, J(t+1) = j, \mathbf{h}(t) = 0)$: the total number of $i \to j$ transitions without an event before the absorption into state 0;
- $N_{(i,j),\mathbf{h}} = \sum_{t=0}^{\tau-1} I(J(t) = i, J(t+1) = j, \mathbf{h}(t) = \mathbf{h})$: the total number of $i \to j$ transitions accompanied by event batch $\mathbf{h}$ before the absorption into state 0;
- $N_{i,0} = I(J(\tau - 1) = i)$.

Apparently, we must have $N_i = N_{i,0} + \sum_{j=1}^{m} \left(N_{(i,j),0} + \sum_{\mathbf{h}:\, \mathbf{h}\in C^0} N_{(i,j),\mathbf{h}}\right)$. The objective of this section is to find some conditional expectations of those random variables. For that purpose, we first define vector set $\{\boldsymbol{\alpha}(\mathbf{y}), \mathbf{y} \geq 0\}$.

Define, for $i = 1, 2, ..., m$,

$$\alpha_i(\mathbf{y}) = \sum_{t=0}^{\infty} \mathbb{P}\{\mathbf{Y}(t) = \mathbf{y}, J(t) = i\}, \tag{9}$$

i.e., the expected number of periods such that, the numbers of occurred events are $\mathbf{y}$, and $J(t) = i$ in those periods, given that $J(0)$ has distribution $(0, \boldsymbol{\beta})$. Let $\boldsymbol{\alpha}(\mathbf{y}) = (\alpha_1(\mathbf{y}), \alpha_2(\mathbf{y}), ..., \alpha_m(\mathbf{y}))$. Then vectors $\{\boldsymbol{\alpha}(\mathbf{y}),\ \mathbf{y} \geq 0\}$ can be obtained recursively as follows:

$$\begin{aligned} &\boldsymbol{\alpha}(0) = \boldsymbol{\beta}(I - B_0)^{-1}; \\ &\boldsymbol{\alpha}(\mathbf{y}) = \sum_{\mathbf{h}\in C^0:\, \mathbf{h}\leq \mathbf{y}} \boldsymbol{\alpha}(\mathbf{y}-\mathbf{h}) B_{\mathbf{h}} (I - B_0)^{-1}, \text{ if } \mathbf{y} \neq 0. \end{aligned} \tag{10}$$

Note that, for $\mathbf{y} \neq 0$, if $\{\mathbf{h} \in C^0 : \mathbf{h} \leq \mathbf{y}\} = \phi$, then $\boldsymbol{\alpha}(\mathbf{y}) = 0$. The term $(I - B_0)^{-1}$ is due to the fact that between two consecutive event batches, there can be any number of transitions without an event. It is easy to see that $\mathbb{P}\{\mathbf{Y} = \mathbf{y}\} = \boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}(\mathbf{y}) = \boldsymbol{\alpha}(\mathbf{y})\mathbf{b}^0$.

Conditional expectations of $\{D_i, N_i, N_{(i,j),0}, N_{(i,j),\mathbf{h}}, \mathbf{h} \in C^0, N_{i,0}\}$ can be expressed in terms of $\{\boldsymbol{\alpha}(\mathbf{y}),\ \mathbf{y} \geq 0\}$ and $\{\mathbf{p}_{\mathbf{Y}}(\mathbf{y}),\ \mathbf{y} \geq 0\}$.

Proposition 3.1 *Given* $\mathbf{Y} = \mathbf{y}$*, we have, for* $i, j = 1, 2, ..., m$*, and* $\mathbf{h} \in C^0$*,*

$$\begin{aligned} &\mathbb{E}[D_i|\mathbf{Y} = \mathbf{y}] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}(\mathbf{y})} \beta_i p_{\mathbf{Y},i}(\mathbf{y}); \\ &\mathbb{E}[N_i|\mathbf{Y} = \mathbf{y}] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}(\mathbf{y})} \sum_{\mathbf{u}:\, \mathbf{u}\leq\mathbf{y}} \alpha_i(\mathbf{u}) p_{\mathbf{Y},i}(\mathbf{y}-\mathbf{u}); \\ &\mathbb{E}[N_{(i,j),0}|\mathbf{Y} = \mathbf{y}] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}(\mathbf{y})} \sum_{\mathbf{u}:\, \mathbf{u}\leq\mathbf{y}} \alpha_i(\mathbf{u})(B_0)_{(i,j)} p_{\mathbf{Y},j}(\mathbf{y}-\mathbf{u}); \\ &\mathbb{E}[N_{(i,j),\mathbf{h}}|\mathbf{Y} = \mathbf{y}] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}(\mathbf{y})} \sum_{\mathbf{u}:\, \mathbf{u}\leq\mathbf{y}-\mathbf{h}} \alpha_i(\mathbf{u})(B_{\mathbf{h}})_{(i,j)} p_{\mathbf{Y},j}(\mathbf{y}-\mathbf{u}-\mathbf{h}); \\ &\mathbb{E}[N_{i,0}|\mathbf{Y} = \mathbf{y}] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}(\mathbf{y})} \alpha_i(\mathbf{y})(\mathbf{b}^0)_i. \end{aligned} \tag{11}$$

Proof The proof of the proposition is analogous to the continuous case considered in Asmussen et al. (1996). First, the expression for $\mathbb{E}[D_i|\mathbf{Y} = \mathbf{y}]$ is obtained directly by definition as follows:

$$\mathbb{E}[D_i|\mathbf{Y}=\mathbf{y}] = \mathbb{P}\{D_i = 1|\mathbf{Y}=\mathbf{y}\} = \frac{\mathbb{P}\{J(0)=i,\mathbf{Y}=\mathbf{y}\}}{\mathbb{P}\{\mathbf{Y}=\mathbf{y}\}} = \frac{\mathbb{P}\{J(0)=i\}\mathbb{P}\{\mathbf{Y}=\mathbf{y}|J(0)=i\}}{\mathbb{P}\{\mathbf{Y}=\mathbf{y}\}}. \tag{12}$$

To find the conditional expectation of N_i, we first note that $N_i = \sum_{t=0}^{\infty} I(J(t) = i)$ and $\mathbb{E}[I(J(t) = i)] = \mathbb{P}\{J(t) = i\}$. Then we have

$$\begin{aligned}
&\mathbb{E}[N_i|\mathbf{Y}=\mathbf{y}]\\
&= \frac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}(\mathbf{y})}\sum_{t=0}^{\infty}\mathbb{P}\{J(t)=i,\mathbf{Y}=\mathbf{y}\}\\
&= \frac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}(\mathbf{y})}\sum_{t=0}^{\infty}\sum_{\mathbf{u}:\ \mathbf{u}\le\mathbf{y}}\mathbb{P}\{\mathbf{Y}(t)=\mathbf{u},\ J(t)=i,\mathbf{Y}=\mathbf{y}\}\\
&= \frac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}(\mathbf{y})}\sum_{t=0}^{\infty}\sum_{\mathbf{u}:\ \mathbf{u}\le\mathbf{y}}\mathbb{P}\{\mathbf{Y}(t)=\mathbf{u},\ J(t)=i\}\mathbb{P}\{\mathbf{Y}=\mathbf{y}|\mathbf{Y}(t)=\mathbf{u},\ J(t)=i\}\\
&= \frac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}(\mathbf{y})}\sum_{\mathbf{u}:\ \mathbf{u}\le\mathbf{y}}\sum_{t=0}^{\infty}\mathbb{P}\{\mathbf{Y}(t)=\mathbf{u},\ J(t)=i\}\mathbb{P}\{\mathbf{Y}=\mathbf{y}-\mathbf{u}|J(0)=i\}\\
&= \frac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}(\mathbf{y})}\sum_{\mathbf{u}:\ \mathbf{u}\le\mathbf{y}}\alpha_i(\mathbf{u})p_{\mathbf{Y},i}(\mathbf{y}-\mathbf{u}).
\end{aligned} \tag{13}$$

For $N_{(i,j),0}$, $N_{(i,j),\mathbf{h}}$, and $N_{i,0}$, we have $N_{(i,j),0} = \sum_{t=0}^{\infty} I(\mathbf{h}(t) = 0, J(t+1) = j|J(t) = i)$, $N_{(i,j),\mathbf{h}} = \sum_{t=0}^{\infty} I(\mathbf{h}(t) = \mathbf{h}, J(t+1) = j|J(t) = i)$, for $\mathbf{h} \in C^0$, and $N_{i,0} = \sum_{t=0}^{\infty} I(J(t) = i, J(t+1) = 0)$. Then the conditional expectations of the random variables can be obtained similarly. Details are omitted. This completes the proof of Proposition 3.1. □

Using the recursive formulas in Eqs. 7 and 10, it can be shown that $\sum_{\mathbf{u}\ge 0}\boldsymbol{\alpha}(\mathbf{u}) = \boldsymbol{\beta}(I-B)^{-1}$ and $\sum_{\mathbf{u}\ge 0}\mathbf{p}_{\mathbf{Y}}(\mathbf{u}) = \mathbf{e}$. By Proposition 3.1, the unconditional expectations of D_i, $N_{(i,j),0}$, $N_{(i,j),\mathbf{h}}$, and $N_{i,0}$, can be obtained.

Corollary 3.2 *For the DMPH* $\mathbf{X}$*, we have*

$$\begin{aligned}
&\{\mathbb{E}[D_i]\}_{i=1,2,\ldots,m} = \boldsymbol{\beta};\\
&\{\mathbb{E}[N_i]\}_{i=1,2,\ldots,m} = \boldsymbol{\beta}(I-B)^{-1};\\
&\{\mathbb{E}[N_{(i,j),0}]\}_{i,j=1,2,\ldots,m} = diag(\boldsymbol{\beta}(I-B)^{-1})B_0;\\
&\{\mathbb{E}[N_{(i,j),\mathbf{h}}]\}_{i,j=1,2,\ldots,m} = diag(\boldsymbol{\beta}(I-B)^{-1})B_{\mathbf{h}},\ for\ \mathbf{h}\in C^0;\\
&\{\mathbb{E}[N_{i,0}]\}_{i=1,2,\ldots,m} = diag(\boldsymbol{\beta}(I-B)^{-1})\mathbf{b}^0,
\end{aligned} \tag{14}$$

where $diag(\boldsymbol{\beta}(I-B)^{-1})$ *is a diagonal matrix with the elements of the vector* $\boldsymbol{\beta}(I-B)^{-1}$ *on its diagonal.*

Note 2 For the DMPH $\mathbf{X}$ considered in Proposition 3.1, the number of periods without a batch event is not recorded. If the number of periods without an event is recorded (i.e., every period is accompanied by an event batch), Proposition 3.1 can be modified as follows. We introduce a fictitous event type $K+1$, set $B_{K+1} = B_0$ (and then set $B_0 = 0$), and define $X_{K+1} = Y_{K+1}$ as the total number of such events. Then all results in Proposition 3.1 hold for the modified DMPH (with $B_0 = 0$).

Note 3 Expectations of $\{D_i, N_{(i,j),0}, N_{(i,j),\mathbf{h}}, \mathbf{h} \in C^0, N_{i,0}\}$, conditioning on $\mathbf{X}$, instead of $\mathbf{Y}$, can be found similarly. In fact, we can treat each event batch $\mathbf{h}$ as a *super* event. Then Proposition 3.1 holds for the super events.

Next, we generalize Proposition 3.1 to the case when only partial information of some types of events are observed. To this end, we divide the type of events into two groups: $\mathcal{K}_1 = \{1, \cdots, K_1\}$ and $\mathcal{K}_2 = \{K_1+1, \cdots, K\}$ and rewrite $\mathbf{Y}(t)$ as $\mathbf{Y}(t) = \left(\mathbf{Y}^{(1)}(t), \mathbf{Y}^{(2)}(t)\right)$, where $\mathbf{Y}^{(1)}(t) = (Y_1(t), ..., Y_{K_1}(t))$ and $\mathbf{Y}^{(2)}(t) = (Y_{K_1+1}(t), ..., Y_K(t))$. We want to find the expectations in Proposition 3.1 under the following two conditions: i) $\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)},\ \mathbf{Y}^{(2)} \le \mathbf{y}^{(2)}\}$ and ii) $\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)},\ \mathbf{Y}^{(2)} \ge \mathbf{y}^{(2)}\}$.

For the case with condition $\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)},\ \mathbf{Y}^{(2)} \le \mathbf{y}^{(2)}\}$, we generalize $\boldsymbol{\alpha}(\mathbf{y})$ and $\mathbf{p}_{\mathbf{Y}}(\mathbf{y})$ as follows. Define, for $i = 1, 2, ..., m$,

$$\begin{aligned}\alpha_{\le,i}(\mathbf{y}) &= \sum_{t=0}^{\infty} \mathbb{P}\{\mathbf{Y}^{(1)}(t) = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)}(t) \le \mathbf{y}^{(2)}, J(t) = i\} \\ &= \sum_{\mathbf{u}^{(2)}:\ \mathbf{u}^{(2)} \le \mathbf{y}^{(2)}} \alpha_i((\mathbf{y}^{(1)}, \mathbf{u}^{(2)})),\end{aligned} \tag{15}$$

and $\boldsymbol{\alpha}_{\le}(\mathbf{y}) = (\alpha_{\le,1}(\mathbf{y}), ..., \alpha_{\le,m}(\mathbf{y}))$. Then we have that

$$\begin{aligned}&\boldsymbol{\alpha}_{\le}(0) = \boldsymbol{\beta}(I - B_0)^{-1}, \ \text{ if } \mathbf{y} = 0; \\ &\boldsymbol{\alpha}_{\le}((0, \mathbf{y}^{(2)})) = \boldsymbol{\beta}(I - B_0)^{-1} \\ &\qquad + \sum_{\mathbf{h} \in C^0:\ \mathbf{h} \le \mathbf{y}} \boldsymbol{\alpha}_{\le}(\mathbf{y} - \mathbf{h}) B_{\mathbf{h}} (I - B_0)^{-1}, \ \text{ if } \mathbf{y}^{(2)} \ne 0; \\ &\boldsymbol{\alpha}_{\le}(\mathbf{y}) = \sum_{\mathbf{h} \in C^0:\ \mathbf{h} \le \mathbf{y}} \boldsymbol{\alpha}_{\le}(\mathbf{y} - \mathbf{h}) B_{\mathbf{h}} (I - B_0)^{-1}, \ \text{ otherwise.}\end{aligned} \tag{16}$$

For $i = 1, 2, ..., m$, define

$$\begin{aligned}p_{\mathbf{Y},\le,i}(\mathbf{y}) &= \mathbb{P}\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \le \mathbf{y}^{(2)} | J(0) = i\} \\ &= \sum_{\mathbf{u}^{(2)}:\ \mathbf{u}^{(2)} \le \mathbf{y}^{(2)}} p_{\mathbf{Y},i}((\mathbf{y}^{(1)}, \mathbf{u}^{(2)})).\end{aligned} \tag{17}$$

and $\mathbf{p}_{\mathbf{Y},\le}(\mathbf{y}) = (p_{\mathbf{Y},\le,1}(\mathbf{y}), ..., p_{\mathbf{Y},\le,m}(\mathbf{y}))$. We have

$$\begin{aligned}&\mathbf{p}_{\mathbf{Y},\le}(0) = (I - B_0)^{-1}\, \mathbf{b}^0, \ \text{ if } \mathbf{y} = 0; \\ &\mathbf{p}_{\mathbf{Y},\le}((0, \mathbf{y}^{(2)})) = (I - B_0)^{-1}\, \mathbf{b}^0 \\ &\qquad + \sum_{\mathbf{h} \in C^0:\ \mathbf{h} \le \mathbf{y}} (I - B_0)^{-1}\, B_{\mathbf{h}} \mathbf{p}_{\mathbf{Y},\le}(\mathbf{y} - \mathbf{h}), \ \text{ if } \mathbf{y}^{(2)} \ne 0; \\ &\mathbf{p}_{\mathbf{Y},\le}(\mathbf{y}) = \sum_{\mathbf{h} \in C^0:\ \mathbf{h} \le \mathbf{y}} (I - B_0)^{-1}\, B_{\mathbf{h}} \mathbf{p}_{\mathbf{Y},\le}(\mathbf{y} - \mathbf{h}), \ \text{ otherwise.}\end{aligned} \tag{18}$$

Similar to Proposition 3.1, the conditional expectations of D_i, N_i, $N_{(i,j),0}$, $N_{(i,j),\mathbf{h}}$, and $N_{i,0}$ can be obtained. First, the probability of $\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \le \mathbf{y}^{(2)}\}$ is given by

$$\mathbb{P}\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \le \mathbf{y}^{(2)}\} = \boldsymbol{\alpha}_{\le}(\mathbf{y})\mathbf{b}^0 = \boldsymbol{\beta}\mathbf{p}_{\mathbf{Y},\le}(\mathbf{y}). \tag{19}$$

The conditional expectations of D_i, N_i, $N_{(i,j),0}$, $N_{(i,j),\mathbf{h}}$, and $N_{i,0}$ can be obtained as follows.

Proposition 3.3 *Given that* $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \le \mathbf{y}^{(2)}$, *we have*

$$\begin{aligned}
&\mathbb{E}[D_i|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \le \mathbf{y}^{(2)}] = \tfrac{\beta_i p_{\mathbf{Y},\le,i}(\mathbf{y})}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y},\le}(\mathbf{y})};\\
&\mathbb{E}\left[N_i|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \le \mathbf{y}^{(2)}\right] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y},\le}(\mathbf{y})}\left(\sum_{\mathbf{u}:\ \mathbf{u}\le\mathbf{y}} \alpha_i(\mathbf{u}) p_{\mathbf{Y},\le,i}(\mathbf{y}-\mathbf{u})\right);\\
&\mathbb{E}[N_{(i,j),0}|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \le \mathbf{y}^{(2)}]\\
&\qquad = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y},\le}(\mathbf{y})}\left(\sum_{\mathbf{u}:\ \mathbf{u}\le\mathbf{y}} \alpha_i(\mathbf{u})(B_0)_{(i,j)} p_{\mathbf{Y},\le,j}(\mathbf{y}-\mathbf{u})\right); \qquad (20)\\
&\mathbb{E}[N_{(i,j),\mathbf{h}}|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \le \mathbf{y}^{(2)}]\\
&\qquad = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y},\le}(\mathbf{y})}\left(\sum_{\mathbf{u}:\ \mathbf{u}\le\mathbf{y}-\mathbf{h}} \alpha_i(\mathbf{u})(B_{\mathbf{h}})_{(i,j)} p_{\mathbf{Y},\le,j}(\mathbf{y}-\mathbf{u}-\mathbf{h})\right);\\
&\mathbb{E}[N_{i,0}|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \le \mathbf{y}^{(2)}] = \tfrac{\alpha_{\le,i}(\mathbf{y})(\mathbf{b}^0)_i}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y},\le}(\mathbf{y})}.
\end{aligned}$$

Lastly, we consider the case with condition $\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \ge \mathbf{y}^{(2)}\}$. Define for $i = 1, 2, ..., m$,

$$\begin{aligned}
\alpha_{\ge,i}(\mathbf{y}) &= \sum_{t=0}^{\infty} \mathbb{P}\left\{\mathbf{Y}^{(1)}(t) = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)}(t) \ge \mathbf{y}^{(2)}, J(t) = i\right\}\\
&= \sum_{\mathbf{u}^{(2)}:\ \mathbf{u}^{(2)}\ge\mathbf{y}^{(2)}} \alpha_i((\mathbf{y}^{(1)}, \mathbf{u}^{(2)})).
\end{aligned} \qquad (21)$$

and $\boldsymbol{\alpha}_{\ge}(\mathbf{y}) = (\alpha_{\ge,1}(\mathbf{y}), ..., \alpha_{\ge,m}(\mathbf{y}))$. Then we have that

$$\begin{aligned}
\boldsymbol{\alpha}_{\ge}(0) &= \boldsymbol{\beta}\left(I - B_0 - \sum_{\mathbf{h}\in C^0:\ \mathbf{h}^{(1)}=0} B_{\mathbf{h}}\right)^{-1};\\
\boldsymbol{\alpha}_{\ge}(\mathbf{y}) &= \left(\sum_{\mathbf{h}\in C^0:\ \mathbf{h}^{(1)}\le\mathbf{y}^{(1)},\ (\mathbf{y}^{(1)}-\mathbf{h}^{(1)},(\mathbf{y}^{(2)}-\mathbf{h}^{(2)})^+)\ne\mathbf{y}} \boldsymbol{\alpha}_{\ge}((\mathbf{y}-\mathbf{h})^+)B_{\mathbf{h}}\right)\\
&\quad\cdot\left(I - B_0 - \sum_{\mathbf{h}\in C^0:\ (\mathbf{y}^{(1)}-\mathbf{h}^{(1)},(\mathbf{y}^{(2)}-\mathbf{h}^{(2)})^+)=\mathbf{y}} B_{\mathbf{h}}\right)^{-1},\ \text{if } \mathbf{y}\ne 0,
\end{aligned} \qquad (22)$$

where $(\mathbf{y}-\mathbf{h})^+ = (\max(0, y_1 - h_1), ..., \max(0, y_K - h_K))$.

For $i = 1, 2, ..., m$, define

$$\begin{aligned}
p_{\mathbf{Y},\ge,i}(\mathbf{y}) &= \mathbb{P}\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \ge \mathbf{y}^{(2)}|J(0) = i\}\\
&= \sum_{\mathbf{u}^{(2)}:\ \mathbf{u}^{(2)}\ge\mathbf{y}^{(2)}} p_{\mathbf{Y},i}((\mathbf{y}^{(1)}, \mathbf{u}^{(2)})).
\end{aligned} \qquad (23)$$

and $\mathbf{p}_{\mathbf{Y},\ge}(\mathbf{y}) = (p_{\mathbf{Y},\ge,1}(\mathbf{y}), ..., p_{\mathbf{Y},\ge,m}(\mathbf{y}))$. We have

$$\begin{aligned}
\mathbf{p}_{\mathbf{Y},\ge}(0) &= \left(I - B_0 - \sum_{\mathbf{h}\in C^0:\ \mathbf{h}^{(1)}=0} B_{\mathbf{h}}\right)^{-1}\mathbf{b}^0;\\
\mathbf{p}_{\mathbf{Y},\ge}(\mathbf{y}) &= \left(I - B_0 - \sum_{\mathbf{h}\in C^0:\ (\mathbf{y}^{(1)}-\mathbf{h}^{(1)},(\mathbf{y}^{(2)}-\mathbf{h}^{(2)})^+)=\mathbf{y}} B_{\mathbf{h}}\right)^{-1}\\
&\quad\cdot\left(\sum_{\mathbf{h}\in C^0:\ \mathbf{h}^{(1)}\le\mathbf{y}^{(1)},\ (\mathbf{y}^{(1)}-\mathbf{h}^{(1)},(\mathbf{y}^{(2)}-\mathbf{h}^{(2)})^+)\ne\mathbf{y}} B_{\mathbf{h}}\mathbf{p}_{\mathbf{Y},\ge}((\mathbf{y}-\mathbf{h})^+)\right),\ \text{if } \mathbf{y}\ne 0.
\end{aligned} \qquad (24)$$

Similar to Propositions 3.1 and 3.3, the conditional expectations of D_i, N_i, $N_{(i,j),0}$, $N_{(i,j),\mathbf{h}}$, and $N_{i,0}$ can be obtained. First, the probability of $\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \geq \mathbf{y}^{(2)}\}$ is given by

$$\mathbb{P}\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \geq \mathbf{y}^{(2)}\} = \boldsymbol{\alpha}_{\geq}(\mathbf{y})\mathbf{b}^0 = \boldsymbol{\beta}\mathbf{p}_{\mathbf{Y},\geq}(\mathbf{y}). \tag{25}$$

To present the formulas for the expectations, we need to introduce more notation. Let

$$\begin{aligned}\Phi &= \{(0, ..., 0, j_{K_1+1}, ..., j_K) : \ j_k = 0 \text{ or } 1, \ k = K_1 + 1, ..., K\} \\ &= \{\boldsymbol{\eta} = (\eta_1, ..., \eta_K) : \eta_k = 0, \text{if } k = 1, ..., K_1; \eta_k = 0 \text{ or } 1, \text{for } k = K_1 + 1, ..., K\}.\end{aligned} \tag{26}$$

It is clear that Φ has 2^{K-K_1} elements. Define, for $\boldsymbol{\eta} \in \Phi$,

$$\Omega_{\boldsymbol{\eta}}(\mathbf{y}) = \{\mathbf{u} : \ u_k = 0, 1, 2, ..., y_k, \text{if } \eta_k = 0; u_k = y_k + 1, \ \text{if } \eta_k = 1\}. \tag{27}$$

Using the above notation, vectors $\boldsymbol{\alpha}_{\geq}(\mathbf{y})$ and $\mathbf{p}_{\mathbf{Y},\geq}(\mathbf{y})$ can be redefined as follows: for $\boldsymbol{\eta} \in \Phi$,

$$\begin{aligned}&\alpha^{\boldsymbol{\eta}}_{\geq,i}(\mathbf{y}) = \sum_{t=0}^{\infty} \mathbb{P}\{Y_k(t) = y_k, \text{if } \eta_k = 0; Y_k(t) \geq y_k + 1, \text{if } \eta_k = 1; J(t) = i\}; \\ &p^{\boldsymbol{\eta}}_{\mathbf{Y},\geq,i}(\mathbf{y}) = \mathbb{P}\{Y_k = y_k, \text{if } \eta_k = 0; Y_k \geq y_k, \text{if } \eta_k = 1 | J(0) = i\}.\end{aligned} \tag{28}$$

For convenience, let $\boldsymbol{\eta}_{\max} = (0, ..., 0, 1, ..., 1)$ with exactly K_1 zero elements, and 0 for the vector of zeros. It is easy to see that $\boldsymbol{\alpha}(\mathbf{y}) = \boldsymbol{\alpha}^0_{\geq}(\mathbf{y})$ and $\mathbf{p}_{\mathbf{Y}}(\mathbf{y}) = \mathbf{p}^0_{\mathbf{Y},\geq}(\mathbf{y})$.

Proposition 3.4 *Given that $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$, $\mathbf{Y}^{(2)} \geq \mathbf{y}^{(2)}$, the conditional expectation of D_i, N_i, $N_{(i,j),0}$, $N_{(i,j),\mathbf{h}}$, and $N_{i,0}$ are given by*

$$\mathbb{E}[D_i | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \geq \mathbf{y}^{(2)}] = \frac{\beta_i p^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq,i}(\mathbf{y})}{\boldsymbol{\beta}\mathbf{p}^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq}(\mathbf{y})}. \tag{29}$$

The conditional expectation of N_i is given by

$$\begin{aligned}&\mathbb{E}\left[N_i | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \geq \mathbf{y}^{(2)}\right] \\ &= \frac{1}{\boldsymbol{\beta}\mathbf{p}^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq}(\mathbf{y})} \sum_{\mathbf{u}\in\Omega_0(\mathbf{y})} \alpha^0_{\geq,i}(\mathbf{u}) p^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq,i}(\mathbf{y} - \mathbf{u}) \\ &\quad + \frac{1}{\boldsymbol{\beta}\mathbf{p}^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq}(\mathbf{y})} \sum_{\boldsymbol{\eta}\in\Phi:\ \boldsymbol{\eta}\neq 0} \sum_{\mathbf{u}\in\Omega_{\boldsymbol{\eta}}(\mathbf{y})} \alpha^{\boldsymbol{\eta}}_{\geq,i}(\mathbf{u}) p^{\boldsymbol{\eta}}_{\mathbf{Y},\geq,i}((\mathbf{y} - \mathbf{u})^+).\end{aligned} \tag{30}$$

The conditional expectation of $N_{(i,j),0}$ is given by

$$\begin{aligned}&\mathbb{E}\left[N_{(i,j),0} | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \geq \mathbf{y}^{(2)}\right] \\ &= \frac{1}{\boldsymbol{\beta}\mathbf{p}^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq}(\mathbf{y})} \sum_{\mathbf{u}\in\Omega_0(\mathbf{y})} \alpha^0_{\geq,i}(\mathbf{u})(B_0)_{(i,j)} p^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq,j}(\mathbf{y} - \mathbf{u}) \\ &\quad + \frac{1}{\boldsymbol{\beta}\mathbf{p}^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq}(\mathbf{y})} \sum_{\boldsymbol{\eta}\in\Phi:\ \boldsymbol{\eta}\neq 0} \sum_{\mathbf{u}\in\Omega_{\boldsymbol{\eta}}(\mathbf{y})} \alpha^{\boldsymbol{\eta}}_{\geq,i}(\mathbf{u})(B_0)_{(i,j)} p^{\boldsymbol{\eta}}_{\mathbf{Y},\geq,i}((\mathbf{y} - \mathbf{u})^+).\end{aligned} \tag{31}$$

The conditional expectation of $N_{(i,j),\mathbf{h}}$, for $\mathbf{h} \in C^0$, is given by

$$\begin{aligned}&\mathbb{E}\left[N_{(i,j),\mathbf{h}} | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \geq \mathbf{y}^{(2)}\right] \\ &= \frac{1}{\boldsymbol{\beta}\mathbf{p}^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq}(\mathbf{y})} \sum_{\mathbf{u}\in\Omega_0(\mathbf{y}-\mathbf{h})} \alpha^0_{\geq,i}(\mathbf{u})(B_{\mathbf{h}})_{(i,j)} p^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq,j}(\mathbf{y} - \mathbf{u} - \mathbf{h}) \\ &\quad + \frac{1}{\boldsymbol{\beta}\mathbf{p}^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq}(\mathbf{y})} \sum_{\boldsymbol{\eta}\in\Phi:\ \boldsymbol{\eta}\neq 0} \sum_{\mathbf{u}\in\Omega_{\boldsymbol{\eta}}(\mathbf{y}-\mathbf{h})} \alpha^{\boldsymbol{\eta}}_{\geq,i}(\mathbf{u})(B_{\mathbf{h}})_{(i,j)} p^{\boldsymbol{\eta}}_{\mathbf{Y},\geq,i}((\mathbf{y} - \mathbf{u} - \mathbf{h})^+).\end{aligned} \tag{32}$$

The conditional expectation of $N_{i,0}$ is given by

$$\mathbb{E}[N_{i,0} | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \geq \mathbf{y}^{(2)}] = \frac{\alpha^{\boldsymbol{\eta}_{\max}}_{\geq,i}(\mathbf{y})(\mathbf{b}^0)_i}{\boldsymbol{\beta}\mathbf{p}^{\boldsymbol{\eta}_{\max}}_{\mathbf{Y},\geq}(\mathbf{y})}. \tag{33}$$

Note 4 The conditions $\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \leq \mathbf{y}^{(2)}\}$ and $\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \geq \mathbf{y}^{(2)}\}$ can be generalized to $\{Y_k \in \Omega_k,\ k = 1, 2, ..., K\}$, where Ω_k is a set of possible values for Y_k. It is clear that the conditional expectations of D_i, N_i, $N_{(i,j),\mathbf{h}}$, and $N_{i,0}$ are obtainable. However, the expressions for the conditional expectation become much more involved. We omit the details.

4 Parameter Estimation

In this section, we present an EM-algorithm to estimate the parameters of the DMPH distribution. Assume that a sample of $\mathbf{Y}$ of size n is observed as $\mathbf{y}_{1...n} = \left\{\mathbf{y}^{[v]} = (y_1^{[v]}, ..., y_K^{[v]}),\ v = 1, 2, ..., n\right\}$. Assume that C^0, the collection of types of the event batches, and m, the number of states of the underlying Markov chain, are given. We will estimate parameters

$$\boldsymbol{\theta} = \{\boldsymbol{\beta}, B_0, B_{\mathbf{h}}, \mathbf{h} \in C^0, \mathbf{b}^0\} \tag{34}$$

for the standard matrix-representation of the model. As in Asmussen et al. (1996), the EM-algorithm is a natural way to estimating parameters here because of the connection between the distribution and the underlying Markov chain. An observation $\mathbf{y}^{[v]}$ of the numbers of individual events before absorbtion can be regarded as an incomplete observation of the Markov chain $\{J(t)\}_{t=0,1,...}$.

The complete information of the Markov chain $\{J(t)\}_{t=0,1,...}$ include the absorption time τ, the states of the process $J_0, J_1, \cdots, J_{\tau-1}$ at time instants $t = 0, 1, ..., \tau - 1$ before absorbtion and the event batches $\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_{\tau-1}$ that are associated with the transitions made at time $1, 2, \cdots, \tau - 1$, where $\mathbf{H}_t = 0$ if there is no event batch in the period. The probability of observing a complete sample path $(j_0, j_1, \cdots, j_{\tau-1}, \mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_{\tau-1})$ is given by

$$\alpha_{j_0}(B_{\mathbf{h}_1})_{(j_0,j_1)} \cdots (B_{\mathbf{h}_{\tau-1}})_{(j_{\tau-2},j_{\tau-1})}(\mathbf{b}^0)_{j_{\tau-1}}. \tag{35}$$

Let

$$\begin{aligned} \mathbf{x} &= \{\mathbf{x}^{[1]}, ..., \mathbf{x}^{[n]}\} \\ &= \{j_0^{[1]}, j_1^{[1]}, \cdots, j_{\tau-1}^{[1]}, \mathbf{h}_1^{[1]}, \mathbf{h}_2^{[1]}, ..., \mathbf{h}_{\tau-1}^{[1]}, \cdots, \\ &\qquad j_0^{[n]}, j_1^{[n]}, \cdots, j_{\tau-1}^{[n]}, \mathbf{h}_1^{[n]}, \mathbf{h}_2^{[n]}, ..., \mathbf{h}_{\tau-1}^{[n]}\} \end{aligned} \tag{36}$$

be the complete information corresponding to the observed sample $\mathbf{y}_{1...n}$. For each point $\mathbf{x}^{[v]}$, $v = 1, \cdots, n$, the conditional expectations of $D_i^{[v]}$, $N_i^{[v]}$, $N_{(i,j),0}^{[v]}$, $N_{(i,j),\mathbf{h}}^{[v]}$, and $N_{i,0}^{[v]}$, as defined in Section 3, can be calculated. Let $D_i = \sum_{v=1}^n D_i^{[v]}$, $N_i = \sum_{v=1}^n N_i^{[v]}$, $N_{(i,j),0} = \sum_{v=1}^n N_{(i,j),0}^{[v]}$, $N_{(i,j),\mathbf{h}} = \sum_{v=1}^n N_{(i,j),\mathbf{h}}^{[v]}$, and $N_{i,0} = \sum_{v=1}^n N_{i,0}^{[v]}$. Then the likelihood of observing $\mathbf{x}$ is given by

$$L(\boldsymbol{\theta}, \mathbf{x}) = \left(\prod_{i=1}^m \beta_i^{D_i}\right)\left(\prod_{\mathbf{h}\in C^0\cup\{0\}} \prod_{i=1}^m \prod_{j=1}^m \{B_{\mathbf{h}}\}_{(i,j)}^{N_{(i,j),\mathbf{h}}}\right)\left(\prod_{i=1}^m (\mathbf{b}^0)_i^{N_{i,0}}\right), \tag{37}$$

and the loglikelihood is defined by $l(\boldsymbol{\theta}, \mathbf{x}) = \log L(\boldsymbol{\theta}, \mathbf{x})$. Maximizing the above loglikelihood under constraints of $\beta_1 + \beta_2 + ... + \beta_m = 1$ and $(B_0 + \sum_{\mathbf{h}\in C^0} B_{\mathbf{h}})\mathbf{e} + \mathbf{b}^0 = \mathbf{e}$, gives the maximum likelihood estimator of the transition probabilities. This yields for $i, j = 1, \cdots, m$, and $\mathbf{h} \in C^0$,

$$\hat{\beta}_i = \frac{D_i}{n}, \quad (\hat{B}_0)_{(i,j)} = \frac{N_{(i,j),0}}{N_i}, \quad (\hat{B}_{\mathbf{h}})_{(i,j)} = \frac{N_{(i,j),\mathbf{h}}}{N_i}, \quad \text{and} \quad (\hat{\mathbf{b}}^0)_i = \frac{N_{i,0}}{N_i}. \tag{38}$$

Notice that $D_1 + D_2 + ... + D_m = n$ and $N_i = N_{i,0} + \sum_{j=1}^{m} \left(N_{(i,j),0} + \sum_{\mathbf{h}:\ \mathbf{h} \in C^0} N_{(i,j),\mathbf{h}} \right)$. A detailed proof of the above formulas can be found in Basawa and Rao (1980). For convenience, we provide a proof in Appendix A.

In EM-algorithm, we first assume an initial values for the parameter vector $\boldsymbol{\theta}$, then it is updated iteratively. Suppose the value of the parameter vector is $\boldsymbol{\theta}^{(s)}$ after the s-th iteration. In the $(s+1)$-st expectation step, we compute the conditional expectation of the loglikelihood $l(\boldsymbol{\theta}, \mathbf{x})$ given the observed sample $\mathbf{y}$ to obtain

$$Q(\boldsymbol{\theta}^{(s)}, \mathbf{y}_{1...n}) = \mathbb{E}[l(\boldsymbol{\theta}^{(s)}, \mathbf{X}) | \mathbf{y}_{1...n}]. \tag{39}$$

Then in the $(s+1)$-st maximization step, a new set of parameter value $\boldsymbol{\theta}$ is obtained by

$$\boldsymbol{\theta}^{(s+1)} = \mathrm{argmax}_{\boldsymbol{\theta}}\, Q(\boldsymbol{\theta}^{(s)}, \mathbf{y}_{1...n}). \tag{40}$$

The above maximization can be carried out simply by using the Eq. 38, which leads to

$$\begin{aligned} \beta_i^{(s+1)} &= \frac{\mathcal{D}_i^{(k+1)}}{n}, \quad (B_0)_{(i,j)}^{(s+1)} = \frac{\mathcal{N}_{(i,j),0}^{(s+1)}}{\mathcal{N}_i^{(s+1)}}, \\ (B_{\mathbf{h}})_{(i,j)}^{(s+1)} &= \frac{\mathcal{N}_{(i,j),\mathbf{h}}^{(s+1)}}{\mathcal{N}_i^{(s+1)}}, \quad (\mathbf{b}^0)_i^{(s+1)} = \frac{\mathcal{N}_{i,0}^{(s+1)}}{\mathcal{N}_i^{(s+1)}}, \end{aligned} \tag{41}$$

where, for $i, j = 1, \cdots, \cdots, m$, and $\mathbf{h} \in C^0$,

$$\begin{aligned} \mathcal{D}_i^{(s+1)} &= \mathbb{E}_{\boldsymbol{\theta}^{(s)}}[D_i | \mathbf{y}_{1...n}] = \sum_{v=1}^{n} \mathbb{E}_{\boldsymbol{\theta}^{(s)}} \left[D_i^{[v]} | \mathbf{Y} = \mathbf{y}^{[v]} \right]; \\ \mathcal{N}_i^{(s+1)} &= \mathbb{E}_{\boldsymbol{\theta}^{(s)}}[N_i | \mathbf{y}_{1...n}] = \sum_{v=1}^{n} \mathbb{E}_{\boldsymbol{\theta}^{(s)}} \left[N_i^{[v]} | \mathbf{Y} = \mathbf{y}^{[v]} \right]; \\ \mathcal{N}_{(i,j),0}^{(s+1)} &= \mathbb{E}_{\boldsymbol{\theta}^{(s)}}[N_{(i,j),0} | \mathbf{y}_{1...n}] = \sum_{v=1}^{n} \mathbb{E}_{\boldsymbol{\theta}^{(s)}} \left[N_{(i,j),0}^{[v]} | \mathbf{Y} = \mathbf{y}^{[v]} \right]; \\ \mathcal{N}_{(i,j),\mathbf{h}}^{(s+1)} &= \mathbb{E}_{\boldsymbol{\theta}^{(s)}}[N_{(i,j),\mathbf{h}} | \mathbf{y}_{1...n}] = \sum_{v=1}^{n} \mathbb{E}_{\boldsymbol{\theta}^{(s)}} \left[N_{(i,j),\mathbf{h}}^{[v]} | \mathbf{Y} = \mathbf{y}^{[v]} \right]; \\ \mathcal{N}_{i,0}^{(s+1)} &= \mathbb{E}_{\boldsymbol{\theta}^{(s)}}[N_{i,0} | \mathbf{y}_{1...n}] = \sum_{v=1}^{n} \mathbb{E}_{\boldsymbol{\theta}^{(s)}} \left[N_{i,0}^{[v]} | \mathbf{Y} = \mathbf{y}^{[v]} \right]. \end{aligned} \tag{42}$$

Note that we use superscript "$[v]$" for sample points and superscript "(s)" for iteration. Each of the conditional expectations on the right–hand–side of Eq. 42 can be computed by applying Proposition 3.1. For the issue of convergence of the EM-algorithm, one is referred to Jeff Wu (1983).

For convenience, we summarize the steps in the EM-aglorithm for computation.

Step 1 Intialize parameters $\{\boldsymbol{\beta}, B_0, B_{\mathbf{h}}, \mathbf{h} \in C^0, \mathbf{b}^0\}$.

Step 2 For each sample point $\mathbf{y}^{[v]}$, a) use formulas given in Eqs. 7 and 10 to compute $\mathbf{p}_{\mathbf{Y}}(\mathbf{u})$ and $\boldsymbol{\alpha}(\mathbf{u})$ for $\mathbf{u} \le \mathbf{y}$, respectively; and b) use formulas given in Eq. 11 to compute conditional expectations of D_i, N_i, $N_{(i,j),0}$, $N_{(i,j),\mathbf{h}}$, and $N_{i,0}$.

Step 3 Use formulas given in Eq. 42 to compute the sums of the conditional expectations for the sample $\mathbf{y}_{1...n}$ and compute the estimates of parameters using formulas given in Eq. 41.

Step 4 Repeat Steps 2 and 3 until the solution is satisfactory according to some pre-given rule, which can be i) the difference between two consecutive solutions $||\boldsymbol{\theta}^{(s+1)} - \boldsymbol{\theta}^{(s)}||$ or the difference between two consecutive loglikelihoods $|\log(\hat{L}(\boldsymbol{\theta}^{(s+1)}, \mathbf{y}_{1...n})) - \log(\hat{L}(\boldsymbol{\theta}^{(s)}, \mathbf{y}_{1...n}))|$, where the likelihood associated with sample $\mathbf{y}_{1...n}$ is defined as

$$\hat{L}(\boldsymbol{\theta}, \mathbf{y}_{1...n}) = \Pi_{v=1}^{n} \mathbb{P}\{\mathbf{Y} = \mathbf{y}^{[v]}\} = \Pi_{v=1}^{n} \boldsymbol{\beta} \mathbf{p}_{\mathbf{Y}}(\mathbf{y}^{[v]}). \tag{43}$$

Note 5 For the cases discussed in Notes 1 to 3, the above EM-algorithm still works with minor modification. All we need to do is to set the initial parameter set $\boldsymbol{\theta}$ properly.

Note 6 If the sample points/observations are in the form $\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \leq \mathbf{y}^{(2)}\}$ or $\{\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \geq \mathbf{y}^{(2)}\}$, the EM-algorithm works by utilizing formulas given in Propositions 3.3 and 3.4 for individual sample points. The computation of the expectations for such cases is much more involved, especially for the latter case. For the case with $K = 2$, explicit formulas are presented in Appendix C for the use of our computation for Section 5.3. For this case, four types of sample points are considered: i) $\{Y_1 = y_1, Y_2 = y_2\}$; ii) $\{Y_1 = y_1, Y_2 \geq y_2\}$; iii) $\{Y_1 \geq y_1, Y_2 = y_2\}$; and iv) $\{Y_1 \geq y_1, Y_2 \geq y_2\}$.

Note 7 For many applications, there are duplications in sample points. For such cases, one perform E-steps for each of the distinct sample points. Then the loglikelihood of the whole sample is just the summation of the loglikelihood of each distinct sample points weighted by the number of duplications.

An interesting property of the proposed EM-algorithm is that either the sample means are preserved in the iteration process or the orders of sample means and the means of the estimated DMPHs are consistent in each of the iteration process.

Proposition 4.1 *Consider collected sample* $\mathbf{y}_{1\ldots n} = \{\mathbf{y}^{[v]}, v = 1, 2, ..., n\}$. *Then we have, for each event type k,*

(1) *If the sample points is of the form* $\{Y_k^{[v]} \leq y_k^{[v]}, v = 1, 2, ..., n\}$*, then we have* $\mathbb{E}_{\boldsymbol{\theta}^{(s)}}[Y_k] \leq \frac{1}{n}\sum_{v=1}^{n} y_k^{[v]}$*, for* $s = 1, 2, \ldots$.
(2) *If the sample points is of the form* $\{Y_k^{[v]} = y_k^{[v]}, v = 1, 2, ..., n\}$*, then we have* $\mathbb{E}_{\boldsymbol{\theta}^{(s)}}[Y_k] = \frac{1}{n}\sum_{v=1}^{n} y_k^{[v]}$*, for* $s = 1, 2, \ldots$.
(3) *If the sample points is of the form* $\{Y_k^{[v]} \geq y_k^{[v]}, v = 1, 2, ..., n\}$*, then we have* $\mathbb{E}_{\boldsymbol{\theta}^{(s)}}[Y_k] \geq \frac{1}{n}\sum_{v=1}^{n} y_k^{[v]}$*, for* $s = 1, 2, \ldots$.

Proof We only give details for the case with $\{Y = y\}$. The proof is similar to that of a similar property for continuous time phase-type distributions given in Asmussen et al. (1996) and Hassan Zadeh and Bilodeau (2013). First, it is easy to see that, for $k = 1, 2, ..., K$, $v = 1, 2, ..., n$, and the sample $\mathbf{y}$,

$$y_k^{[v]} = \sum_{\mathbf{h}\in C^0} h_k \sum_{i=1}^{m}\sum_{j=1}^{m} N_{(i,j),\mathbf{h}}^{[v]}. \tag{44}$$

In each iteration of the EM-algorithm, we have for $v = 1, \cdots, n$

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}^{(s+1)}}[Y_k^{[v]}] &= \mathbb{E}_{\boldsymbol{\theta}^{(s+1)}}\left[\sum_{\mathbf{h}\in C^0} h_k \sum_{i=1}^{m}\sum_{j=1}^{m} N_{(i,j),\mathbf{h}}^{[v]}\right]\\ &= \mathbb{E}_{\boldsymbol{\theta}^{(s)}}\left[\sum_{\mathbf{h}\in C^0} h_k \sum_{i=1}^{m}\sum_{j=1}^{m} N_{(i,j),\mathbf{h}}^{[v]}\middle|\,\mathbf{y}\right]\\ &= \mathbb{E}_{\boldsymbol{\theta}^{(s)}}\left[Y_k^{[v]}\middle|\,\mathbf{y}\right]\\ &= y_k^{[v]}\end{aligned} \tag{45}$$

Summing up both sides of Eq. 45 over $v = 1, \cdots, n$ and then divided by n, one obtain the desired result. For the other two cases, the equality in Eq. 44 becomes inequality. Then the equality in Eq. 45 becomes inequality of the same type accordingly. This completes the proof of Proposition 4.1. □

Proposition 4.1 reveals an intrisic property of the EM-algorithm and it is useful for checking computation accuracy. A few observations associated with the EM-algorithm and the likelihood $\hat{L}(\boldsymbol{\theta}, \mathbf{y}_{1\ldots n})$ can also be useful for checking computation accuracy and the correctness of programming.

- The likelihood $\hat{L}(\boldsymbol{\theta}^{(s)}, \mathbf{y}_{1\ldots n})$ is increasing in s.
- The likelihood $\hat{L}(\boldsymbol{\theta}, \mathbf{y}_{1\ldots n})$ of the optimal solution $\boldsymbol{\theta}$ is increasing in m.

We would like to point out that the elements in the batch set C^0 are determined by the nature of application problems. On the other hand, the order m is a parameter that has to be chosen by the parameter estimation algorithm, unless the type of the DMPH to be fitted is given (see Example 4.1). In principle, m should be chosen to the smallest value such that the model fits the data reasonably well.

Example 4.1 (Example 2.1 continued) For this case, the EM-algorithm can be simplified significantly since $m = 1$. In addition, it is easy to see that $\mathbb{E}[D|\mathbf{Y} = \mathbf{y}] = 1$ and $\mathbb{E}[N|\mathbf{Y} = \mathbf{y}] = y_1 + \ldots + y_K$. However, the computations of other vectors and quantities (e.g., $\boldsymbol{\alpha}(\mathbf{y})$, $\mathbb{E}[N_{\mathbf{h}}|\mathbf{Y} = \mathbf{y}]$) still have to use the formulas given in Section 3.

Example 4.2 (Example 2.2 continued) In the special case, we assumed $B_{\mathbf{h}} = w_{\mathbf{h}} B$, for $\mathbf{h} \in C^0 \cup \{0\}$, which reduced the number parameters greatly. For this case, the E-step remains the same as the general case. The formulas for performing the M-step are given by

$$\begin{aligned} (\hat{B})_{(i,j)} &= \frac{1}{N_i} \sum_{\mathbf{h} \in C^0 \cup \{0\}} N_{(i,j),\mathbf{h}}; \\ \hat{w}_{\mathbf{h}} &= \frac{\sum_{i=1}^m \sum_{j=1}^m N_{(i,j),\mathbf{h}}}{\sum_{\mathbf{h} \in C^0 \cup \{0\}} \sum_{i=1}^m \sum_{j=1}^m N_{(i,j),\mathbf{h}}}. \end{aligned} \tag{46}$$

A proof of Eq. 46 can be found in Appendix B.

5 Numerical Analysis

In this section, we carry out a numerical analysis on the EM-algorithm to address technical issues such as the selection of m and the impact of B_0 (Section 5.1). We also examine the approximation of the well-known uniform distribution by DMPHs (see Section 5.2). In Section 5.3, we examine the generality of the model by using it to fit data from different sources.

We would like to pioint out that, to assess the goodness-of-fit, as discussed in Section 4, we use two criteria: i) The distance between two consecutive solutions $\boldsymbol{\theta}$; and ii) the value of the corresponding likelihood function $\hat{L}(\boldsymbol{\theta}, \mathbf{y}_{1\ldots n})$.

5.1 An Example

We consider an example with $K = 3$ and $C^0 = \{\mathbf{h}_1 = (1, 0, 0), \mathbf{h}_2 = (0, 1, 0), \mathbf{h}_3 = (0, 0, 1), \mathbf{h}_4 = (2, 0, 1), \mathbf{h}_5 = (0, 1, 1)\}$. We consider a sample with $n = 8$ distinct sample

Table 1 Frequency estimates for Section 5.1 ($B_0 = 0$)

m	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_3$	$\hat{r}_4$	$\hat{r}_5$	$\hat{r}_6$	$\hat{r}_7$	$\hat{r}_8$
	(5)	(8)	(4)	(4)	(3)	(2)	(2)	(1)
1	13.88	6.362	4.026	2.136	0.268	2.121	0.076	0.1257
2	15.39	5.016	5.129	1.666	0.5117	1.104	0.0825	0.0975
3	17.222	5.832	3.652	1.418	0.1258	0.7196	0.0229	0.0059
4	6.947	14.88	2.084	4.078	0.2881	0.6141	0.0801	0.0253
6	3.964	11.96	6.999	4.173	1.0767	0.6396	0.1123	0.0452

points $\mathbf{y}_{1...n} = \{(1,1,1), (1,2,2), (2,1,2), (2,3,2), (3,1,4), (3,3,3), (5,5,2), (8,2,4)\}$. The repeating numbers of the sample points $\{r_1, ..., r_8\}$ are given in the second row of Table 1 (in parentheses). The sample point is of the form $\{\mathbf{Y} = \mathbf{y}\}$. We first consider a DMPH model with $B_0 = 0$. For $m = 1, 2, 3,$ and 4, we use the EM-algorithm to estimate $\{\boldsymbol{\beta}, B_{(1,0,0)}, B_{(0,1,0)}, B_{(0,0,1)}, B_{(2,0,1)}, B_{(0,1,1)}, \mathbf{b}^0\}$. Then we compute the probabilites $\mathbb{P}\{\mathbf{Y} = \mathbf{y}^{[v]}\}$ for the eight distinct sample points, and estimate the frequencies for each sample point by: for $v = 1, 2, ..., 8$.

$$\hat{r}_v = \left(\sum_{s=1}^{n} r_s\right) \frac{\mathbb{P}\{\mathbf{Y} = \mathbf{y}^{[v]}\}}{\sum_{s=1}^{n} \mathbb{P}\{\mathbf{Y} = \mathbf{y}^{[s]}\}}. \tag{47}$$

The results are presented in Table 1. As is shown in Table 1, if m increases, the estimates are closer to the original data. The loglikelihoods of the final solutions for $m = 1, 2, 3,$ and 4, are -147.9, -123.64, -101.5, -97.33, and -73.72, respectively, which is increasing in m.

Second, we remove the constraint $B_0 = 0$. The results are given in Table 2. The (log) likelihoods of the final solutions for $m = 1, 2, 3,$ and 4, are -147.9, -123.6, -111.6, -89.3, and -80.0, respectively, which is also increasing in m. Comparing results in Tables 1 and 2, the first model generates slightly better results when m is not small. Thus, the effect of setting the proper B_0 and selecting the proper m on the quality of the fitted DMPH is significant, and shouldn't be ignored in the modeling process. In addition, we note that the likelihood for the case with $m = 4$ and without the constraint is smaller than the likelihood for the case with constraint $B_0 = 0$. This contradicts the theory. A possible explanation is that the initial values for B_0 led to a local maxima (e.g., Wu 1983).

For this arbitrarily generated example, fitting of the sample distribution/frequencies seems inadquate. The reason is that distinct sample points are scattered in a space of three dimensions. Thus, fitting a multivariate distribution for this set of sample points is

Table 2 Frequency estimates for Section 5.1 ($B_0 \neq 0$)

m	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_3$	$\hat{r}_4$	$\hat{r}_5$	$\hat{r}_6$	$\hat{r}_7$	$\hat{r}_8$
	(5)	(8)	(4)	(4)	(3)	(2)	(2)	(1)
1	13.89	6.349	4.017	2.144	0.273	2.122	0.078	0.1224
2	15.37	5.000	5.168	1.662	0.5117	1.104	0.0825	0.096
3	8.232	7.480	10.35	1.754	0.337	0.730	0.068	0.043
4	5.203	13.24	6.612	3.024	0.106	0.757	0.470	0.0075
6	6.265	9.252	8.253	3.430	0.194	1.417	0.176	0.0103

challenging. A larger m may lead to better fitting, though. As will be shown in Sections 5.2 and 5.3, the fitting results are much better if the set of distinct sample points are located in a geometrically compact region, even for small m.

5.2 Approximating a Bivariate Uniform Distribution

As another numerical experiment, we fit the DMPH distribution to a hypothetical data set, where there are 5 observations on each point of a $(0, 1, 2, 3, 4) \times (0, 1, 2, 3, 4)$ grid (i.e., $n = 25$ distinct sample points, and repeating number $r_v = 5$, for $v = 1, ..., 25$). By definition, we must have $K = 2$. We assume that $C^0 = \{(1, 0), (0, 1), (1, 1)\}$. Since the purpose of this exercise is to examine whether the DMPH distribution can approximate a bivariate uniform distribution, we do not restrict the number of parameters and so no structure for the matrices are assumed. We fit the DMPH distribution with $m = 1, 4, 8$, and 12, and report the approximation results $\hat{r}_v$ in Table 3. In Table 3, the first row is for y_1 and the first column

Table 3 Parameters fitting of the bivariate uniform distribution data

(y_1, y_2)	0	1	2	3	4	m
	(5)	(5)	(5)	(5)	(5)	
	30.988	12.393	4.956	1.982	0.792	1
0	5.502	5.591	3.871	3.195	1.353	4
	5.198	5.197	5.198	5.184	3.251	8
	5.000	4.999	4.998	4.982	4.094	12
	(5)	(5)	(5)	(5)	(5)	
	12.393	9.916	5.949	3.172	1.586	1
1	5.591	7.742	9.604	7.014	3.088	4
	5.197	5.197	5.201	6.504	4.458	8
	5.000	5.000	5.048	4.885	6.477	12
	(5)	(5)	(5)	(5)	(5)	
	4.956	5.949	4.759	3.173	1.903	1
2	3.871	9.604	11.321	7.476	3.419	4
	5.198	5.201	6.508	6.227	3.733	8
	4.999	4.986	5.378	4.737	4.200	12
	(5)	(5)	(5)	(5)	(5)	
	1.982	3.172	3.173	2.538	1.777	1
3	3.195	7.014	4.476	4.997	2.448	4
	5.183	6.505	6.228	7.463	3.928	8
	5.001	5.002	4.739	5.571	5.543	12
	(5)	(5)	(5)	(5)	(5)	
	0.792	1.586	1.903	1.777	1.421	1
4	1.353	3.088	3.419	2.448	1.307	4
	3.250	4.458	3.734	3.928	2.857	8
	4.258	5.990	4.121	6.270	3.710	12

is for y_2. The repeating numbers of sample points are in the first row of each block (in fact, all of them are 5), given in parenthses, and the estimates of the repeating numbers are given in rows 2 to 5 in each block for $m = 1, 4, 8$, and 12, respectively. The loglikelihoods of the four cases are −484.08, −435.25, −403.97, and −403.3. As is shown by Table 3, the fitting is quite good when m is 12.

5.3 An Auto Insurance Claims Example

We begin with the auto insurance property damage and bodily injury claim data used in David Cummins and Wiltbank (1983), in which the authors concluded that none of the well–known bivariate discrete distributions were suitable. The authors argued that this is because the marginal distributions of the number of the two types of events seem to come from different distribution families. However, for most commonly used bivariate distributions, the marginal distributions are from the same family. The data is listed in Table 4. Notice that the data set covers all four scenarios: $\{Y_1 = y_1, Y_2 = y_2\}$, $\{Y_1 = y_1, Y_2 \geq y_2\}$, $\{Y_1 \geq y_1, Y_2 = y_2\}$, and $\{Y_1 \geq y_1, Y_2 \geq y_2\}$.

We fit the DMPH distribution to the data, assuming that $K = 2$ and $C^0 = \{(1, 0), (0, 1), (1, 1), (1, 2)\}$. That implies that the possible combinations of claims are (1) an accident causes one property damage claim only; (2) an accident causes one bodily injury claim only; (3) an accident causes one property damage claim and one bodily injury claim; and (4) an accident causes one property damage claim and two bodily injury claim. Notice that the claim batches are selected using intuition here. For example, (0, 2) is not selected because it is unlikely that two people are injured yet there is no car damages. The expected number of events are reported in parenthesis in Table 5, where the first row of each block gives the observed numbers of claim and the second to fifth row presents the expected numbers of claims by the fitted distribution with $m = 1$, $m = 2$, $m = 3$, and $m = 4$ (denoted by model 1, 2, 3 and 4) respectively. From the table, we do see that with $m = 4$, one could capture the pattern of the bivariate distribution. This shows that the DMPH distribution is flexible enough to allow dissimilar marginal distributions.

The loglikelihoods obtained for the four models are −374.57, −280.13, −277.55, and −277.24, for $m = 1, 2, 3$, and 4, respectively. To determine which model is "better" in describing the data, we use Akaike's Information Criterion (AIC). AIC is defined by $AIC = 2K - 2\ln(\hat{L})$, where K is the number of parameters in the model and $\hat{L}$ is the maximized value of the likelihood function for the estimated model. For the DMPH models, the

Table 4 Auto insurance claim data

	Property damage events					
	(y_1, y_2)	0	1	2	≥ 3	Totals
Bodily injury events	0	44	49	2	1	96
	1	10	20	2	1	33
	2	2	6	1	1	10
	≥ 3	0	4	5	1	10
	Total	56	79	10	4	149

Table 5 Parameters fitting of the auto insurance claim data

		Property damage events				
		0	1	2	≥ 3	m
Bodily injury events		(44)	(49)	(2)	(1)	
		49.666	8.2777	1.3796	0.2759	1
	0	44.6260	45.7184	1.6086	0.9120	2
		44.7899	48.9682	1.9799	0.9382	3
		43.9454	49.0076	1.9931	0.9872	4
		(10)	(20)	(2)	(1)	
		8.2777	11.0370	11.7268	4.7183	1
	1	8.8724	22.2463	3.1715	1.0594	2
		8.9647	20.1446	1.9899	1.2284	3
		10.3500	20.1218	1.9991	1.0200	4
		(2)	(6)	(1)	(1)	
		1.3796	3.4490	5.7484	10.8788	1
	2	1.7887	8.0554	2.4288	0.8638	2
		1.7942	5.9752	1.0274	0.7153	3
		1.4922	5.9449	1.0068	0.8747	4
		(0)	(4)	(5)	(1)	
		0.2759	1.0761	2.6010	28.2309	1
	≥ 3	0.4640	3.6590	2.2598	1.2651	2
		0.4489	3.9821	4.9075	1.1448	3
		0.2071	3.9829	4.9462	1.1200	4

number of parameters K is given by $K = m - 1 + 5m^2$. The AIC values for the four models are found to be 760, 602, 648, and 720 for $m = 1, 2, 3$, and 4, respectively. Thus, the model with $m = 2$ is the best model according to AIC. For the $m = 2$ case, we present the final representation of the DMPH:

$$\begin{aligned}
&\boldsymbol{\beta} = (0.0299,\ 0.9701);\ B_0 = \begin{pmatrix} 0.0357 & 0.0041 \\ 0.0039 & 0.0357 \end{pmatrix};\\
&B_{(1,0)} = \begin{pmatrix} 0.2430 & 0.0871 \\ 0.0000 & 0.1831 \end{pmatrix};\ B_{(0,1)} = \begin{pmatrix} 0.0005 & 0.0344 \\ 0.4860 & 0.0141 \end{pmatrix};\\
&B_{(1,1)} = B_{(1,2)} = 0;\ \mathbf{b}^0 = \begin{pmatrix} 0.5948 \\ 0.2769 \end{pmatrix}.
\end{aligned} \tag{48}$$

We also used the special case of the model introduced in Example 2.2 to fit the data with $m = 2$, $m = 3$, and $m = 4$. It is found that increasing the dimension of the model does not help improving the fit much. This indicates that it is unreasonable to assume that the matrices $\{B_{\mathbf{h}}, \mathbf{h} \in C^0\}$ are proportional and it verifies the statement in David Cummins and Wiltbank (1983) that the marginal distributions of the number of the two types of events seem to come from different distribution families.

6 Future Research

(1) While the computation for observations of the form $\{\mathbf{Y} = \mathbf{y}\}$, $\{\mathbf{Y} \leq \mathbf{y}\}$, and their mixed case can be done in a straightforward manner, the computation for observation of the form $\{\mathbf{Y} \geq \mathbf{y}\}$ (and its mixed cases) is much more involved and tedious. How to design the computation procedure for such cases is an interesting issue for future research. More generally, how to do parameter estimation under general conditions is a challenging issue.

(2) Our numerical examples demonstrate that the EM-algorithm generates better solutions with respect to the likelihood function or the distance between distributions, if the phase parameter m increases. This observation is intuitive since, if m increases, one does not lose any of the original parameters. Yet a formal proof would be an interesting issue for future research. However, if there are local maxima, the initial values matters, as in the example in Section 5.1.

(3) In our numerical experiments, most of the final estimates produced by the EM-algorithm have the bi-diagonal form. Thus, the initial matrices of the EM-algorithm may be set to have the bi-diagonal structure, which may simplify the algorithm significantly.

Appendix A

Writing $\beta_m = 1 - \sum_{i=1}^{m-1} \beta_i$ and $(\mathbf{b}^0)_i = 1 - \sum_{\mathbf{h}\in C^0\cup\{0\}} \sum_{j=1}^{m} (B_{\mathbf{h}})_{(i,j)}$, the loglikelihood function corresponding to Eq. 37 can be expressed as:

$$L(\boldsymbol{\theta}, \mathbf{x}) = \sum_{j=1}^{m-1} \ln(\beta_j) D_j + \ln\left(1 - \sum_{i=1}^{m-1} \beta_i\right) D_m + \sum_{\mathbf{h}\in C^0\cup\{0\}} \sum_{i=1}^{m} \sum_{j=1}^{m} \ln((B_{\mathbf{h}})_{(i,j)}) N_{(i,j),\mathbf{h}} + \sum_{i=1}^{m} \left(\ln\left(1 - \sum_{\mathbf{h}\in C^0\cup\{0\}} \sum_{j=1}^{m} (B_{\mathbf{h}})_{(i,j)}\right)\right) N_{i,0}. \tag{49}$$

Taking derivative with respect to $(B_{\mathbf{h}})_{(i,j)}$ and setting to zero, one obtains

$$\frac{N_{(i,j),\mathbf{h}}}{(B_{\mathbf{h}})_{(i,j)}} = \frac{N_{i,0}}{1 - \sum_{\mathbf{h}\in C^0\cup\{0\}} \sum_{j=1}^{m} (B_{\mathbf{h}})_{(i,j)}} \equiv \Lambda_i. \tag{50}$$

To identify Λ_i, we rewrite the above equation as

$$\Lambda_i = \frac{N_{i,0}}{1 - \sum_{\mathbf{h}\in C^0\cup\{0\}} \sum_{j=1}^{m} \frac{N_{(i,j),\mathbf{h}}}{\Lambda_i}}, \tag{51}$$

which leads to

$$\Lambda_i = N_{i,0} + \sum_{\mathbf{h}\in C^0\cup\{0\}} \sum_{j=1}^{m} N_{(i,j),\mathbf{h}} = N_i \tag{52}$$

Combining Eqs. 50 and 52, we obtain that

$$(\hat{B}_{\mathbf{h}})_{(i,j)} = \frac{N_{(i,j),\mathbf{h}}}{N_i}. \tag{53}$$

Derivations of other quantities are similar and omitted.

Appendix B

First, we note that $\sum_{\mathbf{h}\in C^0\cup\{0\}} w_{\mathbf{h}} = 1$ and $B\mathbf{e} + \mathbf{b}^0 = \mathbf{e}$. Similar to general case in Appendix A, we get

$$L(\boldsymbol{\theta}, \mathbf{x}) = \sum_{i=1}^{m-1} \ln(\beta_i) D_i + \ln\left(1 - \sum_{j=1}^{m-1} \beta_j\right) D_m + \sum_{\mathbf{h}\in C^0\cup\{0\}} \sum_{i=1}^{m}\sum_{j=1}^{m} \ln(w_{\mathbf{h}} B_{(i,j)}) N_{(i,j),\mathbf{h}} + \left(\sum_{i=1}^{m}\left(\ln\left(1 - \sum_{j=1}^{m} B_{(i,j)}\right)\right) N_{i,0}\right). \tag{54}$$

Taking derivative with respect to $B_{(i,j)}$ and setting to zero, we obtain

$$\frac{\sum_{\mathbf{h}\in C^0\cup\{0\}} N_{(i,j),\mathbf{h}}}{B_{(i,j)}} = \frac{\sum_{\mathbf{h}\in C^0\cup\{0\}} N_{i,0}}{1 - \sum_{j=1}^{m} B_{(i,j)}} \equiv \Lambda_i, \tag{55}$$

which leads to $\Lambda_i = N_{(i,0)} + \sum_{j=1}^{m} \sum_{\mathbf{h}\in C^0} N_{(i,j),\mathbf{h}} = N_i$. Thus $\hat{B}_{(i,j)} = N_{(i,j)}/N_i$, where $N_{(i,j)} = \sum_{\mathbf{h}\in C^0} N_{(i,j),\mathbf{h}}$. The formula for $w_{\mathbf{h}}$ can be obtained similarly.

Appendix C

In this appendix, for the case with $K = 2$, we summarize formulas for computing the expectations under four types of conditions: i) $\{Y_1 = y_1, Y_2 = y_2\}$; ii) $\{Y_1 = y_1, Y_2 \geq y_2\}$; iii) $\{Y_1 \geq y_1, Y_2 = y_2\}$; and iv) $\{Y_1 \geq y_1, Y_2 \geq y_2\}$. For this case, $\Phi = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, and

$$\begin{aligned}
&\Omega_{(0,0)}(\mathbf{y}) = \{\mathbf{u} : \mathbf{u} \leq \mathbf{y}\};\\
&\Omega_{(0,1)}(\mathbf{y}) = \{\mathbf{u} : u_1 = 0, 1, ..., y_1; u_2 = y_2 + 1\};\\
&\Omega_{(1,0)}(\mathbf{y}) = \{\mathbf{u} : u_1 = y_1 + 1; u_2 = 0, 1, ..., y_2\};\\
&\Omega_{(1,1)}(\mathbf{y}) = \{(y_1 + 1, y_2 + 1)\}.
\end{aligned} \tag{56}$$

a) $\boldsymbol{\alpha}$ related vectors:

- For condition $\{Y_1 = y_1, Y_2 = y_2\}$, we have

$$\begin{aligned}
&\boldsymbol{\alpha}^{(0,0)}(0) = \boldsymbol{\beta}(I - B_0)^{-1};\\
&\boldsymbol{\alpha}^{(0,0)}(\mathbf{y}) = \sum_{\mathbf{h}\in C^0:\ \mathbf{h}\leq\mathbf{y}} \boldsymbol{\alpha}^{(0,0)}(\mathbf{y} - \mathbf{h}) B_{\mathbf{h}} (I - B_0)^{-1}, \text{ if } \mathbf{y} \neq 0.
\end{aligned} \tag{57}$$

- For condition $\{Y_1 = y_1, Y_2 \geq y_2\}$, we have

$$\begin{aligned}
\boldsymbol{\alpha}_{\geq}^{(0,1)}(0) &= \boldsymbol{\beta}\left(I - B_0 - \sum_{\mathbf{h}\in C^0:\ h_1=0} B_{\mathbf{h}}\right)^{-1};\\
\boldsymbol{\alpha}_{\geq}^{(0,1)}((y_1,0)) &= \left(\sum_{\mathbf{h}\in C^0:\ 1\leq h_1\leq y_1} \boldsymbol{\alpha}_{\geq}^{(0,1)}((y_1-h_1,0))B_{\mathbf{h}}\right)\\
&\quad\cdot\left(I - B_0 - \sum_{\mathbf{h}\in C^0:\ h_1=0} B_{\mathbf{h}}\right)^{-1},\ \text{if } y_1 \geq 1;\\
\boldsymbol{\alpha}_{\geq}^{(0,1)}(\mathbf{y}) &= \left(\sum_{\mathbf{h}\in C^0:\ h_1\leq y_1} \boldsymbol{\alpha}_{\geq}^{(0,1)}((\mathbf{y}-\mathbf{h})^+)B_{\mathbf{h}}\right)(I-B_0)^{-1},\ \text{if } y_2 \geq 1.
\end{aligned} \tag{58}$$

- For condition $\{Y_1 \geq y_1, Y_2 = y_2\}$, formulas for $\boldsymbol{\alpha}_{\geq}^{(1,0)}(\mathbf{y})$ are symmetric to that of $\boldsymbol{\alpha}^{(0,1)}\geq(\mathbf{y})$.
- For condition $\{Y_1 \geq y_1, Y_2 \geq y_2\}$, we have

$$\begin{aligned}
\boldsymbol{\alpha}_{\geq}^{(1,1)}(0) &= \boldsymbol{\beta}\left(I - B_0 - \sum_{\mathbf{h}\in C^0} B_{\mathbf{h}}\right)^{-1};\\
\boldsymbol{\alpha}_{\geq}^{(1,1)}(\mathbf{y}) &= \left(\sum_{\mathbf{h}\in C^0:\ (\mathbf{y}-\mathbf{h})^+\neq\mathbf{y}} \boldsymbol{\alpha}_{\geq}^{(1,1)}((\mathbf{y}-\mathbf{h})^+)B_{\mathbf{h}}\right)\\
&\quad\cdot\left(I - B_0 - \sum_{\mathbf{h}\in C^0:\ (\mathbf{y}-\mathbf{h})^+=\mathbf{y}} B_{\mathbf{h}}\right)^{-1},\ \text{if } \mathbf{y}\neq 0.
\end{aligned} \tag{59}$$

b) $\mathbf{p_Y}$ related vectors:

- For condition $\{Y_1 = y_1, Y_2 = y_2\}$, we have

$$\begin{aligned}
\mathbf{p}_{\mathbf{Y}}^{(0,0)}(0) &= (I-B_0)^{-1}\mathbf{b}^0;\\
\mathbf{p}_{\mathbf{Y}}^{(0,0)}(\mathbf{y}) &= \sum_{\mathbf{h}\in C^0:\ \mathbf{h}\leq\mathbf{y}} (I-B_0)^{-1}B_{\mathbf{h}}\mathbf{p}_{\mathbf{Y}}^{(0,0)}(\mathbf{y}-\mathbf{h}),\ \text{if } \mathbf{y}\neq 0.
\end{aligned} \tag{60}$$

- For condition $\{Y_1 = y_1, Y_2 \geq y_2\}$, we have

$$\begin{aligned}
\mathbf{p}_{\mathbf{Y},\geq}^{(0,1)}(0) &= \left(I - B_0 - \sum_{\mathbf{h}\in C^0:\ h_1=0} B_{\mathbf{h}}\right)^{-1}\mathbf{b}^0;\\
\mathbf{p}_{\mathbf{Y},\geq}^{(0,1)}((y_1,0)) &= \left(I - B_0 - \sum_{\mathbf{h}\in C^0:\ h_1=0} B_{\mathbf{h}}\right)^{-1}\\
&\quad\cdot\left(\sum_{\mathbf{h}\in C^0:\ 1\leq h_1\leq y_1} B_{\mathbf{h}}\mathbf{p}_{\mathbf{Y},\geq}^{(0,1)}((y_1-h_1,0))\right),\ \text{if } y_1\geq 1;\\
\mathbf{p}_{\mathbf{Y},\geq}^{(0,1)}(\mathbf{y}) &= (I-B_0)^{-1}\left(\sum_{\mathbf{h}\in C^0:\ h_1\leq y_1} B_{\mathbf{h}}\mathbf{p}_{\mathbf{Y},\geq}^{(0,1)}((\mathbf{y}-\mathbf{h})^+)\right),\ \text{if } y_2\geq 1.
\end{aligned} \tag{61}$$

- For condition $\{Y_1 \geq y_1, Y_2 = y_2\}$, formulas for $\mathbf{p}_{\mathbf{Y},\geq}^{(1,0)}(\mathbf{y})$ are symmetric to that of $\mathbf{p}_{\mathbf{Y},\geq}^{(0,1)}(\mathbf{y})$.

- For condition $\{Y_1 \geq y_1, Y_2 \geq y_2\}$, we have

$$
\begin{aligned}
\mathbf{p}^{(1,1)}_{\mathbf{Y},\geq}(0) &= \left(I - B_0 - \sum_{\mathbf{h}\in C^0} B_{\mathbf{h}}\right)^{-1} \mathbf{b}^0; \\
\mathbf{p}^{(1,1)}_{\mathbf{Y},\geq}(\mathbf{y}) &= \left(I - B_0 - \sum_{\mathbf{h}\in C^0:\ (\mathbf{y}-\mathbf{h})^+=\mathbf{y}} B_{\mathbf{h}}\right)^{-1} \\
&\quad \cdot \left(\sum_{\mathbf{h}\in C^0:\ (\mathbf{y}-\mathbf{h})^+\neq\mathbf{y}} B_{\mathbf{h}} \mathbf{p}^{(1,1)}_{\mathbf{Y},\geq}((\mathbf{y}-\mathbf{h})^+)\right),\ \text{if } \mathbf{y} \neq 0.
\end{aligned}
\tag{62}
$$

c) Conditional expectataions

- For condition $\{Y_1 = y_1, Y_2 = y_2\}$, we have

$$
\begin{aligned}
&\mathbb{E}[D_i|\mathbf{Y}=\mathbf{y}] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}^{(0,0)}_{\mathbf{Y}}(\mathbf{y})} \beta_i p^{(0,0)}_{\mathbf{Y},i}(\mathbf{y}); \\
&\mathbb{E}[N_i|\mathbf{Y}=\mathbf{y}] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}^{(0,0)}_{\mathbf{Y}}(\mathbf{y})} \sum_{\mathbf{u}:\ \mathbf{u}\leq\mathbf{y}} \alpha^{(0,0)}_i(\mathbf{u}) p^{(0,0)}_{\mathbf{Y},i}(\mathbf{y}-\mathbf{u}); \\
&\mathbb{E}[N_{(i,j),0}|\mathbf{Y}=\mathbf{y}] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}^{(0,0)}\mathbf{Y}(\mathbf{y})} \sum_{\mathbf{u}:\ \mathbf{u}\leq\mathbf{y}} \alpha^{(0,0)}_i(\mathbf{u})(B_0)_{(i,j)} p^{(0,0)}_{\mathbf{Y},j}(\mathbf{y}-\mathbf{u}); \\
&\mathbb{E}[N_{(i,j),\mathbf{h}}|\mathbf{Y}=\mathbf{y}] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}^{(0,0)}\mathbf{Y}(\mathbf{y})} \sum_{\mathbf{u}:\ \mathbf{u}\leq\mathbf{y}-\mathbf{h}} \alpha^{(0,0)}_i(\mathbf{u})(B_{\mathbf{h}})_{(i,j)} p^{(0,0)}_{\mathbf{Y},j}(\mathbf{y}-\mathbf{u}-\mathbf{h}); \\
&\mathbb{E}[N_{i,0}|\mathbf{Y}=\mathbf{y}] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}^{(0,0)}\mathbf{Y}(\mathbf{y})} \alpha^{(0,0)}_i(\mathbf{y})(\mathbf{b}^0)_i.
\end{aligned}
\tag{63}
$$

- For condition $\{Y_1 = y_1, Y_2 \geq y_2\}$, we have

$$
\begin{aligned}
&\mathbb{E}[D_i|Y_1=y_1, Y_2\geq y_2] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}^{(0,1)}_{\mathbf{Y}},\geq(\mathbf{y})} \beta_i p^{(0,1)}_{\mathbf{Y},i}(\mathbf{y}); \\
&\mathbb{E}[N_i|Y_1=y_1, Y_2\geq y_2] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}^{(0,1)}_{\mathbf{Y}}(\mathbf{y})} \left(\sum_{\mathbf{u}:\ \mathbf{u}\leq\mathbf{y}} \alpha^{(0,0)}_i(\mathbf{u}) p^{(0,1)}_{\mathbf{Y},\geq,i}(\mathbf{y}-\mathbf{u})\right. \\
&\qquad\qquad \left. + \sum_{u_1=0}^{y_1} \alpha^{(0,1)}_{\geq,i}((u_1, y_2+1)) p^{(0,1)}_{\mathbf{Y},\geq,i}((y_1-u_1, 0))\right); \\
&\mathbb{E}[N_{(i,j),0}|Y_1=y_1, Y_2\geq y_2] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}^{(0,1)}_{\mathbf{Y}}(\mathbf{y})} \left(\sum_{\mathbf{u}:\ \mathbf{u}\leq\mathbf{y}} \alpha^{(0,0)}_i(\mathbf{u})(B_0)_{(i,j)} p^{(0,1)}_{\mathbf{Y},\geq,j}(\mathbf{y}-\mathbf{u})\right. \\
&\qquad\qquad \left. + \sum_{u_1=0}^{y_1} \alpha^{(0,1)}_{\geq,i}((u_1, y_2+1))(B_0)_{(i,j)} p^{(0,1)}_{\mathbf{Y},\geq,j}((y_1-u_1, 0))\right); \\
&\mathbb{E}[N_{(i,j),\mathbf{h}}|Y_1=y_1, Y_2\geq y_2] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}^{(0,1)}_{\mathbf{Y}}(\mathbf{y})} \left(\sum_{\mathbf{u}:\ \mathbf{u}\leq\mathbf{y}-\mathbf{h}} \alpha^{(0,0)}_i(\mathbf{u})(B_{\mathbf{h}})_{(i,j)} p^{(0,1)}_{\mathbf{Y},\geq,j}(\mathbf{y}-\mathbf{u}-\mathbf{h})\right. \\
&\qquad\qquad \left. + \sum_{u_1=0}^{y_1-h_1} \alpha^{(0,1)}_{\geq,i}((u_1, y_2-h_2+1))(B_{\mathbf{h}})_{(i,j)} p^{(0,1)}_{\mathbf{Y},\geq,j}((y_1-h_1-u_1, 0))\right); \\
&\mathbb{E}[N_{i,0}|Y_1=y_1, Y_2\geq y_2] = \tfrac{1}{\boldsymbol{\beta}\mathbf{p}^{(0,1)}_{\mathbf{Y}}(\mathbf{y})} \alpha^{(0,1)}_{\geq,i}(\mathbf{y})(\mathbf{b}^0)_i.
\end{aligned}
\tag{64}
$$

- The case with $\{Y_1 \geq y_1, Y_2 = y_2\}$ is symmetric to the case with $\{Y_1 = y_1, Y_2 \geq y_2\}$.

- For condition $\{Y_1 \geq y_1, Y_2 \geq y_2\}$, we have

$$
\begin{aligned}
\mathbb{E}[D_i|Y_1 \geq y_1, Y_2 \geq y_2] &= \frac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}^{(1,1)}(\mathbf{y})}\beta_i p_{\mathbf{Y},\geq,i}^{(1,1)}(\mathbf{y});\\
\mathbb{E}[N_i|Y_1 \geq y_1, Y_2 \geq y_2] &= \frac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}^{(1,1)}(\mathbf{y})}\left(\sum_{\mathbf{u}:\ \mathbf{u}\leq\mathbf{y}} \alpha_i^{(0,0)}(\mathbf{u}) p_{\mathbf{Y},\geq,i}^{(1,1)}(\mathbf{y}-\mathbf{u})\right.\\
&\quad + \sum_{u_1=0}^{y_1} \alpha_{\geq,i}^{(0,1)}((u_1, y_2+1)) p_{\mathbf{Y},\geq,i}^{(1,1)}((y_1-u_1, 0))\\
&\quad + \sum_{u_2=0}^{y_2} \alpha_{\geq,i}^{(1,0)}((y_1+1, u_2)) p_{\mathbf{Y},\geq,i}^{(1,1)}((0, y_2-u_2))\\
&\quad \left. + \alpha_{\geq,i}^{(1,1)}((y_1+1, y_2+1)) p_{\mathbf{Y},\geq,i}^{(1,1)}((0,0))\right);\\
\mathbb{E}[N_{(i,j),0}|Y_1 \geq y_1, Y_2 \geq y_2] &= \frac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}^{(1,1)}(\mathbf{y})}\left(\sum_{\mathbf{u}:\ \mathbf{u}\leq\mathbf{y}} \alpha_i^{(0,0)}(\mathbf{u})(B_0)_{(i,j)} p_{\mathbf{Y},\geq,j}^{(1,1)}(\mathbf{y}-\mathbf{u})\right.\\
&\quad + \sum_{u_1=0}^{y_1} \alpha_{\geq,i}^{(0,1)}((u_1, y_2+1))(B_0)_{(i,j)} p_{\mathbf{Y},\geq,j}^{(1,1)}((y_1-u_1, 0))\\
&\quad + \sum_{u_2=0}^{y_2} \alpha_{\geq,i}^{(1,0)}((y_1+1, u_2))(B_0)_{(i,j)} p_{\mathbf{Y},\geq,j}^{(1,1)}((0, y_2-u_2))\\
&\quad \left. + \alpha_{\geq,i}^{(1,1)}((y_1+1, y_2+1))(B_0)_{(i,j)} p_{\mathbf{Y},\geq,j}^{(1,1)}((0,0))\right);\\
\mathbb{E}[N_{(i,j),\mathbf{h}}|Y_1 \geq y_1, Y_2 \geq y_2] &= \frac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}^{(1,1)}(\mathbf{y})}\left(\sum_{\mathbf{u}:\ \mathbf{u}\leq\mathbf{y}-\mathbf{h}} \alpha_i^{(0,0)}(\mathbf{u})(B_{\mathbf{h}})_{(i,j)} p_{\mathbf{Y},\geq,j}^{(1,1)}(\mathbf{y}-\mathbf{u}-\mathbf{h})\right.\\
&\quad + \sum_{u_1=0}^{y_1-h_1} \alpha_{\geq,i}^{(0,1)}((u_1, y_2-h_2+1))(B_{\mathbf{h}})_{(i,j)} p_{\mathbf{Y},\geq,j}^{(0,1)}((y_1-h_1-u_1, 0))\\
&\quad + \sum_{u_2=0}^{y_2-h_2} \alpha_{\geq,i}^{(1,0)}((y_1-h_1+1, u_2))(B_{\mathbf{h}})_{(i,j)} p_{\mathbf{Y},\geq,j}^{(1,1)}((0, y_2-h_2-u_2))\\
&\quad \left. + \alpha_{\geq,i}^{(1,1)}((y_1-h_1+1, y_2-h_2+1))(B_{\mathbf{h}})_{(i,j)} p_{\mathbf{Y},\geq,j}^{(1,1)}((0,0))\right);\\
\mathbb{E}[N_{i,0}|Y_1 \geq y_1, Y_2 \geq y_2] &= \frac{1}{\boldsymbol{\beta}\mathbf{p}_{\mathbf{Y}}^{(1,1)}(\mathbf{y})}\alpha_{\geq,i}^{(1,1)}(\mathbf{y})(\mathbf{b}^0)_i.
\end{aligned}
\tag{65}
$$

References

Asmussen S, Nerman O, Olsson M (1996) Fitting phase-type distributions via the EM-algorithm. Scand J Stat 23(4):419–441

Assaf D, Langberg NA, Savits TH, Shaked M (1984) Multivariate phase-type distributions. Oper Res 32(3):688–702

Basawa IV, Rao PBLS (1980) Statistical inference for stochastic processes. Academic Press, London ; Toronto

Breuer L (2003) An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. Ann Oper Res 112:123–138

Buchholz P, Kemper P, Kriege J (2010) Multi-class Markovian arrival processes and their parameter fitting. Perform Eval 67(11):1092–1106

Cai J, Li H (2005) Conditional tail expectations for multivariate phase-type distributions. J Appl Probab 42(3):810–825

David Cummins J, Wiltbank LJ (1983) Estimating the total claims distribution using multivariate frequency and severity distributions. The Journal of Risk and Insurance 50(3):377–403

He Q-M, Neuts MF (1998) Markov chains with marked transitions. Stoch Process Appl 74(1):37–52

He Q-M, Ren J (2014) Analysis of a multivariate claim process. Methodol Comput Appl Probab. (Accepted)

Herbertsson A (2011) Modelling default contagion using multivariate phase-type distributions. Rev Deriv Res 14(1):1–36

Kulkarni VG (1989) A new class of multivariate phase type distributions. Oper Res 37(1):151–158

Olsson M (1996) Estimation of phase-type distributions from censored data. Scand J Stat 23:443–460

Ren J (2010) Recursive formulas for compound phase distributions. ASTIN Bulletin-Actuarial Studies in non LifeInsurance 40(2):615

Jeff Wu CF (1983) On the convergence properties of the EM algorithm. Ann Stat 11(1):95–103

Hassan Zadeh A., Bilodeau M (2013) Fitting bivariate losses with phase type distributions. Scand Actuar J 2013(4):241–262