Scandinavian Actuarial Journal

ISSN: 0346-1238 (Print) 1651-2030 (Online) Journal homepage: http://www.tandfonline.com/loi/sact20

Fitting bivariate losses with phase-type distributions

Amin Hassan Zadeh & Martin Bilodeau

To cite this article: Amin Hassan Zadeh & Martin Bilodeau (2013) Fitting bivariate losses with phase-type distributions, Scandinavian Actuarial Journal, 2013:4, 241-262, DOI: 10.1080/03461238.2011.602196

To link to this article: http://dx.doi.org/10.1080/03461238.2011.602196

Scandinavian Actuarial Journal, 2013
Vol. 2013, No. 4, 241–262, http://dx.doi.org/10.1080/03461238.2011.602196

Taylor & Francis
Taylor & Francis Group

Original Article

Fitting bivariate losses with phase-type distributions

AMIN HASSAN ZADEH[a]* AND MARTIN BILODEAU[b]

[a]Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario Canada N6A 5B7

[b]Département de mathématiques et de statistique, Université de Montréal, Montréal, Québec Canada H3C 3J7

(*Accepted 1 June 2011*)

Maximum likelihood estimation and (parametric bootstrap) goodness-of-fit test are considered for bivariate phase-type distributions introduced by Assaf and Colleagues. In a special case, the dependence structure of bivariate phase-type distributions is revealed. The results are used to fit a real bi-dimensional data set related to insurance losses (LOSS) and allocated loss adjustment expenses (ALAE). The fitted bivariate phase-type is used to obtain conditional quantiles and mean of ALAE for a given value of LOSS. The bivariate phase-type distribution meets all the requirements listed in the study by Klugman and Parsa.

Keywords: bivariate phase-type distributions; dependent losses; continuous-time Markov chain; copulas; EM algorithm; parametric bootstrap goodness-of-fit test

1. Introduction

Phase-type (PH) random variables are defined as the time until absorption in a set of absorbing states in a continuous time Markov chain environment. Coxian, Erlang-*n*, hyperexponential and mixture of Erlang-*n* distributions are special cases of PH random variables. Neuts (1981) defines the PH random variable and establishes its theoretical properties. PH distributions are dense among all distributions with positive support. In addition, they have density, Laplace transform and all their moments in closed form and thus, various probability quantities can be obtained easily. Despite the interesting properties of PH variables, some difficulties arise in statistical estimation. Nonuniqueness of representations in some PH models, as discussed in O'Cinneide (1989), and over-parametrization is briefly mentioned in Asmussen *et al.* (1996). Asmussen *et al.* (1996) study the parameter estimation by the EM algorithm, as well as fitting other densities on the positive line with PH distributions. In Assaf and Levikson (1982) some properties of PH variables in reliability are investigated.

Phase-type distributions have many appealing features which make it very attractive to actuaries. Many systems with PH input variables yield an output variable which is also

*Corresponding author. E-mail: ahassan@stats.uwo.ca

PH. For example, in Drekic *et al.* (2004), in the Sparre Andersen renewal models with PH distributed claims, the distribution of deficit at ruin is PH with an interpretable representation. Asmussen (2000) applies PH distributions to risk theory. Li and Garrido (2004) consider ruin probability in risk theory for Erlang-*n* distributions, a special case of PH distributions. Tractability and ease of calculations of PH distributions incited some researchers to generalize the notion to the multivariate case.

In the study by Assaf *et al.* (1984), a multivariate PH distribution is defined. In Kulkarni (1989), a new class of multivariate PH distribution is introduced. In the multivariate case, the structure of dependence under some conditions is studied by Li (2003). The conditional tail expectation for multivariate PH distributions is obtained in Cai and Li (2005a). Multivariate PH risk model is the subject of Cai and Li (2005b).

This paper is organized as follows. Univariate and multivariate PH variables, with their properties, are briefly defined in Section 2. Section 3 covers parameter estimation of bivariate PH (BPH) distributions via the EM algorithm. In Section 4, a method to simulate a BPH distribution is used in a small simulation study on the bias and standard deviation of the EM estimator. A (parametric bootstrap) goodness-of-fit test for BPH distributions is proposed in Section 5. Section 6 includes a data analysis of the ALAE data by fitting a BPH distribution. It also gives expressions for the conditional quantiles and conditional mean. This article extends the works of Asmussen *et al.* (1996), Assaf *et al.* (1984) and Åhlström *et al.* (1999) to problems of statistical nature in BPH distributions, namely, the statistical estimation by the EM algorithm and a parametric bootstrap goodness-of-fit test. To our knowledge, this is the first paper in which BPH distributions are applied in the context of a real data analysis.

2. Preliminaries

Consider $\{J_t, t \geq 0\}$ a right continuous Markov process on the finite state space $\Gamma = \{1, 2, \ldots, m, m+1\}$ with initial probability vector α and infinitesimal generator matrix A. Suppose that Γ_1 and Γ_2 are two nonempty stochastically closed subsets ($E \subset \Gamma$ is said to be stochastically closed if, once J_t has entered E, it never leaves) of Γ such that $\Gamma_1 \cap \Gamma_2 = \{m+1\}$. States $1, \ldots, m$ are transient and state $m+1$ is absorbing. Hence, absorption into state $m+1$ is certain to happen. As a convention, all vectors are column vectors and superscript T denotes the transpose of a matrix. The matrix A can be written as

$$\mathbf{A} = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0}^T & 0 \end{pmatrix}, \tag{1}$$

where the matrix $\mathbf{T} = (t_{ij})$ is $m \times m$ and $\mathbf{t} = (t_j)$ is an m-dimensional vector. These elements satisfy $t_{ii} < 0$, $i = 1,\ldots,m, t_{ij} \geq 0$, $i \neq j$, and $\mathbf{Te} + \mathbf{t} = \mathbf{0}$, where $\mathbf{e}$ is a vector of ones. States 1,

..., m are transient if and only if $\mathbf{T}$ is nonsingular (see Neuts 1994). In this article, we always suppose that $\alpha_{m+1}=0$, and hence, $\boldsymbol{\alpha}$ can be written as

$$\boldsymbol{\alpha}=\begin{pmatrix}\boldsymbol{\pi}\\0\end{pmatrix}.$$

Let X_1 and X_2 be the times until absorption in Γ_1 and Γ_2, respectively. We call the joint distribution of (X_1, X_2) a bivariate PH (BPH) distribution with representation ($\boldsymbol{\pi}$, $\mathbf{T}$, Γ_1, Γ_2). The marginal distributions of X_1 and X_2 have univariate PH distributions.

If (X_1, X_2) has a BPH distribution, by using Markov chain theory, it is shown in Assaf *et al.* (1984) that the joint survival function is

$$\mathbb{P}(X_1>x_1,X_2>x_2)=\begin{cases}\boldsymbol{\pi}^{\mathrm{T}}e^{\mathbf{T}x_1}\mathbf{g}_1e^{\mathbf{T}(x_2-x_1)}\mathbf{g}_2\mathbf{e}, & x_2\geq x_1\geq 0\\ \boldsymbol{\pi}^{\mathrm{T}}e^{\mathbf{T}x_2}\mathbf{g}_2e^{\mathbf{T}(x_1-x_2)}\mathbf{g}_1\mathbf{e}, & x_1\geq x_2\geq 0\end{cases}, \tag{2}$$

where $\mathbf{g}_k$, $k=1, 2$, is an $m\times m$ diagonal matrix whose ith diagonal element is 1 if $i\in\Gamma_k^c$, and 0 otherwise. Assaf *et al.* (1984) also provide the Laplace transform

$$\begin{aligned}\phi(u_1,u_2)&=\mathbb{E}\left[e^{-u_1X_1-u_2X_2}\right]\\&=\boldsymbol{\pi}^{\mathrm{T}}[(u_1+u_2)\mathbf{I}-\mathbf{T}]^{-1}\{\mathbf{G}_2[u_1\mathbf{I}-\mathbf{T}]^{-1}\mathbf{T}\mathbf{g}_1\\&\quad+\mathbf{G}_1[u_2\mathbf{I}-\mathbf{T}]^{-1}\mathbf{T}\mathbf{g}_2-[\mathbf{T}\mathbf{g}_1\mathbf{g}_2-\mathbf{G}_1-\mathbf{G}_2]\}\mathbf{e}\end{aligned}$$

where $\mathbf{G}_k=\mathbf{T}\mathbf{g}_k-\mathbf{g}_k\mathbf{T}$, $k=1, 2$, is the commutator.

In general, the joint distribution F has a singular component on the set $x_1=x_2$, which can be avoided by supposing that $t_i=0$, for $i\in\Gamma_1^c\cap\Gamma_2^c$, see Assaf *et al.* (1984). Hereafter, we also suppose that $\pi_i=0$, for $i\in\Gamma_1\cup\Gamma_2$, and as a result $\mathbb{P}(X_1>0,X_2>0)=1$. By imposing this structure on the initial probability vector $\boldsymbol{\pi}$ we have that $\mathbf{g}_k\boldsymbol{\pi}=\boldsymbol{\pi}$, $k=1, 2$, and hence, the marginal survival functions can be obtained easily from Equation (2).

Using the fact that $\mathrm{d}e^{\mathbf{T}x}/\mathrm{d}x=\mathbf{T}e^{\mathbf{T}x}=e^{\mathbf{T}x}\mathbf{T}$, the density of the absolutely continuous component can be derived from Equation (2):

$$f(x_1,x_2)=\begin{cases}\boldsymbol{\pi}^{\mathrm{T}}e^{\mathbf{T}x_1}\mathbf{G}_1e^{\mathbf{T}(x_2-x_1)}\mathbf{T}\mathbf{g}_2\mathbf{e}, & x_2\geq x_1\geq 0\\ \boldsymbol{\pi}^{\mathrm{T}}e^{\mathbf{T}x_2}\mathbf{G}_2e^{\mathbf{T}(x_1-x_2)}\mathbf{T}\mathbf{g}_1\mathbf{e}, & x_1\geq x_2\geq 0\end{cases}. \tag{3}$$

The singular component on $x_1=x_2$ may be useful in some applications related to life insurance. It is given in Assaf *et al.* (1984) with a further simplification as

$$\begin{aligned}\mathbb{P}(X_1=X_2>x)&=\boldsymbol{\pi}^{\mathrm{T}}e^{\mathbf{T}x}\mathbf{T}^{-1}\left[\mathbf{T}\mathbf{g}_1\mathbf{g}_2-\mathbf{G}_1-\mathbf{G}_2\right]\mathbf{e}\\&=\boldsymbol{\pi}^{\mathrm{T}}e^{\mathbf{T}x}\mathbf{T}^{-1}\mathbf{g}_1\mathbf{g}_2\mathbf{T}\mathbf{e}.\end{aligned} \tag{4}$$

Hence,

$$\mathbb{P}(X_1=X_2)=\boldsymbol{\pi}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{g}_1\mathbf{g}_2\mathbf{T}\mathbf{e},$$

is obtained with the evaluation of Equation (4) at $x=0$. Thus, with a correction to the statement made in Assaf *et al.* (1984), the singular part is zero if and only if

$[\mathbf{T}\mathbf{g}_1\mathbf{g}_2 - \mathbf{G}_1 - \mathbf{G}_2]\ \mathbf{e} = 0$, which is equivalent to $t_i = 0,\ i \in \Gamma_1^c \cap \Gamma_2^c$. When $\pi_i = 0$, for $i \in \Gamma_1 \cup \Gamma_2$ one can assume without loss of generality that

$$\boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{\pi}^{(1,2)} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

$$\mathbf{T} = \begin{pmatrix} \mathbf{A}^{(1,2)} & \mathbf{B}^{(1)} & \mathbf{B}^{(2)} \\ \mathbf{0} & \mathbf{A}^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}^{(2)} \end{pmatrix}$$

where the partition corresponds to the three subsets $\Gamma_1^c \cap \Gamma_2^c$, $\Gamma_1 \backslash \{m+1\}$, and $\Gamma_2 \backslash \{m+1\}$. Then, the joint density Equation (3) can be rewritten, with a correction to Assaf *et al.* (1984), as

$$f(x_1, x_2) = \begin{cases} -\boldsymbol{\pi}^{(1,2)T} e^{\mathbf{A}^{(1,2)} x_1} \mathbf{B}^{(1)} e^{\mathbf{A}^{(1)}(x_2 - x_1)} \mathbf{A}^{(1)} \mathbf{e}, & x_2 \geq x_1 \geq 0, \\ -\boldsymbol{\pi}^{(1,2)T} e^{\mathbf{A}^{(1,2)} x_2} \mathbf{B}^{(2)} e^{\mathbf{A}^{(2)}(x_1 - x_2)} \mathbf{A}^{(2)} \mathbf{e}, & x_1 \geq x_2 \geq 0. \end{cases} \tag{5}$$

The multivariate version of phase-type random variables (MPH) is also defined in Assaf *et al.* (1984). Suppose that $\{J_t,\ t \geq 0\}$ is a right continuous Markov chain on a finite state space Γ. Let $\Gamma_1, \ldots, \Gamma_n$ be nonempty stochastically closed subsets of Γ, such that $\cap_{i=1}^n \Gamma_i$ has just one member, namely $m+1$, and absorption into $m+1$ is certain. The matrix $\mathbf{A}$ is still the infinitesimal generator as in Equation (1). Define $X_k = \inf\{t \geq 0 \mid J_t \in \Gamma_k\}$, $k = 1,2,\ldots,n$.

We assume that $\pi_i = 0$ for $i \in \cup_{i=1}^n \Gamma_i$. The joint MPH distribution of $(X_1 \ldots X_n)$ has the representation $(\boldsymbol{\pi}, \mathbf{T}, \Gamma_1, \ldots, \Gamma_n)$. For $0 < x_1 \leq x_2 \leq \ldots \leq x_n$,

$$\begin{aligned} S(x_1, \ldots, x_n) &= P(X_1 > x_1, \ldots, X_n > x_n) \\ &= \boldsymbol{\pi}^{\mathrm{T}} e^{\mathbf{T} x_1} \mathbf{g}_1 e^{\mathbf{T}(x_2 - x_1)} \mathbf{g}_2 \ldots e^{\mathbf{T}(x_n - x_{n-1})} \mathbf{g}_n \mathbf{e}. \end{aligned}$$

The MPH distribution is absolutely continuous if and only if $t_{ij} = 0$, whenever $i \in \Gamma_k^c \cap \Gamma_l^c$ and $j \in \Gamma_k \cap \Gamma_l$, where $k \neq l$. Laplace transform is given in Assaf *et al.* (1984) and it can be used to calculate all the moments. As in the univariate case, MPH has the closure property. Let $\mathbf{T} = (T_1, \ldots, T_n)$ and $\mathbf{W} = (W_1, \ldots, W_m)$ be independent MPH random vectors. Then, the conjunction $(\mathbf{T}, \mathbf{W}) = (T_1, \ldots, T_n, W_1, \ldots, W_m)$ is an MPH random vector. See Marshall and Shaked (1986) for a proof. Moreover, MPH distributions are closed under finite mixture and convolution, see Assaf *et al.* (1984) and Kulkarni (1989). Cai and Li (2005b) give an explicit representation for the convolution of MPH distributions. As for univariate PH, the class of n-dimensional MPH distributions is dense in the set of all distributions on $[0, \infty]^n$, see Assaf *et al.* (1984) for a proof.

3. EM algorithm

3.1. General EM algorithm

The EM (Expectation-Maximization) algorithm of Dempster *et al.* (1977) is a general iterative method for finding the maximum likelihood estimate of the parameters, when the data are incomplete or have missing values. It finds its usefulness when the likelihood function of the incomplete (observed) data is intractable but that of the complete (unobserved or missing) data is of a simpler form which can be analytically optimized. The EM algorithm is not guaranteed to find the global maximum, it may converge to a local maximum or even a saddle point of the likelihood surface, see Wu (1983).

Assume that the data $\mathbf{x}$ is observed and generated by some distributions, say $f(\mathbf{x}|\phi)$ with log-likelihood function $L(\phi) = \log f(\mathbf{x}|\phi)$. We call $\mathbf{x}$ the incomplete data and refer to $L(\phi)$ as the incomplete log-likelihood function. Suppose that an unobserved (complete) data $\mathbf{y}$, where $\mathbf{x} = \mathbf{x}(\mathbf{y})$, has pdf $g(\mathbf{y}|\phi)$. Assume

$$Q(\phi|\phi^{(p)}) = \mathbb{E}_{\phi^{(p)}}[\log g(\mathbf{y}|\phi)|\mathbf{x}],$$

exists for all pairs $(\phi^{(p)}, \phi)$. This notation is for the conditional expectation of $\log g(\mathbf{y}|\phi)$ given $\mathbf{x}$ and the current value $\phi^{(p)}$ of the parameter. Another notation often used is $\mathbb{E}[\log g(\mathbf{y}|\phi)|\mathbf{x}, \phi^{(p)}]$. The EM iteration $\phi^{(p)} \rightarrow \phi^{(p+1)}$ is defined as follows:

E-step: Compute $Q(\phi|\phi^{(p)})$.
M-step: Find $\phi^{(p+1)}$ that maximizes $Q(\phi|\phi^{(p)})$ over ϕ.

Simplifications occur when the complete data density function is a member of the exponential family

$$g(\mathbf{y}|\phi) = b(\mathbf{y}) \exp[\phi^{\mathrm{T}} \mathbf{t}(\mathbf{y})] / a(\phi), \tag{6}$$

where ϕ is the vector parameter, $\mathbf{t}(\mathbf{y})$ is the vector of complete data sufficient statistic. If Equation (6) holds, Dempster *et al.* (1977) present simplified expressions for the E and M steps:
E-step: Estimate the complete data sufficient statistics $\mathbf{t}(\mathbf{x})$ by finding

$$\mathbf{t}^{(p)} = \mathbb{E}_{\phi^{(p)}}[\mathbf{t}(\mathbf{y})|\mathbf{x}].$$

M-step: Determine $\phi^{(p+1)}$ as the solution for ϕ of the equation

$$\mathbb{E}_{\phi}[\mathbf{t}(\mathbf{y})] = \mathbf{t}^{(p)}.$$

3.2. EM algorithm for BPH distribution

EM algorithms in a Markovian chain environment are not new. There are some works done in this context, including those of Breuer (2002), Ryden (1996), Roberts *et al.* (2006), Roberts and Ephraim (2008), Asmussen *et al.* (1996) and Åhlström *et al.* (1999). The

latter one is the EM algorithm for a special case of BPH distributions satisfying $x_1 < x_2$ and with censoring conditions on the data. Our work compared to Åhlström *et al.* (1999) might be considered as an incremental work to general BPH distributions.

By definition a BPH random vector has two components representing the times until absorption in two stochastically closed subsets Γ_1 and Γ_2. This can be considered as an incomplete data in the sense that they only provide information about the time of hitting Γ_1 and Γ_2, not about the whole path of J_t. The initial state, the states that have been visited, and the time spent in each visited state are not observed. Hence, the hidden information can help to maximize the incomplete likelihood function which is untractable. For the case $x_1 < x_2$, the complete path can be formulated by the embedded Markov chain of visited states

$$i_0, i_1, \ldots, i_{m_1-1}, \ldots, i_{m_2-1}(i_{m_2} = m+1),$$

and the sojourn times

$$s_0, s_1, \ldots, s_{m_1-1}, \ldots, s_{m_2-1}(s_{m_2} = \infty),$$

where m_1 is the number of jumps until hitting Γ_1 and m_2 is the number of jumps until hitting the absorbing state $m+1$.

Given a realized observation (x_1, x_2) of the BPH distribution, a complete observation of the process J_t on the interval $(0, x_2]$ is represented by

$$\mathbf{y} = \left(i_0, \ldots, i_{m_1-1}, \ldots, i_{m_2-1}, s_0, \ldots, s_{m_1-1}, \ldots, s_{m_2-1}\right),$$

where, $x_1 = s_0 + \ldots + s_{m_1-1}$ and $x_2 = s_0 + \ldots + s_{m_2-1}$. To get the probability density function of $\mathbf{y}$, one needs the probability p_{jk} of jumping from j to k which is given by

$$p_{jk} = \mathbb{P}(i_{n+1} = k | i_n = j) = \begin{cases} \frac{t_{jk}}{-t_{jj}}, & j \neq k; j, k = 1, \ldots, m, \\ \frac{t_j}{-t_{jj}}, & k = m+1;\ j = 1, \ldots, m. \end{cases}$$

The density of $\mathbf{y}$ can be derived by Markov chain properties and considering that the time spent in each state i has an exponential distribution with mean $1/\lambda_i$, where $\lambda_i = -t_{ii}$, as in Asmussen *et al.* (1996). Thus,

$$g(\mathbf{y}|\boldsymbol{\theta}) = \pi_{i_0} \exp\left\{-\lambda_{i_0} s_0\right\} t_{i_0, i_1} \ldots \exp\left\{-\lambda_{i_{m_2-1}} s_{m_2-1}\right\} t_{i_{m_2-1}},$$

where $\boldsymbol{\theta} = (\boldsymbol{\pi}, \mathbf{T})$ is the parameter.

Let $J_t^{[1]}, \ldots, J_t^{[n]}$ be n independent realizations of the process. This gives n embedded Markov chains

$$i_0^{[v]}, \ldots, i_{m_1^{[v]}-1}^{[v]}, \ldots, i_{m_2^{[v]}-1}^{[v]},$$

with corresponding holding times

$$s_0^{[v]}, \ldots, s_{m_1^{[v]}-1}^{[v]}, \ldots, s_{m_2^{[v]}-1}^{[v]}, \quad v = 1, \ldots, n.$$

The complete data become $\mathbf{y} = (\mathbf{y}^{[1]}, \ldots, \mathbf{y}^{[n]})$, where

$$\mathbf{y}^{[v]} = \left(i_0^{[v]}, \ldots, i_{m_1^{[v]}-1}^{[v]}, \ldots, i_{m_2^{[v]}-1}^{[v]}, s_0^{[v]}, \ldots, s_{m_1^{[v]}-1}^{[v]}, \ldots, s_{m_2^{[v]}-1}^{[v]} \right), \quad v = 1, \ldots, n.$$

The observed incomplete data are the following function of the complete data

$$\mathbf{x}^{[v]} = \left(x_1^{[v]}, x_2^{[v]} \right) = \left(s_0^{[v]} + \ldots + s_{m_1^{[v]}-1}^{[v]}, s_0^{[v]} + \ldots + s_{m_2^{[v]}-1}^{[v]} \right).$$

Define

$$B_i^{[v]} = 1\left\{ i_0^{[v]} = i \right\}$$

$$Z_i^{[v]} = \sum_{k=0}^{m_2^{[v]}-1} 1\left\{ i_k^{[v]} = i \right\} s_k^{[v]},$$

$$N_{ij}^{[v]} = \sum_{k=0}^{m_2^{[v]}-1} 1\left\{ i_k^{[v]} = i, i_{k+1}^{[v]} = j \right\}.$$

Let

$$B_i = \sum_{v=1}^{n} B_i^{[v]},$$

$$Z_i = \sum_{v=1}^{n} Z_i^{[v]},$$

$$N_{ij} = \sum_{v=1}^{n} N_{ij}^{[v]},$$

be the number of Markov processes starting from state i, the total time spent in each state i, and the number of jumps from state i to state j, respectively.

Then, the density of the complete data $\mathbf{y}$ is the product of n densities as in Equation (7)

$$g(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^{m} \left[\pi_i^{B_i} \exp\{t_{ii} Z_i\} \prod_{j=1, j\neq i}^{m+1} t_{ij}^{N_{ij}} \right], \tag{8}$$

where $t_{i,m+1} = t_{i\cdot}$ The density Equation (8) is a member of a curved multi-parameter exponential family with sufficient statistics

$$B_i, Z_i, N_{ij},$$

where $i = 1, \ldots, m$, $j = 1, \ldots, m+1$, $i \neq j$. For the complete data, maximum likelihood estimates of the unknown parameters were obtained by Asmussen *et al.* (1996). They can be used in our context and, therefore, the M-step is given by

$$\hat{\pi}_i = \frac{B_i}{n}, \hat{t}_{ij} = \frac{N_{ij}}{Z_i}, i \neq j, \hat{t}_i = \frac{N_{i,m+1}}{Z_i}, \hat{t}_{ii} = -\left(\hat{t}_i + \sum_{j=1, j\neq i}^{m} \hat{t}_{ij} \right).$$

However, the E-step differs considerably since the observed and incomplete data are now bivariate. As noted in Section 3, the E-step for an exponential family consists of computing the conditional expectation of the sufficient statistics, given the complete data and the current parameter estimates.

If the current parameter estimates at step k of the algorithm is $\boldsymbol{\theta}^{(k)}$, the complete sufficient statistics at the $k+1$ E-step consist in the evaluation of the following conditional expectations

$$B_i^{(k+1)} = \sum_{v=1}^{n} \mathbb{E}_{\theta^{(k)}}\left[B_i^{[v]} \middle| \mathbf{x}^{[v]}\right], \tag{9}$$

$$Z_i^{(k+1)} = \sum_{v=1}^{n} \mathbb{E}_{\theta^{(k)}}\left[Z_i^{[v]} \middle| \mathbf{x}^{[v]}\right], \tag{10}$$

$$N_{ij}^{(k+1)} = \sum_{v=1}^{n} \mathbb{E}_{\theta^{(k)}}\left[N_{ij}^{[v]} \middle| \mathbf{x}^{[v]}\right], \tag{11}$$

for $i = 1,\ldots,m$, $j = 1,\ldots,m+1$, $i \neq j$. The most complicated part is the E-step which is derived in the Appendix. All calculations are for the case $x_1 < x_2$; the other case is similar. The final results are given here. For convenience, the same notations are used as in the univariate case in Asmussen *et al.* (1996). To simplify formulas we define the sets $\Gamma_k' = \Gamma_k \backslash \{m+1\}, k = 1, 2$, the matrix $\mathbf{E}_{ij}$ with a one in position (i, j) and zeros elsewhere, and the vector $\mathbf{e}_i$ with a one in position i and zeros elsewhere. Also, let

$$C_1(a, b, i, j, \mathbf{T}) = \int_a^b e^{\mathbf{T}(u-a)} \mathbf{E}_{ij} e^{\mathbf{T}(b-u)} \mathrm{du}.$$

The conditional expectations in Equations (9–11) are given as follows:

$$\mathbb{E}_{\theta^{(k)}}\left[B_i^{[v]} \middle| \mathbf{x}^{[v]}\right] = \frac{\pi_i^{(k)} \mathbf{e}_1^{\mathrm{T}} e^{\mathbf{T}^{(k)} x_1^{[v]}} \mathbf{G}_1 \mathbf{e}^{\mathbf{T}^{(k)} (x_2^{[v]} - x_1^{[v]})} \mathbf{T}^{(k)} \mathbf{g}_2 \mathbf{e}}{f\left(\mathbf{x}^{[v]} \middle| \boldsymbol{\theta}^{(k)}\right)},$$

where $\pi_i^{(k)}$ is the ith component of $\boldsymbol{\pi}^{(k)}$,

$$\mathbb{E}_{\theta^{(k)}}\left[Z_i^{[v]} \middle| \mathbf{x}^{[v]}\right] = \begin{cases} \dfrac{\boldsymbol{\pi}^{(k)\mathrm{T}} C_1\left(0, x_1^{[v]}, i, i, \mathbf{T}^{(k)}\right) \mathbf{G}_1 e^{\mathbf{T}^{(k)}\left(x_2^{[v]} - x_1^{[v]}\right)} \mathbf{T}^{(k)} \mathbf{g}_2 \mathbf{e}}{f\left(\mathbf{x}^{[v]} \middle| \boldsymbol{\theta}^{(k)}\right)}, & i \in \Gamma_1^c \cap \Gamma_2^c, \\ \dfrac{\boldsymbol{\pi}^{(k)\mathrm{T}} e^{\mathbf{T}^{(k)} x_1^{[v]}} \mathbf{G}_1 C_1\left(x_1^{[v]}, x_2^{[v]}, i, i, \mathbf{T}^{(k)}\right) \mathbf{T}^{(k)} \mathbf{g}_2 \mathbf{e}}{f\left(\mathbf{x}^{[v]} \middle| \boldsymbol{\theta}^{(k)}\right)}, & i \in \Gamma_1', \\ 0, \quad \text{otherwise}, \end{cases}$$

$$\mathbb{E}_{\theta^{(k)}}\left[N_{ij}^{[v]} \middle| \mathbf{x}^{[v]}\right] = \begin{cases} t_{ij} \dfrac{\boldsymbol{\pi}^{(k)\mathrm{T}} C_1\left(0, x_1^{[v]}, i, j, \mathbf{T}^{(k)}\right) \mathbf{G}_1 e^{\mathbf{T}^{(k)}\left(x_2^{[v]} - x_1^{[v]}\right)} \mathbf{T}^{(k)} \mathbf{g}_2 \mathbf{e}}{f\left(\mathbf{x}^{[v]} \middle| \boldsymbol{\theta}^{(k)}\right)}, & i, j \in \Gamma_1^c \cap \Gamma_2^c, \\ -t_{ij} \dfrac{\boldsymbol{\pi}^{(k)\mathrm{T}} e^{\mathbf{T}^{(k)} x_1^{[v]}} \mathbf{E}_{ij} e^{\mathbf{T}^{(k)}\left(x_2^{[v]} - x_1^{[v]}\right)} \mathbf{T}^{(k)} \mathbf{g}_2 \mathbf{e}}{f\left(\mathbf{x}^{[v]} \middle| \boldsymbol{\theta}^{(k)}\right)}, & i \in \Gamma_1^c \cap \Gamma_2^c, j \in \Gamma_1', \\ t_{ij} \dfrac{\boldsymbol{\pi}^{(k)\mathrm{T}} e^{\mathbf{T}^{(k)} x_1^{[v]}} \mathbf{G}_1 C_1\left(x_1^{[v]}, x_2^{[v]}, i, j, \mathbf{T}^{(k)}\right) \mathbf{T}^{(k)} \mathbf{g}_2 \mathbf{e}}{f\left(\mathbf{x}^{[v]} \middle| \boldsymbol{\theta}^{(k)}\right)}, & i, j \in \Gamma_1', \\ 0, \quad \text{otherwise}, \end{cases}$$

$$\mathbb{E}_{\theta^{(k)}}\left[N_{i,m+1}^{[v]}\middle|\mathbf{x}^{[v]}\right] = -t_i \frac{\boldsymbol{\pi}^{(k)^T} e^{\mathbf{T}^{(k)}} x_1^{[v]} \mathbf{G}_1 e^{\mathbf{T}^{(k)}(x_2^{[v]}-x_1^{[v]})} \mathbf{e}_i}{f\left(\mathbf{x}^{[v]}\middle|\boldsymbol{\theta}^{(k)}\right)}.$$

The function $C_1(a, b, i, j, \mathbf{T})$ can be evaluated by numerical methods in ordinary or partial differential equations such as the Runge–Kutta method of order four. Note that for $a < b$,

$$C_1(a, b, i, j, \mathbf{T}) = C_1'(0, b-a, i, j, \mathbf{T})$$

where

$$C_1'(0, r, i, j, \mathbf{T}) = \int_0^r e^{\mathbf{T}(r-u)} \mathbf{E}_{ij} e^{\mathbf{T}u} \mathrm{d}u,$$

and

$$\frac{\mathrm{d}C_1'(0, r, i, j, \mathbf{T})}{\mathrm{d}r} = \mathbf{E}_{ij} e^{\mathbf{T}r} + \mathbf{T} C_1'(0, r, i, j, \mathbf{T})$$

with initial conditions $C_1'(0, 0, i, j, \mathbf{T}) = \mathbf{0}$.

We end this section with an interesting property that holds when one fits a BPH distribution by the EM algorithm: at each iteration of the EM algorithm, the mean of the fitted BPH distribution equals the sample mean. This property was given by Asmussen *et al*. (1996) for PH distribution and it also holds in the bivariate case as is now shown. The observations are linear functions of the sufficient statistics,

$$\left(\mathrm{X}_1^{[v]}, \mathrm{X}_2^{[v]}\right) = \left(\sum_{i\in\Gamma\setminus\Gamma_1} Z_i^{[v]}, \sum_{i\in\Gamma\setminus\Gamma_2} Z_i^{[v]}\right).$$

For the component X_1, this implies

$$n\bar{X}_1 = \sum_{v=1}^{n} \sum_{i\in\Gamma\setminus\Gamma_1} Z_i^{[v]}.$$

Taking conditional expectations, given $\mathbf{x}$, on both sides yields

$$\begin{aligned} n\bar{X}_1 &= \sum_{v=1}^{n} \sum_{i\in\Gamma\setminus\Gamma_1} \mathbb{E}_{\theta^{(k)}}\left[Z_i^{[v]}\middle|\mathbf{x}^{[v]}\right] \\ &= \sum_{v=1}^{n} \sum_{i\in\Gamma\setminus\Gamma_1} \mathbb{E}_{\theta^{(k+1)}}\left[Z_i^{[v]}\right] \\ &= \sum_{v=1}^{n} \mathbb{E}_{\theta^{(k+1)}}[X_1^{[v]}] \\ &= n\,\mathbb{E}_{\theta^{(k+1)}}[X_1]. \end{aligned}$$

The same argument applies to the component X_2.

4. Simulated data from a BPH distribution

Consider a BPH distribution with state space $\Gamma = \{1, 2, 3, 4\}$, closed subsets $\Gamma_1 = \{2, 4\}$ and $\Gamma_2 = \{3, 4\}$, $\boldsymbol{\pi} = (1, 0, 0)^{\mathrm{T}}$, and the matrix

$$\mathbf{T} = \begin{pmatrix} -a_1 & pa_1 & qa_1 \\ 0 & -a_2 & 0 \\ 0 & 0 & -a_3 \end{pmatrix}, \tag{12}$$

where $0 < p < 1$, $q = 1 - p$, $a_i > 0$, $i = 1, 2, 3$. This is a special case of the Marshall–Olkin distribution, see Marshall and Olkin (1967). The joint survival function can be written without matrix exponential using the simplified joint density Equation (5)

$$S_\theta(x_1, x_2) = \begin{cases} \frac{p}{a_1 - a_2}\left[a_1 e^{-x_1(a_1-a_2)-x_2 a_2} - a_2 e^{-a_1 x_2}\right] + q e^{-a_1 x_2}, & x_2 \geq x_1 \geq 0, \\ \frac{q}{a_1 - a_3}\left[a_1 e^{-x_2(a_1-a_3)-x_1 a_3} - a_3 e^{-a_1 x_1}\right] + p e^{-a_1 x_1}, & x_1 \geq x_2 \geq 0. \end{cases}$$

The Pearson correlation coefficient,

$$r_\theta = \frac{a_2 a_3 - pq a_1^2}{[a_2^2 + (1 - q^2)a_1^2]^{1/2}[a_3^2 + (1 - p^2)a_1^2]^{1/2}},$$

is obtained from the Laplace transform. It is estimated from the data using the usual sample correlation coefficient $\hat{r}$. The maximum correlation of 1 is obtained by letting a_1 approach 0, whereas the minimum of $-1/3$ is reached by letting a_2 and a_3 approach 0 and choosing $p = 1/2$. Also, if $a_2 = qa_1$ and $a_3 = pa_1$ then, $r = 0$. In fact, the last case corresponds to two independent exponentials. A more appropriate measure of dependence for distributions with nonlinear regression is Spearman correlation

$$\rho_{\boldsymbol{\theta}} = -3 + 12 \int_0^\infty \int_0^\infty S_{\boldsymbol{\theta}}(x_1, x_2) \mathrm{d}F_{\boldsymbol{\theta},1}(x_1) \mathrm{d}F_{\boldsymbol{\theta},2}(x_2).$$

where $F_{\theta,i}$, $i = 1, 2$, are the two marginal distributions of the joint BPH distribution F_θ. In the special case $a_2 = a_3$ it is given by

$$\rho_{\boldsymbol{\theta}} = \frac{7a_1 a_2^2 + 2a_2^3 - 6a_1^3 pq + a_1^2 a_2 (4p - 3)(4p - 1)}{2a_1^3 + 7a_1 a_2 (a_1 + a_2) + 2a_2^3}.$$

The minimum Spearman correlation of $-3/4$ and the maximum of 1 are obtained under the same circumstances as for Pearson correlation. Another appropriate measure is Kendall correlation

$$\tau_{\boldsymbol{\theta}} = -1 + 4 \int_0^{\infty} \int_0^{\infty} S_{\theta}(x_1, x_2) \mathrm{d}F_{\theta}(x_1, x_2),$$

which for this model is

$$\tau_{\boldsymbol{\theta}} = \frac{a_2 a_3 - 2a_1^2 pq + a_1(a_3 q^2 + a_2 p^2)}{(a_1 + a_2)(a_1 + a_3)}.$$

Kendall $\tau_{\boldsymbol{\theta}}$ for this model varies between $-1/2$ and 1. For any bivariate distribution, Spearman $\rho_{\boldsymbol{\theta}}$ and Kendall $\tau_{\boldsymbol{\theta}}$ can be consistently estimated from the data using their empirical versions

$$\hat{\rho} = \frac{12}{n(n+1)(n-1)} \sum_{v=1}^{n} R_v S_v - 3\frac{n+1}{n-1}$$

$$\hat{\tau} = \frac{4}{n(n-1)} P_n - 1,$$

where (R_v, S_v) are the pairs of ranks and P_n is the number of concordant pairs. Here, two pairs $\left(x_1^{[v]}, x_2^{[v]}\right)$ and $\left(x_1^{[\gamma]}, x_2^{[\gamma]}\right)$ are said to be concordant when

$$\left(x_1^{[v]} - x_1^{[\gamma]}\right)\left(x_2^{[v]} - x_2^{[\gamma]}\right) > 0.$$

The EM algorithm was run on 10,000 data sets of size $n = 50$, 100, 200 and 400 generated from a BPH distribution with sub-intensity matrix $\mathbf{T}$ as in Equation (12) with $p = 0.5$, $a_1 = 0.05$, $a_2 = 0.1$, and $a_3 = 0.1$, which gives correlation coefficients of $r_{\boldsymbol{\theta}} = 0.7836$, $\rho_{\boldsymbol{\theta}} = 0.675$, and $\tau_{\boldsymbol{\theta}} = 0.5$. For the model Equation (12), it can be observed that since each of the conditional expectations of the E-step does not depend on $\boldsymbol{\theta}$ then, the EM algorithm converges in one iteration. This remark gives also an explicit expression for the EM estimator

$$\hat{\mathbf{T}} = \begin{bmatrix} -\frac{n}{\sum_{v=1}^{n} \min\left(x_1^{[v]}, x_2^{[v]}\right)} & \frac{\sum_{v=1}^{n} 1\left\{x_1^{[v]} < x_2^{[v]}\right\}}{\sum_{v=1}^{n} \min\left(x_1^{[v]}, x_2^{[v]}\right)} & \frac{\sum_{v=1}^{n} 1\left\{x_1^{[v]} > x_2^{[v]}\right\}}{\sum_{v=1}^{n} \min\left(x_1^{[v]}, x_2^{[v]}\right)} \\ 0 & -\frac{\sum_{v=1}^{n} 1\left\{x_1^{[v]} < x_2^{[v]}\right\}}{\sum_{v=1}^{n} \max\left(0, x_2^{[v]} - x_1^{[v]}\right)} & 0 \\ 0 & 0 & -\frac{\sum_{v=1}^{n} 1\left\{x_1^{[v]} > x_2^{[v]}\right\}}{\sum_{v=1}^{n} \max\left(0, x_1^{v} - x_2^{v}\right)} \end{bmatrix}$$

Table 1. Bias and standard deviation of the EM estimator.

n	Bias	Standard deviation
50	$\begin{pmatrix} 0.0010 & 0.0005 & 0.0005 \\ 0 & 0.0040 & 0 \\ 0 & 0 & 0.0042 \end{pmatrix}$	$\begin{pmatrix} 0.0074 & 0.0053 & 0.0051 \\ 0 & 0.0218 & 0 \\ 0 & 0 & 0.0218 \end{pmatrix}$
100	$\begin{pmatrix} 0.0005 & 0.0002 & 0.0002 \\ 0 & 0.0021 & 0 \\ 0 & 0 & 0.0019 \end{pmatrix}$	$\begin{pmatrix} 0.0051 & 0.0036 & 0.0036 \\ 0 & 0.0149 & 0 \\ 0 & 0 & 0.0147 \end{pmatrix}$
200	$\begin{pmatrix} 0.0002 & 0.0001 & 0.0001 \\ 0 & 0.0011 & 0 \\ 0 & 0 & 0.0011 \end{pmatrix}$	$\begin{pmatrix} 0.0036 & 0.0025 & 0.0025 \\ 0 & 0.0104 & 0 \\ 0 & 0 & 0.0103 \end{pmatrix}$
400	$.001\begin{pmatrix} 0.0991 & 0.0461 & 0.0530 \\ 0 & 0.6161 & 0 \\ 0 & 0 & 0.5165 \end{pmatrix}$	$\begin{pmatrix} 0.0025 & 0.0018 & 0.0018 \\ 0 & 0.0072 & 0 \\ 0 & 0 & 0.0071 \end{pmatrix}$

In this special case, it can be easily seen that the EM estimator coincides with the maximum likelihood estimator (MLE). To this end, note that the probability density function equals

$$f_{\boldsymbol{\theta}}(x_1, x_2) = \begin{cases} pa_1a_2e^{-x_1a_1-a_2(x_2-x_1)}, & x_2 \geq x_1 \geq 0 \\ (1-p)a_1a_3e^{-x_2a_1-a_3(x_1-x_2)}, & x_1 \geq x_2 \geq 0 \end{cases}.$$

The remaining calculations of the MLE are straightforward. Table 1 reports absolute value of biases and standard deviations of the EM estimate computed over the 10,000 replicates. For example, for $n = 50, \hat{\mathbf{T}}_{11} = -n \Big/ \sum_{v=1}^{n} \min\left(x_1^{[v]}, x_2^{[v]}\right)$ has a bias of 0.0010 and a standard deviation of 0.0074, whereas $\hat{\mathbf{T}}_{12}$ has a bias of 0.0005 and a standard deviation of 0.0053. As for any regular maximum likelihood estimator, bias decreases as n^{-1} and becomes negligible compared to standard deviation which decreases as $n^{-1/2}$.

5. Goodness-of-fit test

The statistic used for testing goodness-of-fit is

$$V_n^2 = \sum_{v=1}^{n} \left[S_{\hat{\boldsymbol{\theta}}_n}\left(x_1^{[v]}, x_2^{[v]}\right) - \hat{S}_n\left(x_1^{[v]}, x_2^{[v]}\right) \right]^2, \tag{13}$$

where $\hat{S}_n$ is the empirical survival function, *i.e.*

$$\hat{S}_n(x_1, x_2) = \frac{1}{n}\sum_{v=1}^{n} 1\left\{X_1^{[v]} > x_1, X_2^{[v]} > x_2\right\},$$

and $S_{\hat{\boldsymbol{\theta}}_n}$ is the parametric survival function Equation (2) with $\boldsymbol{\theta}$ estimated by the EM algorithm. Large values of V_n^2 are evidence against the parametric model. For regular parametric models, Stute *et al.* (1993) established that the parametric bootstrap of

$$W_n^2 = n\int \left[F_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}) - \hat{F}_n(\mathbf{x})\right]^2 dF_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x})$$

is consistent for testing goodness-of-fit. An argument similar to the Lemma in Section 2 of Kiefer (1959) can be used to show that the following version

$$\begin{aligned} W_n^2 &= n\int \left[F_{\hat{\boldsymbol{\theta}}_n}(\mathrm{x}) - \hat{F}_n(\mathrm{x})\right]^2 d\hat{F}_n(\mathrm{x}) \\ &= \sum_{v=1}^{n}\left[F_{\hat{\boldsymbol{\theta}}_n}\left(\mathrm{x}^{[v]}\right) - \hat{F}_n\left(\mathrm{x}^{[v]}\right)\right]^2 \end{aligned}$$

can also be bootstrapped consistently. The proposed V_n^2 is obtained by replacing the distribution function by the survival function which has a simpler expression for BPH distributions.

The goodness-of-fit bootstrap test of significance level α is performed as follows. Given a sample of size n, $(x_1^{[v]}, x_2^{[v]}), v = 1, \ldots, n$, estimate $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}_n$ using the EM algorithm for BPH distributions, and calculate the goodness-of-fit statistic V_n^2 in Equation (13). Then, repeat a large number, say B, of times the following 3 steps.

(1) Generate a bootstrap sample of size n from the BPH distribution with parameter $\hat{\boldsymbol{\theta}}_n$ denoted $(\tilde{x}_1^{[v]}, \tilde{x}_2^{[v]}), v = 1, \ldots, n$.
(2) Find the EM estimate $\tilde{\boldsymbol{\theta}}_n$ from the bootstrap sample.
(3) Compute the goodness-of-fit statistic

$$\tilde{V}_n^2 = \sum_{v=1}^{n}\left[S_{\hat{\boldsymbol{\theta}}_n}\left(\tilde{x}_1^{[v]}, \tilde{x}_2^{[v]}\right) - \tilde{S}_n\left(\tilde{x}_1^{[v]}, \tilde{x}_2^{[v]}\right)\right]^2,$$

where $\tilde{S}_n$ is the empirical survival function computed from the bootstrap sample.

After repeating the previous loop B times, this Monte Carlo simulation produces B (ordered) values:

$$\tilde{V}_{n,(1)}^2 \leq \tilde{V}_{n,(2)}^2 \leq \cdots \leq \tilde{V}_{n,(B)}^2.$$

The bootstrap test rejects the model when V_n^2 exceeds the $\lceil(1-\alpha)B\rceil$ order statistic, *i.e.* when $V_n^2 > \tilde{V}_{n,(\lceil(1-\alpha)B\rceil)}^2$.

A simulation was conducted to verify the significance level of the bootstrap test. It consisted in generating 2,000 samples of size $n = 200$ from the BPH distribution Equation (12), each with a different sub-intensity matrix $\mathbf{T}$ generated at random. The parameters were independently generated from the following distributions: a_1, a_2, a_3 uniform on the interval (0,10) and p uniform on (0,1). Each bootstrap test was done at level $\alpha = 0.05$ with $B = 1{,}000$ bootstrap samples. The proportion of significant bootstrap tests obtained, out of 2,000, was 0.054, which is close to the intended 0.05 level.

6. Fitting ALAE data with a BPH model

This section consists of fitting a BPH distribution to insurance company indemnity claims. The data set contains 1,500 bi-dimensional observations. The variables are LOSS or indemnity payment and ALAE, allocated loss adjustment expenses, which covers expenses attributed to the settlement of individual claims such as claim investigation expenses. See Klugman *et al*. (2004) for more information on the data. Frees and Valdez (1998) and Klugman and Parsa (1999) fitted this data set using copulas. The variable LOSS treated in these two papers is the loss incurred to the insured so that this variable is censored when the claim exceeds the policy limit. Here, LOSS is always the indemnity payment so that there is no censoring. It can be observed from Figure 1 that there is a moderate positive sample Pearson correlation of $\hat{r} = 0.4022$. The descriptive statistics are given in Table 2.

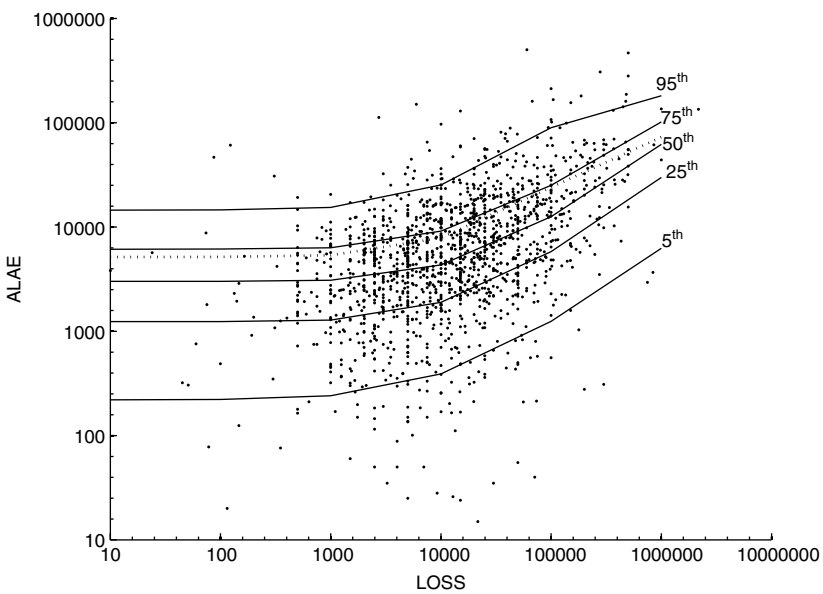

Figure 1. ALAE versus LOSS with curves for conditional quantiles and mean. The dotted curve is the conditional mean of ALAE given LOSS.

Table 2. Descriptive statistics of ALAE data.

	LOSS	ALAE
Mean	41,208	12,588
Median	12,000	5,471
Standard deviation	102,748	28,146
Minimum	10	15
Maximum	2,173,595	501,863
0.25 quantile	4,000	2,333
0.75 quantile	35,000	12,577
Pearson $\hat{r}$	0.4022	
Spearman $\hat{\rho}$	0.4519	
Kendall $\hat{\tau}$	0.3154	

The variables were rescaled as LOSS/100,000 and ALAE/10,000 so that they are about the same order of magnitude. This rescaling also makes the elements of the sub-intensity matrix **T** not too close to zero, thus improving the numerical stability. A few choices were tried for Γ_1 and Γ_2 and it was found that the choice $\Gamma = \{1,\ldots 12\}$, $\Gamma_1 = \{5, 6, 7, 12\}$ and $\Gamma_2 = \{8, 9, 10, 11, 12\}$ yields a good fit. The EM algorithm was iterated 300 times starting with random values of $\boldsymbol{\pi}$ and **T**. The fitted BPH distribution could capture the essential characteristics of the joint distribution, such as the dependence structure.

As shown in Table 3, the fitted BPH survival function is close to the empirical survival function over the whole domain. The fitted BPH distribution has standard deviations for LOSS and ALAE of 102,750 and 28,146, respectively. Pearson, Spearman, and Kendall correlation coefficients are $r_{\hat{\boldsymbol{\theta}}_n} = 0.3932$, $\rho_{\hat{\boldsymbol{\theta}}_n} = 0.4512$ (0.0116), and $\tau_{\hat{\boldsymbol{\theta}}_n} = 0.2968$ (0.0042), respectively. These are very close to the sample statistics in Table 2. Since there is no explicit expression for the latter two coefficients, they can be computed numerically, with any desired degree of accuracy, through a simulation. The evaluation of Kendall $\tau_{\hat{\boldsymbol{\theta}}_n}$, for example, was done by jointly simulating 50,000 data points from the fitted (joint) BPH distribution, $F_{\hat{\boldsymbol{\theta}}_n}$, and averaging, over these 50,000 values, the survival function, $S_{\hat{\boldsymbol{\theta}}_n}$. As a

Table 3. Comparison of survival functions.

(LOSS,ALAE)	$\hat{S}_n$	$S_{\hat{\boldsymbol{\theta}}_n}$	G-H	Frank
(1e3,0)	0.9300	0.9262	0.9257	0.9286
(8e3,0)	0.5927	0.5981	0.6029	0.6137
(1e4,0)	0.5200	0.5425	0.5469	0.5582
(1e6,0)	0.0007	0.0007	0.0082	0.0088
(1e7,0)	0.0000	0.0000	0.0000	0.0000
(0,1e3)	0.8793	0.8819	0.8659	0.8742
(0,1e4)	0.3120	0.3150	0.3237	0.3380
(0,1e5)	0.0133	0.0154	0.0121	0.0112
(1e2,1e1)	0.9932	0.9910	0.9907	0.9910
(1e2,1e2)	0.9827	0.9807	0.9780	0.9790
(1e3,1e2)	0.9203	0.9162	0.9148	0.9177
(1e4,1e5)	0.0120	0.0142	0.0115	0.0097
(1e3,1e3)	0.8253	0.8233	0.8157	0.8255
(1e4,1e3)	0.4847	0.5047	0.5078	0.5267
(1.5e4,5e3)	0.3013	0.3088	0.3103	0.3356
(5e3,1.5e4)	0.1940	0.1933	0.1953	0.2070
(4e4,1.2e4)	0.1280	0.1263	0.1255	0.1221
(1e6,1e5)	0.0007	0.0002	0.0039	0.0003
V_n^2		0.1280	0.1502	0.6429

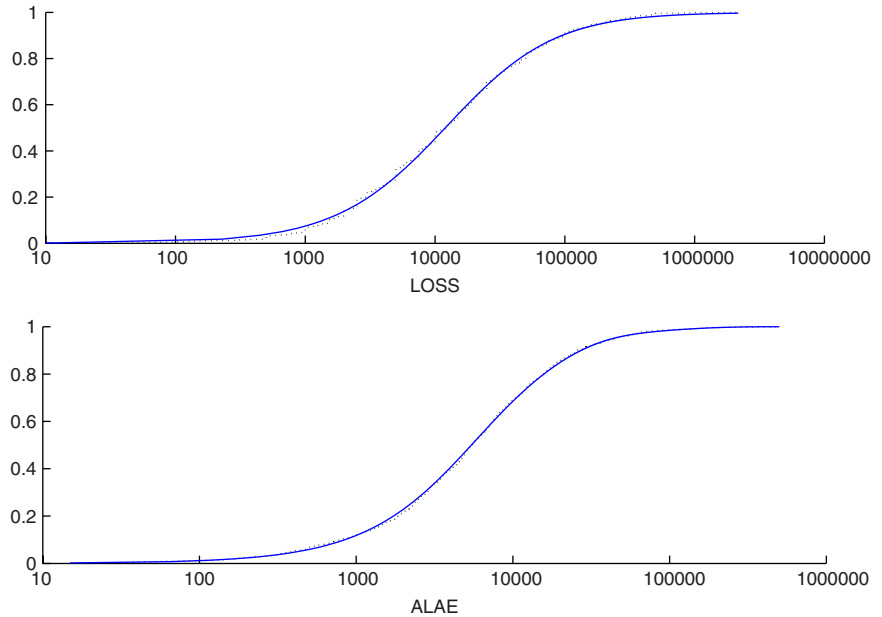

Figure 2. Marginal fitted distribution functions (smooth) and empirical distribution functions.

measure of accuracy, the standard error of the mean accompanies in parentheses this average. Spearman correlation can be evaluated similarly, with the exception that the two components are independently generated from the marginal PH distributions of the fitted BPH distribution. Also, the means of the fitted BPH distribution always equal the sample means as shown in Section 3.2. The marginals of the fitted BPH distribution are plotted in Figure 2. The columns G-H and Frank of Table 3 give the two fits obtained in Frees and Valdez (1998) using Pareto distributed marginals in both cases with either the Gumbel-Hougaard copula or the Frank copula. The estimates of the parameters of these two fits were computed by Frees and Valdez (1998) and were simply used here to compute the survival functions from the distribution functions. The last row is the measure of global fit used for the goodness-of-fit test in Section 5 and computed here over all the $n = 1{,}500$ data points. The minimum value is obtained for the BPH distribution. The Gumbel-

Table 4. Conditional quantiles and mean of ALAE given LOSS.

	Quantiles					
LOSS	0.05	0.25	0.50	0.75	0.95	Mean
1e1	220	1,240	3,010	6,120	14,590	5,146
1e2	222	1,240	3,012	6,140	14,590	5,164
1e3	241	1,278	3,088	6,316	15,400	5,356
5e3	318	1,500	3,529	7,324	19,610	6,342
1e4	390	1,895	6,316	9,128	25,320	7,852
2.5e4	643	3,321	7,364	15,070	36,280	12,243
5e4	954	3,869	9,592	18,914	45,990	15,981
1e5	1,246	5,814	12,567	25,115	89,590	24,034
1e6	6,250	29,840	61,930	101,660	181,500	71,721

Hougaard and Frank models with the censored variable LOSS give fits quite similar to the one obtained with the BPH model without censoring. Strictly speaking however, such comparisons are difficult to interpret since the variable LOSS has different meanings.

6.1. Conditional quantiles and mean

The conditional survival function of X_2, given $X_1 = x_1$, can be obtained from Equation (2),

$$\mathbb{P}(X_2 > x_2 | x_1) = \begin{cases} \boldsymbol{\pi}^{\mathrm{T}} e^{\mathbf{T}x_1} \mathbf{G}_1 e^{\mathbf{T}(x_2 - x_1)} \mathbf{g}_2 \mathbf{e} / [\boldsymbol{\pi}^{\mathrm{T}} e^{\mathbf{T}x_1} \mathbf{T}\mathbf{g}_1 \mathbf{e}], & x_1 < x_2, \\ \boldsymbol{\pi}^{\mathrm{T}} e^{\mathbf{T}x_2} \mathbf{g}_2 e^{\mathbf{T}(x_1 - x_2)} \mathbf{T}\mathbf{g}_1 \mathbf{e} / [\boldsymbol{\pi}^{\mathrm{T}} e^{\mathbf{T}x_1} \mathbf{T}\mathbf{g}_1 \mathbf{e}], & x_1 > x_2. \end{cases}$$

For a given value of LOSS, quantiles of ALAE were calculated based on the conditional survival function Equation (14) and presented in Table 4 and Figure 1. The conditional mean of X_2 given $X_1 = x_1$ is somewhat tedious. For this purpose, by using either Equation (14) directly or the conditional probability density function derived from Equation (14), one can write a system of differential equations of the first order which can be numerically solved by, *e.g.*, Runge-Kutta methods. The probability density function of X_2 given $X_1 = x_1$ was used here. After some straightforward calculations, the conditional mean is given by

$$\mathbb{E}(X_2 | x_1) = \frac{\boldsymbol{\pi}^T e^{\mathbf{T}x_1} \mathbf{G}_1 (-\mathbf{T}^{-1} + x_1 \mathbf{I}) \mathbf{g}_2 \mathbf{e} + \boldsymbol{\pi}^T C_2(x_1, \mathbf{G}_2, \mathbf{T}) \mathbf{T}\mathbf{g}_1 \mathbf{e}}{-\boldsymbol{\pi}^{\mathrm{T}} e^{\mathbf{T}x_1} \mathbf{T}\mathbf{g}_1 \mathbf{e}},$$

where $C_2(x, \mathbf{G}, \mathbf{T})$ is a function satisfying the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}x} C_2(x, \mathbf{G}, \mathbf{T}) = C_2(x, \mathbf{G}, \mathbf{T}) \mathbf{T} + x e^{\mathbf{T}x} \mathbf{G},$$

with initial conditions $C_2(0, \mathbf{G}, \mathbf{T}) = \mathbf{0}$. The conditional mean is given in the last column of Table 4. It is observed in Figure 1 that the mean is always greater than the median which reflects the right-skewness of ALAE.

References

Åhlström, L., Olsson, M., & Nerman, O. (1999). A parametric estimation procedure for relapse time distributions. *Lifetime Data Analysis*, **2**, 113–132.

Assaf, D., & Levikson, B. (1982). Closure of phase type distributions under operations arising in reliability theory. *Annals of Probability*, **10**, 265–269.

Assaf, D., Langberg, N. A., Savits, T. H., & Shaked, M. (1984). Multivariate phase-type distributions. *Operation Research*, **32**, 688–702.

Asmussen, S., Nerman, O., & Olsson, M. (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, **23**, 419–441.

Asmussen, S. (2000). *Ruin probabilities*. Advanced Series on Statistical Science & Applied Probability, 2. River Edge, NJ: World Scientific Publishing Co., Inc.

Breuer, L. (2002). An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. *Annals of Operations Research*, **112**, 123–138.
Cai, J., & Li, H. (2005a). Conditional tail expectations for multivariate phase-type distributions. *Journal of Applied Probability*, **42**, 810–825.
Cai, J., & Li, H. (2005b). Multivariate risk model of phase type. *Insurance: Mathematics and Economics*, **36**, 137–152.
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society Series B*, **39**, 1–38.
Drekic, S., Dickson, D. C. M., Stanford, D. A., & Willmot, G. E. (2004). On the distribution of the deficit at ruin when claims are phase-type. *Scandinavian Actuarial Journal*, **2**, 105–120.
Frees, E. W., & Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, **2**, 1–25.
Kiefer, J. (1959). *K*-sample analogues of the Kolmogorov-Smirnov and Cramr-V. Mises tests. *The Annals of Mathematical Statistics*, **30**, 420–447.
Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2004). *Loss models: from data to decision* (2nd ed). Hoboken, NJ: John Wiley & Sons.
Klugman, S. A., & Parsa, R. (1999). Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics*, **24**, 139–148.
Kulkarni, V. G. (1989). A new class of multivariate phase type distributions. *Operation Research*, **37**, 151–158.
Li, H. (2003). Association of multivariate phase-type distributions, with applications to shock models. *Statistics & Probability Letters*, **64**, 381–392.
Li, S., & Garrido, J. (2004). On ruin for the Erlang(*n*) risk process. *Insurance: Mathematics and Economics*, **34**, 391–408.
Marshall, A. W., & Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, **62**, 30–44.
Marshall, A. W., & Shaked, M. (1986). Multivariate new better than used distributions. *Mathematics of Operations Research*, **11**, 110–116.
Neuts, M. F. (1981). *Matrix-geometric solutions in stochastic models. An algorithmic approach*. Baltimore, MD: Johns Hopkins University Press.
Neuts, M. F. (1994). *Matrix-geometric solutions in stochastic models. An algorithmic approach*. New York: Dover Publications Inc.
O'Cinneide, C. A. (1989). On nonuniqueness of representations of phase-type distributions. *Communications in Statistics. Stochastic Models*, **5**, 247–259.
Roberts, W. J. J., & Ephraim, Y. (2008). An EM algorithm for ion-channel current estimation. *IEEE Transactions on Signal Processing*, **56**, 26–33.
Roberts, W. J. J., Ephraim, Y., & Dieguez, E. (2006). On Rydén's EM algorithm for estimating MMPPs. *IEEE Signal Processing Letters*, **13**, 373–376.
Ryden, T. (1996). An EM algorithm for estimation in Markov-modulated Poisson processes. *Computational Statistics & Data Analysis*, **21**, 431–447.
Stute, W., Gonzáles, M. W., & Wenceslao, P. Q. M. (1993). Bootstrap based goodness-of-fit tests. *Metrika*, **40**, 243–256.
Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics*, **11**, 95–103.

Appendix

We assume a *BPH* distribution for which $\mathbb{P}(X_1 > 0, X_2 > 0)$ As a result, the initial state can be only in $\Gamma_1^c \cap \Gamma_2^c$. All calculations are for the case $x_1 < x_2$. The expressions for the case $x_1 > x_2$ are given without further explanations.

1. Calculations of $\mathbb{E}[B_i|\mathrm{x}]$:

For $x_1 < x_2,\ i \in \Gamma_1^c \cap \Gamma_2^c$,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}[B_i|\mathrm{x}] &= \mathbb{E}_{\boldsymbol{\theta}}[1\{i_0 = i\}|\mathbf{x}] \\ &= \mathbb{P}_{\boldsymbol{\theta}}(i_0 = i|\mathbf{x}) \\ &= \frac{\mathbb{P}_{\boldsymbol{\theta}}(i_0 = i)\,\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{X} \in \mathrm{d}\mathbf{x}|i_0 = i)}{\mathbb{P}_{\theta}(\mathbf{X} \in d\mathbf{x})} \\ &= \frac{\pi_i \mathbf{e}_i^{\mathrm{T}} e^{\mathbf{T}x_1} \mathbf{G}_1 e^{\mathbf{T}(x_2 - x_1)} \mathbf{T}\mathbf{g}_2 \mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}.\end{aligned}$$

For $x_1 > x_2, i \in \Gamma_1^c \cap \Gamma_2^c$,

$$\mathbb{E}_{\theta}[B_i|\mathrm{x}] = \frac{\pi_i \mathbf{e}_i^{\mathrm{T}} e^{\mathbf{T}x_2} \mathbf{G}_2 e^{\mathbf{T}(x_1 - x_2)} \mathbf{T}\mathbf{g}_1 \mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}.$$

2. Calculations of $\mathbb{E}[Z_i|\mathrm{x}]$:
For $x_1 < x_2, i \in \Gamma_1^c \cap \Gamma_2^c$,

$$\begin{aligned}\mathbb{E}_{\theta}[Z_i|\mathbf{x}] &= \mathbb{E}_{\boldsymbol{\theta}}\left[\int_0^{\infty} 1\{J_u = i\}\mathrm{d}u|\mathbf{x}\right] \\ &= \int_0^{\infty} \mathbb{P}_{\theta}(J_u = i|\mathbf{x})\mathrm{d}u \\ &= \int_0^{\infty} \frac{\mathbb{P}_{\theta}(J_u = i)\mathbb{P}_{\theta}(\mathbf{X} \in \mathrm{d}\mathbf{x}|J_u = i)}{\mathbb{P}_{\theta}(\mathbf{X} \in \mathrm{d}\mathbf{x})}\mathrm{d}u \\ &= \frac{\boldsymbol{\pi}^{\mathrm{T}} C_1(0, x_1, i, i, \mathbf{T})\, \mathbf{G}_1 \mathbf{e}^{\mathbf{T}(x_2 - x_1)}\, \mathbf{T}\mathbf{g}_2 \mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}.\end{aligned}$$

For $x_1 > x_2, i \in \Gamma_1^c \cap \Gamma_2^c$,

$$\mathbb{E}_{\theta}[Z_i|\mathbf{x}] = \frac{\boldsymbol{\pi}^{\mathrm{T}} C_1(0, x_2, i, i, \mathbf{T}) \mathbf{G}_2 \mathbf{e}^{\mathbf{T}(x_1 - x_2)} \mathbf{T}\mathbf{g}_1 \mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}.$$

For $x_1 < x_2, i \in \Gamma_1'$,

$$\begin{aligned}\mathbb{E}_{\theta}[Z_i|\mathbf{x}] &= \int_0^{\infty} \mathbb{P}_{\theta}(J_u = i|\mathbf{x})\mathrm{d}u \\ &= \int_0^{\infty} \frac{\mathbb{P}_{\theta}(X_1 \in \mathrm{d}x_1)\mathbb{P}_{\theta}(J_u = i|X_1 \in \mathrm{d}x_1)\mathbb{P}_{\theta}(X_2 \in \mathrm{d}x_2|J_u = i)}{\mathbb{P}_{\theta}(\mathbf{X} \in \mathrm{d}\mathbf{x})} \\ &= \frac{\boldsymbol{\pi}^{\mathrm{T}} e^{\mathbf{T}x_1} \mathbf{G}_1 C_1(x_1, x_2, i, i, \mathbf{T}) \mathbf{T}\mathbf{g}_2 \mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}.\end{aligned}$$

For $x_1 > x_2, i \in \Gamma_2'$,

$$\mathbb{E}_\theta[Z_i|\mathrm{x}] = \frac{\boldsymbol{\pi}^{\mathrm{T}} e^{\mathbf{T}x_2} \mathbf{G}_2 C_1(x_2, x_1, i, i, \mathbf{T}) \mathbf{T}\mathbf{g}_1 \mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}$$

3. Calculations of $\mathbb{E}\left[N_{ij}|x\right]$:
For every small $\varepsilon > 0$, N_{ij} can be approximated by

$$N_{ij}^{\varepsilon} = \sum_{k=0}^{\infty} 1\left\{J_{k\varepsilon} = i, J_{(k+1)\varepsilon} = j\right\}.$$

For each case we have calculated $\mathbb{E}_\theta\left[N_{ij}^{\varepsilon}|\mathbf{x}\right]$. The exact value can be obtained by letting $\varepsilon \downarrow 0$.
For $x_1 < x_2, i,j \in \Gamma_1^c \cap \Gamma_2^c$,

$$\begin{aligned}
\mathbb{E}_\theta\left[N_{ij}^{\varepsilon}|\mathbf{x}\right] &= \mathbb{E}_\theta\left[\sum_{k=0}^{\infty} 1\left\{J_{k\varepsilon} = i, J_{(k+1)\varepsilon} = j\right\}|\mathbf{x}\right] \\
&= \sum_{k=0}^{\lfloor x_1/\varepsilon \rfloor - 1} \frac{\mathbb{P}_\theta(J_{k\varepsilon} = i, J_{(k+1)\varepsilon} = j, \mathbf{X} \in d\mathbf{x})}{\mathbb{P}_\theta(\mathbf{X} \in d\mathbf{x})} \\
&= \sum_{k=0}^{\lfloor x_1/\varepsilon \rfloor - 1} \frac{\mathbb{P}_\theta(J_{k\varepsilon} = i)\, \mathbb{P}_\theta(J_{(k+1)\varepsilon} = j | J_{k\varepsilon} = i)\, \mathbb{P}_\theta(\mathbf{X} \in \mathrm{d}\mathbf{x} \Big| J_{(k+1)\varepsilon} = j)}{\mathbb{P}_\theta(\mathbf{X} \in \mathrm{d}\mathbf{x})}.
\end{aligned}$$

Since $e^{\mathbf{T}u}$ is a continuous function and

$$\frac{e^{\mathbf{T}\varepsilon} - \mathbf{I}}{\varepsilon} \to \mathbf{T}, \quad as\ \varepsilon \to 0,$$

then,

$$\mathbb{P}_\theta(J_{(K+1)\varepsilon} = j \big| J_{k\varepsilon} = i) \to t_{ij},$$

and

$$\mathbb{E}_\theta\left[N_{ij}^{\varepsilon}|\mathbf{x}\right] \to t_{ij} \frac{\boldsymbol{\pi}^{\mathrm{T}} C_1(0, x_1, i, j, \mathbf{T}) \mathbf{G}_1 e^{\mathbf{T}(x_2 - x_1)} \mathbf{T}\mathbf{g}_2 \mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}.$$

For $x_1 > x_2, i,j \in \Gamma_1^c \cap \Gamma_2^c$,

$$\mathbb{E}_\theta\left[N_{ij}|\mathbf{x}\right] = t_{ij} \frac{\boldsymbol{\pi}^{\mathrm{T}} C_1(0, x_2, i, j, \mathbf{T}) \mathbf{G}_2 e^{\mathbf{T}(x_1 - x_2)} \mathbf{T}\mathbf{g}_1 \mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}.$$

For $x_1 < x_2, i \in \Gamma_1^c \cap \Gamma_2^c, j \in \Gamma_1'$,

$$\begin{aligned}\mathbb{E}_\theta\left[N_{ij}^{\varepsilon}|\mathbf{x}\right] &= \mathbb{P}_\theta(J_{x_1-\varepsilon} = i, J_{x_1} = j|\mathbf{x}) \\ &= \frac{\mathbb{P}_\theta(J_{x_1-\varepsilon} = i)\mathbb{P}_\theta(J_{x_1} = j|J_{x_1-\varepsilon} = i)\mathbb{P}_\theta(X_2 \in dx_2|J_{x_1} = j)}{\mathbb{P}_\theta(\mathbf{X} \in \mathrm{d}\mathbf{x})} \\ &= \frac{\boldsymbol{\pi}^T e^{\mathbf{T}(x_1-\varepsilon)}\mathbf{E}_{ii}e^{\mathbf{T}\varepsilon}\mathbf{e}_j\mathbb{P}_\theta(X_2 \in \mathrm{d}x_2\Big|J_{x_1} = j)}{\mathbb{P}_\theta(\mathbf{X} \in \mathrm{d}\mathbf{x})} \rightarrow \\ &- t_{ij}\frac{\boldsymbol{\pi}^T e^{\mathbf{T}x_1}\mathbf{E}_{ij}e^{\mathbf{T}(x_2-x_1)}\mathbf{T}\mathbf{g}_2\mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}.\end{aligned}$$

For $x_1 > x_2, i \in \Gamma_1^c \cap \Gamma_2^c, j \in \Gamma_2'$,

$$\mathbb{E}_\theta\left[N_{ij}|\mathrm{x}\right] = -t_{ij}\frac{\boldsymbol{\pi}^T e^{\mathbf{T}x_2}\mathbf{E}_{ij}e^{\mathbf{T}(x_1-x_2)}\mathbf{T}\mathbf{g}_1\mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}.$$

For $x_1 < x_2, i,j \in \Gamma_1'$,

$$\begin{aligned}\mathbb{E}_\theta\left[N_{ij}^{\varepsilon}|\mathbf{x}\right] &= \sum_{k=[x_1/\varepsilon]}^{[x_2/\varepsilon]-1}\mathbb{P}_\theta(J_{k\varepsilon} = i, J_{(k+1)\varepsilon} = j|\mathbf{x}) \\ &= \sum_{k=[x_1/\varepsilon]}^{[x_2/\varepsilon]-1}\frac{\mathbb{P}_\theta(X_1 \in \mathrm{d}x_1, J_{k\varepsilon} = i, J_{(k+1)\varepsilon} = j, X_2 \in \mathrm{d}x_2)}{\mathbb{P}_\theta(\mathbf{X} \in \mathrm{d}\mathbf{x})} \\ &= \sum_{k=[x_1/\varepsilon]}^{[x_2/\varepsilon]-1}\mathbb{P}_\theta(X_1 \in \mathrm{d}x_1)\mathbb{P}_\theta(J_{k\varepsilon} = i, J_{(k+1)\varepsilon} = j|X_1 \in \mathrm{d}x_1) \\ &\quad .\frac{\mathbb{P}_\theta(X_2 \in \mathrm{d}x_2|J_{(k+1)\varepsilon} = j)}{\mathbb{P}_\theta(\mathbf{X} \in \mathrm{d}\mathbf{x})} \\ &\rightarrow t_{ij}\frac{\boldsymbol{\pi}^{\mathrm{T}} e^{\mathbf{T}x_1}\mathbf{G}_1 C_1(x_1, x_2, i, j, \mathbf{T})\mathbf{T}\mathbf{g}_2\mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}.\end{aligned}$$

For $x_1 > x_2, i,j \in \Gamma_2'$,

$$\mathbb{E}_\theta\left[N_{ij}|\mathbf{x}\right] = t_{ij}\frac{\boldsymbol{\pi}^{\mathrm{T}} e^{\mathbf{T}x_2}\mathbf{G}_2 C_1(x_2, x_1, i, j, \mathbf{T})\mathbf{T}\mathbf{g}_1\mathbf{e}}{f(\mathbf{x}|\boldsymbol{\theta})}.$$

For $x_1 < x_2, i \in \Gamma_1', j = m+1$,

$$\begin{aligned}\mathbb{E}_\theta\left[N_{ij}^{\varepsilon}|\mathbf{x}\right] &= \frac{\mathbb{P}_{\boldsymbol{\theta}}(X_1 \in \mathrm{d}x_1)\mathbb{P}_{\boldsymbol{\theta}}(J_{x_2-\varepsilon} = i|X_1 \in \mathrm{d}x_1)\mathbb{P}_{\boldsymbol{\theta}}(J_{x_2} = j|J_{x_2-\varepsilon} = i)}{\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{X} \in \mathrm{d}\mathbf{x})} \\ &\rightarrow -t_i\frac{\boldsymbol{\pi}^T e^{\mathbf{T}x_1}\mathbf{G}_1 e^{\mathbf{T}(x_2-x_1)}\mathbf{e}_i}{f(\mathbf{x}|\boldsymbol{\theta})}\end{aligned}$$

For $x_1 > x_2, i \in \Gamma_2', j = m+1$,

$$\mathbb{E}_\theta\left[N_{ij}|\mathrm{x}\right] = -t_i\frac{\boldsymbol{\pi}^T e^{\mathbf{T}x_2}\mathbf{G}_2 e^{\mathbf{T}(x_1-x_2)}\mathbf{e}_i}{f(\mathbf{x}|\boldsymbol{\theta})}.$$

The *BPH* distribution with a mass on $X_1 = X_2$ occurs rarely in application but is still worthy of mention. Based on the interpretation of the sub-intensity matrix $\mathbf{T}$, this case happens when there is a positive probability of moving from some state in $\Gamma_1^c \cap \Gamma_2^c$ directly to the absorbing state $m+1$. Indeed, all calculations will be on the set $\Gamma_1^c \cap \Gamma_2^c$ because the only state visited on $\Gamma_1 \cap \Gamma_2$ is the absorbing state $m+1$. This case is very similar to the univariate case which is treated in details in Asmussen *et al.* (1996). The expectations of the E-step are

$$\mathbb{E}_\theta[\mathrm{B}_i|\mathrm{x}] = \frac{\pi_i \mathbf{e}_i e^{\mathbf{T}x} \mathbf{g}_1 \mathbf{g}_2 \mathbf{T}\mathbf{e}}{\boldsymbol{\pi}^{\mathbf{T}} e^{\mathbf{T}x} \mathbf{g}_1 \mathbf{g}_2 \mathbf{T}\mathbf{e}}, i \in \Gamma_1^c \cap \Gamma_2^c,$$

$$\mathbb{E}_\theta[Z_i|\mathrm{x}] = \frac{\boldsymbol{\pi}^T C_1(0, x, i, i, \mathbf{T}) \mathbf{g}_1 \mathbf{g}_2 \mathbf{T}\mathbf{e}}{\boldsymbol{\pi}^T e^{\mathbf{T}x} \mathbf{g}_1 \mathbf{g}_2 \mathbf{T}\mathbf{e}} i \in \Gamma_1^c \cap \Gamma_2^c,$$

$$\mathbb{E}_\theta\left[N_{ij}|\mathrm{x}\right] = t_{ij} \frac{\boldsymbol{\pi}^T C_1(0, x, i, j, \mathbf{T}) \mathbf{g}_1 \mathbf{g}_2 \mathbf{T}\mathbf{e}}{\boldsymbol{\pi}^T e^{\mathbf{T}x} \mathbf{g}_1 \mathbf{g}_2 \mathbf{T}\mathbf{e}}, i, j \in \Gamma_1^c \cap \Gamma_2^c,$$

$$\mathbb{E}_\theta\left[N_{ij}|\mathrm{x}\right] = -t_i \frac{\boldsymbol{\pi}^T e^{\mathbf{T}x} \mathbf{g}_1 \mathbf{g}_2 \mathbf{T}\mathbf{e}_i}{\boldsymbol{\pi}^T e^{\mathbf{T}x} \mathbf{g}_1 \mathbf{g}_2 \mathbf{T}\mathbf{e}}, i \in \Gamma_1^c \cap \Gamma_2^c, j = m+1.$$