Astin Bulletin (2026), **56**, pp. 101–126
doi:10.1017/asb.2025.10062

International Actuarial Association
Association Actuarielle Internationale

RESEARCH ARTICLE

Modeling discrete common-shock risks through matrix distributions

Martin Bladt[1], Eric C. K. Cheung[2], Oscar Peralta[3] and Jae-Kyung Woo[2]

[1]Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark
[2]School of Risk and Actuarial Studies, UNSW Business School, University of New South Wales, Sydney, Australia
[3]Department of Actuarial and Insurance Sciences, Autonomous Technological Institute of México, México City, México
Corresponding author: Oscar Peralta; Email: oscar.peralta@itam.mx

Received: 28 February 2025; **Revised:** 22 June 2025; **Accepted:** 6 August 2025; **First published online:** 16 September 2025

Keywords: Phase-type distribution; discrete bivariate distribution; common shocks; expectation-maximization algorithm

Abstract

We introduce a novel class of bivariate common-shock discrete phase-type (CDPH) distributions to describe dependencies in loss modeling, with an emphasis on those induced by common shocks. By constructing two jointly evolving terminating Markov chains that share a common evolution up to a random time corresponding to the common shock component, and then proceed independently, we capture the essential features of risk events influenced by shared and individual-specific factors. We derive explicit expressions for the joint distribution of the termination times and prove various class and distributional properties, facilitating tractable analysis of the risks. Extending this framework, we model random sums where aggregate claims are sums of continuous phase-type random variables with counts determined by these termination times and show that their joint distribution belongs to the multivariate phase-type or matrix-exponential class. We develop estimation procedures for the CDPH distributions using the expectation-maximization algorithm and demonstrate the applicability of our models through simulation studies and an application to bivariate insurance claim frequency data. In particular, the distribution of the latent common shock component present in correlated count data can be estimated as well.

1 Introduction

Ever since their conception in Jensen (1954), phase-type distributions, which represent the time until absorption in a finite state-space Markov chain, have been extensively used due to their flexibility in approximating a wide range of distributions and their analytical tractability (see Bladt and Nielsen, 2017 for a review). Such a class of distributions has also gained popularity in actuarial science in the last two decades. In insurance ruin problems, phase-type distributions can be used to model the claim amounts (e.g., Drekic *et al*., 2004; Frostig *et al*., 2012) and/or the inter-arrival times (an assumption implicitly embedded in the class of risk processes with Markovian claim arrivals; see, e.g., Cheung and Landriault, 2010). Interested readers are referred to Asmussen and Albrecher (2010) and Badescu and Landriault (2009) for comprehensive reviews. Outside ruin theory, Hassan Zadeh and Stanford (2016) developed Bayesian and Bühlmann credibility theories under phase-type claims, whereas Wang *et al*. (2018) derived the distribution of the discounted aggregate claims under phase-type renewal process. Generalizations of phase-type distributions have also been proved to be successful for fitting insurance loss data. In particular, Ahn *et al*. (2012) utilized log phase-type distribution to fit the well-known Danish fire data, and Bladt and Yslas (2023a) proposed a phase-type mixture-of-experts regression model to fit the severities from the French Motor Third Party Liability dataset. Apart from modeling insurance losses and interclaim times, phase-type mortality law was studied by Lin and Liu (2007), and a review of the use of phase-type distributions in health care systems can be found in Fackrell (2009). While all the above works are concerned with continuous phase-type distributions, discrete phase-type (DPH)

© The Author(s), 2025. Published by Cambridge University Press on behalf of The International Actuarial Association.

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

distributions have also been applied to model claim frequencies. For example, Panjer-type recursions for the calculations of compound distributions under DPH claim counts were developed by Eisele (2006), Wu and Li (2010), and Ren (2010).

In modern risk theory and insurance mathematics, there is increasing need for researchers and practitioners to accurately model the dependencies among multiple risk factors or risk events, as this is crucial for effective risk management. Such dependence modeling is particularly important in scenarios where different risks are influenced by shared factors or catastrophic events – commonly referred to as *common shocks*, leading to realization of random variables that are strongly dependent. For example, an accident or a natural disaster can cause claims in both personal injury and property damage, resulting in correlated claim counts and dependent losses in different insurance contracts or business lines. Capturing dependencies is essential for a realistic assessment of joint behavior of risks. While the use of copulas is popular (see, e.g., Frees and Valdez, 1998 for a review), specific models for multivariate claim counts (e.g., Shi and Valdez, 2014; Pechon *et al*., 2018; Bolancé and Vernic, 2019; Fung *et al*., 2019) and multivariate losses (e.g., Lee and Lin, 2012; Willmot and Woo, 2015) have been developed as well. The presence of dependency has also led to related research topics in actuarial science, such as multivariate risk measures (see, e.g., Cossette *et al*., 2016; Landsman *et al*., 2016) and multivariate ruin problems (see, e.g., Badila *et al*., 2015; Albrecher *et al*., 2022).

Dependencies also commonly arise in the applied probability literature, and several approaches to constructing continuous multivariate phase-type distributions have been proposed in the field. These models often introduce dependencies by coupling the underlying Markov processes in specific ways, namely, by employing the exit times from different sets as in Assaf *et al*. (1984), or by using accumulating reward systems as in Kulkarni (1989). In Bladt and Nielsen (2010), a class of multivariate matrix-exponential (MME) distributions, characterized by rational multivariate Laplace transforms, was introduced. However, the multivariate matrix distributions in Kulkarni (1989) and Bladt and Nielsen (2010) do not necessarily admit closed-form probability density function in general, possibly limiting their applications. With an emphasis on explicit results and ease of applications in an actuarial context, recent efforts have also been made in Cheung *et al*. (2022), Bladt (2023), Albrecher *et al*. (2023) and Bladt and Yslas (2023b) to introduce new classes of distributions which are conditionally independent mixtures of matrix distributions. Specifically, in Bladt (2023), the author proposed to employ a system of Markov jump processes that agree on the initial state and evolve independently from that point onward.

Despite the increasing interest in continuous multivariate matrix distributions mentioned above, extensions of the DPH class to a multivariate setting, particularly in the context of matrix distributions, are relatively scarce in the literature. One of the first attempts is the work by He and Ren (2016a), who constructed a multivariate DPH distribution via the numbers of different types of batches that have arrived before absorption of a discrete-time Markov chain, with an accompanying estimation scheme discussed in He and Ren (2016b). A discrete analog of Kulkarni (1989)'s continuous multivariate phase-type distributions was considered in Navarro (2019); however, such a class, like its continuous counterpart, does not admit closed-form expressions for the joint probability mass function (pmf). A more recent development was presented by Bladt and Yslas (2023b), which is a discrete version of Bladt (2023)'s work, where the authors investigated the termination time of discrete-time Markov chains that begin in the same initial state but otherwise behave independently. However, many of these models impose dependencies that do not necessarily explicitly capture the common-shock scenarios that arise.

In this paper, we introduce a novel class of bivariate *common-shock discrete phase-type* (CDPH) distributions. Our approach extends the DPH framework to a bivariate setting by constructing two temporarily jointly evolving Markov chains that share a common evolution up to a random time – the occurrence of the common shock (or the number of common shocks) – after which they evolve independently. This construction effectively models the dependency induced by common shocks while retaining the tractability of DPH distributions. More precisely, we consider two discrete-time Markovian processes, M_1 and M_2, which evolve identically until a decoupling time $\tau_{1,2}$, after which they proceed independently and terminate at possibly different times τ_1 and τ_2. We derive explicit expressions for the joint distribution of (τ_1, τ_2), which characterize the CDPH distributions. This framework allows

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

us to model scenarios where two types of risks are affected by shared events before diverging due to independent factors. We can show that, in the bivariate case, our proposed class of CDPH distributions is larger than the class of multivariate DPH distributions in Bladt and Yslas (2023b) and is a subclass of that considered by Navarro (2019). Building on this foundation, we explore applications of the CDPH distributions in risk modeling involving random sums with common shocks. In particular, we consider aggregate risks represented as sums of random variables whose counts are determined by τ_1 and τ_2. This setup is particularly relevant in insurance, where aggregate claims may depend on both shared catastrophic events and independent claim occurrences. We characterize the joint distributions of these random sums, demonstrating the practicality and analytical convenience of our models.

Our contributions can be summarized as follows: We introduce the CDPH distributions, a new class of bivariate DPH distributions designed to effectively model dependencies induced by common shocks. To facilitate their applications in loss modeling, we provide explicit formulas for the joint pmf and probability generating function (pgf) of the CDPH class and prove various closure properties. Characterizations of the associated random sums are also provided. Furthermore, we develop estimation procedures for the CDPH distributions using the expectation-maximization (EM) algorithm, enabling their statistical implementation. Finally, we illustrate the applicability and effectiveness of our models through simulation studies, highlighting their potential in practical risk assessment scenarios. One of the key contributions of this work is to provide a robust and flexible framework for modeling discrete data subject to common shocks. Our construction not only offers an intuitive dependence mechanism and a tractable joint distribution, but most importantly, it also allows for the data-driven estimation of the underlying common shock distribution without restrictive parametric assumptions. This provides a powerful tool for analyzing latent shock components in correlated count data. In essence, our work extends the versatility of phase-type distributions in modeling the inherent complex dependencies arising in actuarial science and provides an explicit common-shock mechanism and tractable distributional forms for discrete data, thereby offering both theoretical insights and practical tools for actuaries and risk managers dealing with such correlated events.

While other powerful dependence models, such as copulas, are widely used, our CDPH framework offers a distinct mechanistic approach. Instead of merely describing the statistical association between variables, the CDPH model provides an interpretable probabilistic story, enabling the decomposition of each risk into an explicit common shock component and an individual-specific component. This not only facilitates a deeper understanding of the dependence structure but also maintains analytical tractability for related quantities, such as the aggregate loss distributions discussed in Section 3.

The remainder of the paper is organized as follows. In Section 2, we define the CDPH distributions and discuss their fundamental properties. Section 3 applies these distributions to analyze random sums. In Section 4, we address the estimation of the CDPH distributions from data, proposing an EM algorithm tailored to our models. Section 5 presents simulation studies that demonstrate the implementation and performance of our estimation procedures, as well as an application to a dataset comprising bivariate insurance claim frequencies. Finally, Section 6 concludes the paper and suggests directions for future research. The proofs of various propositions and some additional numerical examples are collected in the Appendix.

2 A new class of CDPH distributions

Throughout this paper, we let $\mathbb{N} := \{0, 1, 2, \dots\}$ denote the set of nonnegative integers and $\mathbb{N}_+ := \{1, 2, 3, \dots\}$ the set of positive integers. We use the notation $a\mathbb{N} + b$ for $a > 0$ to represent the arithmetic progression or lattice $\{b, a + b, 2a + b, 3a + b, \dots\}$.

This section provides some of the essential properties of CDPH distributions, starting from basic principles and then treating more intricate class properties. Throughout, we link our class to related constructions in the literature.

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

2.1 Background on DPH

Consider a Markov chain $M = \{M(n)\}_{n\in\mathbb{N}}$ evolving in a finite state-space $\mathcal{E} = \{1, 2, \ldots, d\}$ according to some initial probability vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ and subprobability matrix $\boldsymbol{P} = \{p_{ij}\}_{i,j\in\mathcal{E}}$. Here, all states in $\mathcal{E}$ are assumed to be transient, and thus, the process M will be (almost surely) terminating. The transience property is mathematically equivalent to the condition that the eigenvalues of $\boldsymbol{P}$ are strictly contained in the complex unit circle (see e.g., Corollary 1.2.63 in Bladt and Nielsen, 2017).

An alternative setup is to consider the augmented state-space $\mathcal{E} \cup \Delta$, where Δ is an absorbing cemetery state used to indicate the termination of M. In this setup, M evolves according to the probability vector $(\alpha_1, \alpha_2, \ldots, \alpha_d, 0)$ and a block-partitioned transition matrix

$$\begin{pmatrix} \boldsymbol{P} & \mathbf{1} - \boldsymbol{P}\mathbf{1} \\ \mathbf{0} & 1 \end{pmatrix},$$

where $\mathbf{1}$ denotes a column vector of ones of appropriate dimension (d in this case), $\boldsymbol{P}\mathbf{1}$ corresponds to the sum of probabilities from each state to all others in $\mathcal{E}$, and $\mathbf{0}$ denotes a zero matrix of appropriate dimension (a zero row vector of dimension d in this case). The final "1" is a scalar representing the probability of remaining in the state Δ once entered. Both setups – the terminating and the absorbing ones – are interchangeable for our purposes, and we switch back and forth depending on the situation.

Define τ to be the step in which M gets terminated, that is, let

$$\tau := \inf\{n \in \mathbb{N}_+ : M(n) \notin \mathcal{E}\}.$$

Then τ follows what is known as a DPH distribution, which we denote by DPH $(\boldsymbol{\alpha}, \boldsymbol{P})$. The class of DPH distributions is particularly tractable (see, e.g., Chapter 1.2.6 in Bladt and Nielsen, 2017). For instance, its pmf over $\mathbb{N}_+$ is explicitly given by

$$f_\tau(\ell) := \mathbb{P}(\tau = \ell) = \boldsymbol{\alpha}\boldsymbol{P}^{\ell-1}\left(\mathbf{1} - \boldsymbol{P}\mathbf{1}\right).$$

Likewise, the pgf takes the tractable form

$$\mathbb{E}(x^\tau) = x\boldsymbol{\alpha}(\boldsymbol{I} - x\boldsymbol{P})^{-1}\left(\mathbf{1} - \boldsymbol{P}\mathbf{1}\right),$$

where $\boldsymbol{I}$ denotes the identity matrix. Note that the above formula is valid for all $x \in \mathbb{R}$ on the real line such that the eigenvalues of $x\boldsymbol{P}$ lie within the complex unit circle (the set that contains the interval $[-1, 1]$).

2.2 Definition, joint probability, and joint moments of CDPH

As mentioned in the introduction, literature on the construction of multivariate DPH distributions is rather limited. Our paper is inspired by the approach in Bladt and Yslas (2023b), adding the flexibility that the Markov chains evolve synchronously for some period of time (rather than just starting in the same initial state) before proceeding independently. The exact details behind our novel construction are as follows. Consider two jointly evolving terminating processes, $M_1 = \{M_1(n)\}_{n\in\mathbb{N}}$ and $M_2 = \{M_2(n)\}_{n\in\mathbb{N}}$, each taking values in the state space $\mathcal{E} \cup \mathcal{S}$, where $\mathcal{E} \cap \mathcal{S} = \emptyset$ and $|\mathcal{E}| \wedge |\mathcal{S}| \geq 1$. We construct the joint process (M_1, M_2) such that each marginal process is a Markov chain, with a specific joint dependence structure: we let $M_1(n) = M_2(n)$ for all n up to the first time they exit the set $\mathcal{E}$. After this point, they evolve independently within $\mathcal{S}$ and eventually terminate, possibly at different times.

More explicitly, we let M_1 be a terminating Markov chain that evolves according to the block-partitioned initial probability vector $(\boldsymbol{\alpha}, \mathbf{0})$ and subprobability matrix

$$\boldsymbol{P}_1 := \begin{pmatrix} \boldsymbol{P} & \boldsymbol{U} \\ \mathbf{0} & \boldsymbol{Q}_1 \end{pmatrix}, \tag{2.1}$$

where $\boldsymbol{P}$ is a subprobability matrix over $\mathcal{E}$ (like before), $\boldsymbol{U}$ denotes the transition probabilities from $\mathcal{E}$ to $\mathcal{S}$, and $\boldsymbol{Q}_1$ is a subprobability matrix over $\mathcal{S}$. It is further assumed that each row of the submatrix $(\boldsymbol{P} \;\; \boldsymbol{U})$

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

sums to one, so that the Markov chain M_1 (and also M_2 below) must enter $\mathcal{S}$ before termination. For M_2, we define $M_2(n) = M_1(n)$ for all $n \leq \tau_{1,2}$, where

$$\tau_{1,2} := \inf\{n \in \mathbb{N}_+ : M_1(n) \notin \mathcal{E}\}.$$

Obviously, $\tau_{1,2}$ follows a DPH $(\boldsymbol{\alpha}, \boldsymbol{P})$ distribution. After time $\tau_{1,2}$, we let M_2 evolve independently as a terminating Markov chain within $\mathcal{S}$ according to the subprobability matrix $\boldsymbol{Q}_2$. This construction yields M_2 as a Markov chain in $\mathcal{E} \cup \mathcal{S}$ that evolves according to the block-partitioned initial probability vector $(\boldsymbol{\alpha}, \boldsymbol{0})$ and subprobability matrix

$$\boldsymbol{P}_2 := \begin{pmatrix} \boldsymbol{P} & \boldsymbol{U} \\ \boldsymbol{0} & \boldsymbol{Q}_2 \end{pmatrix}.$$

We now define the bivariate random variable (τ_1, τ_2) via the termination times of M_1 and M_2 so that, for $k \in \{1, 2\}$,

$$\tau_k := \inf\{n \in \mathbb{N}_+ : M_k(n) \notin \mathcal{E} \cup \mathcal{S}\}.$$

For each $k \in \{1, 2\}$, the marginal distribution of τ_k is clearly DPH $((\boldsymbol{\alpha}, \boldsymbol{0}), \boldsymbol{P}_k)$. Moreover, τ_1 and τ_2 are greater than $\tau_{1,2}$. Specifically, by writing $\tau_1 = \tau_{1,2} + (\tau_1 - \tau_{1,2})$ and $\tau_2 = \tau_{1,2} + (\tau_2 - \tau_{1,2})$, one observes that τ_1 and τ_2 are dependent for two reasons:

- Both τ_1 and τ_2 share the term $\tau_{1,2}$, the time until which the Markov chains M_1 and M_2 evolve identically. The random variable $\tau_{1,2}$ can be regarded as the common shock component.
- Although the Markov chains M_1 and M_2 evolve independently after time $\tau_{1,2}$, the remaining times until they get absorbed (namely $\tau_1 - \tau_{1,2}$ and $\tau_2 - \tau_{1,2}$) are only conditionally independent given the state $M_1(\tau_{1,2})$ of the Markov chain M_1 at the time $\tau_{1,2}$. Unconditionally, the variables $\tau_1 - \tau_{1,2}$ and $\tau_2 - \tau_{1,2}$ are dependent.

The joint distribution of (τ_1, τ_2) is provided in the following proposition.

Proposition 2.1. *For $n_1, n_2 \geq 2$, $m, z_1, z_2 \geq 1$, and $i \in \mathcal{E}$, define the unconditional and conditional joint pmf's of (τ_1, τ_2) by*

$$f_{\tau_1,\tau_2}(n_1, n_2) := \mathbb{P}(\tau_1 = n_1, \tau_2 = n_2),$$
$$f_{\tau_1,\tau_2|M_1,M_2}(n_1, n_2|i) := \mathbb{P}(\tau_1 = n_1, \tau_2 = n_2 \mid M_1(0) = M_2(0) = i),$$

and define the auxiliary functions

$$\psi(m, z_1, z_2) := \mathbb{P}(\tau_{1,2} = m, \tau_1 - \tau_{1,2} = z_1, \tau_2 - \tau_{1,2} = z_2),$$
$$\psi_i(m, z_1, z_2) := \mathbb{P}(\tau_{1,2} = m, \tau_1 - \tau_{1,2} = z_1, \tau_2 - \tau_{1,2} = z_2 \mid M_1(0) = M_2(0) = i).$$

Then, we have

$$f_{\tau_1,\tau_2}(n_1, n_2) = \sum_{m=1}^{\min\{n_1,n_2\}-1} \psi(m, n_1 - m, n_2 - m), \tag{2.2}$$

$$f_{\tau_1,\tau_2|M_1,M_2}(n_1, n_2|i) = \sum_{m=1}^{\min\{n_1,n_2\}-1} \psi_i(m, n_1 - m, n_2 - m), \tag{2.3}$$

and

$$\psi(m, z_1, z_2) = \boldsymbol{\alpha}\boldsymbol{P}^{m-1}\boldsymbol{U}\left((\boldsymbol{Q}_1^{z_1-1}\boldsymbol{q}_1) \odot (\boldsymbol{Q}_2^{z_2-1}\boldsymbol{q}_2)\right), \tag{2.4}$$

$$\psi_i(m, z_1, z_2) = \boldsymbol{e}_i^{\top}\boldsymbol{P}^{m-1}\boldsymbol{U}\left((\boldsymbol{Q}_1^{z_1-1}\boldsymbol{q}_1) \odot (\boldsymbol{Q}_2^{z_2-1}\boldsymbol{q}_2)\right), \tag{2.5}$$

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

where $\boldsymbol{q}_1 = \mathbf{1} - \boldsymbol{Q}_1\mathbf{1}$ *and* $\boldsymbol{q}_2 = \mathbf{1} - \boldsymbol{Q}_2\mathbf{1}$ *denote the one-step probabilities of termination from the states in* $\mathcal{S}$ *for* M_1 *and* M_2*, respectively;* $\boldsymbol{e}_i$ *denotes the* i*-th canonical column vector, and* $\odot$ *denotes the Hadamard (entrywise) product of two vectors.*

Using Proposition 2.1, one can readily show the following corollary.

Corollary 2.2. *For* $\zeta_0, \zeta_1, \zeta_2 \in [0,1]$*, the joint pgf of* $(\tau_{1,2}, \tau_1 - \tau_{1,2}, \tau_2 - \tau_{1,2})$ *is given by*

$$\mathbb{E}\left(\zeta_0^{\tau_{1,2}}\zeta_1^{\tau_1-\tau_{1,2}}\zeta_2^{\tau_2-\tau_{1,2}}\right) = \sum_{j\in\mathcal{S}} (\boldsymbol{\alpha}(\boldsymbol{I}\zeta_0^{-1} - \boldsymbol{P})^{-1}\boldsymbol{U})_j((\boldsymbol{I}\zeta_1^{-1} - \boldsymbol{Q}_1)^{-1}\boldsymbol{q}_1)_j((\boldsymbol{I}\zeta_2^{-1} - \boldsymbol{Q}_2)^{-1}\boldsymbol{q}_2)_j. \tag{2.6}$$

Moreover, the joint pgf of (τ_1, τ_2) *is given by*

$$\mathbb{E}\left(\zeta_1^{\tau_1}\zeta_2^{\tau_2}\right) = \sum_{j\in\mathcal{S}} (\boldsymbol{\alpha}(\boldsymbol{I}(\zeta_1\zeta_2)^{-1} - \boldsymbol{P})^{-1}\boldsymbol{U})_j((\boldsymbol{I}\zeta_1^{-1} - \boldsymbol{Q}_1)^{-1}\boldsymbol{q}_1)_j((\boldsymbol{I}\zeta_2^{-1} - \boldsymbol{Q}_2)^{-1}\boldsymbol{q}_2)_j. \tag{2.7}$$

Employing Corollary 2.2, the following result regarding the corresponding joint factorial moments can be obtained, which can in turn be used to compute the joint moments of τ_1 and τ_2 (see Remark 2.1) Here, for $n \in \mathbb{N}$, we let $x^{[n]} = x(x-1)(x-2)\cdots(x-n+1)$ denote the falling factorial of x.

Corollary 2.3. *For* $n_0, n_1, n_2 \in \mathbb{N}$*, the joint factorial moments of* $(\tau_{1,2}, \tau_1 - \tau_{1,2}, \tau_2 - \tau_{1,2})$ *are given by*

$$\mathbb{E}\big(\tau_{1,2}^{[n_0]}\,(\tau_1 - \tau_{1,2})^{[n_1]}\,(\tau_2 - \tau_{1,2})^{[n_2]}\big) = \sum_{j\in\mathcal{S}} (\boldsymbol{\alpha}\widetilde{\boldsymbol{P}}(n_0)\boldsymbol{U})_j(\widetilde{\boldsymbol{Q}}_1(n_1)\boldsymbol{q}_1)_j(\widetilde{\boldsymbol{Q}}_2(n_2)\boldsymbol{q}_2)_j, \tag{2.8}$$

where

$$\begin{aligned}
\widetilde{\boldsymbol{P}}(n_0) &:= \frac{\partial^{n_0}}{\partial\zeta_1^{n_0}}(\boldsymbol{I}\zeta_0^{-1} - \boldsymbol{P})^{-1}\Big|_{\zeta_0=1} = n_0!\boldsymbol{P}^{n_0-1}(\boldsymbol{I}-\boldsymbol{P})^{-n_0-1}, \qquad n_0 \in \mathbb{N}_+,\\
\widetilde{\boldsymbol{Q}}_1(n_1) &:= \frac{\partial^{n_1}}{\partial\zeta_1^{n_1}}(\boldsymbol{I}\zeta_1^{-1} - \boldsymbol{Q}_1)^{-1}\Big|_{\zeta_1=1} = n_1!\boldsymbol{Q}_1^{n_1-1}(\boldsymbol{I}-\boldsymbol{Q}_1)^{-n_1-1}, \qquad n_1 \in \mathbb{N}_+,\\
\widetilde{\boldsymbol{Q}}_2(n_2) &:= \frac{\partial^{n_2}}{\partial\zeta_2^{n_2}}(\boldsymbol{I}\zeta_2^{-1} - \boldsymbol{Q}_2)^{-1}\Big|_{\zeta_2=1} = n_2!\boldsymbol{Q}_2^{n_2-1}(\boldsymbol{I}-\boldsymbol{Q}_2)^{-n_2-1}, \qquad n_2 \in \mathbb{N}_+,
\end{aligned}$$

with $\widetilde{\boldsymbol{P}}(0) = (\boldsymbol{I}-\boldsymbol{P})^{-1}$*,* $\widetilde{\boldsymbol{Q}}_1(0) = (\boldsymbol{I}-\boldsymbol{Q}_1)^{-1}$ *and* $\widetilde{\boldsymbol{Q}}_2(0) = (\boldsymbol{I}-\boldsymbol{Q}_2)^{-1}$*.*

Remark 2.1. The joint moments of τ_1 and τ_2 can indeed be obtained from Corollary 2.3 as follows. For $n_1, n_2 \in \mathbb{N}$, we start by writing $\tau_1 = \tau_{1,2} + (\tau_1 - \tau_{1,2})$ and $\tau_2 = \tau_{1,2} + (\tau_2 - \tau_{1,2})$ followed by applying binomial expansions of their power to see that

$$\mathbb{E}\big(\tau_1^{n_1}\,\tau_2^{n_2}\big) = \sum_{j_1=0}^{n_1}\binom{n_1}{j_1}\sum_{j_2=0}^{n_2}\binom{n_2}{j_2}\mathbb{E}\big(\tau_{1,2}^{j_1+j_2}(\tau_1 - \tau_{1,2})^{n_1-j_1}(\tau_2 - \tau_{1,2})^{n_2-j_2}\big). \tag{2.9}$$

Further using the identity

$$x^n = \sum_{k=0}^{n}\left\{{n \atop k}\right\}x^{[k]}$$

for $n \in \mathbb{N}$, where

$$\left\{{n \atop k}\right\} := \frac{1}{k!}\sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}j^n$$

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

is the Stirling number of the second kind (see, e.g., Chapter 24.1.4 in Abramowitz and Stegun, 1972), (2.9) can be rewritten as

$$\mathbb{E}\big(\tau_1^{n_1}\ \tau_2^{n_2}\big) = \sum_{j_1=0}^{n_1}\binom{n_1}{j_1}\sum_{j_2=0}^{n_2}\binom{n_2}{j_2}\sum_{i=0}^{j_1+j_2}\left\{{j_1+j_2 \atop i}\right\}\sum_{\ell_1=0}^{n_1-j_1}\left\{{n_1-j_1 \atop \ell_1}\right\}\sum_{\ell_2=0}^{n_2-j_2}\left\{{n_2-j_2 \atop \ell_2}\right\}$$
$$\times\ \mathbb{E}\big(\tau_{1,2}^{[i]}(\tau_1-\tau_{1,2})^{[\ell_1]}(\tau_2-\tau_{1,2})^{[\ell_2]}\big).$$

That is, the joint moments of τ_1 and τ_2 can be represented as a linear combination of the joint factorial moments of $(\tau_{1,2}, \tau_1-\tau_{1,2}, \tau_2-\tau_{1,2})$ available in (2.8).

Note that the support of (τ_1, τ_2) is $(\mathbb{N}+2)\times(\mathbb{N}+2)$. For modeling purposes, it is often useful to consider discrete random variables that have a lattice different from $\mathbb{N}+2$. Thus, for $k_1, k_2 \in \mathbb{R}$ and $c_1, c_2 > 0$, one can define the vector (N_1, N_2) where

$$N_1 := c_1(\tau_1-2)+k_1, \qquad N_2 := c_2(\tau_2-2)+k_2. \tag{2.10}$$

We then say in a broad sense (see Remark 2.3) that the bivariate random variable (N_1, N_2) follows a bivariate CDPH distribution with support on $(c_1\mathbb{N}+k_1)\times(c_2\mathbb{N}+k_2)$ and parameters $(\boldsymbol{\alpha}, \boldsymbol{P}, \boldsymbol{U}, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$. Employing Proposition 2.1, we obtain the following immediate result.

Corollary 2.4. *Let (N_1, N_2) follow a bivariate* CDPH *distribution with support $(c_1\mathbb{N}+k_1)\times(c_2\mathbb{N}+k_2)$ and parameters $(\boldsymbol{\alpha}, \boldsymbol{P}, \boldsymbol{U}, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$. Then, the joint pmf of (N_1, N_2) for $x_1 \in c_1\mathbb{N}+k_1$ and $x_2 \in c_2\mathbb{N}+k_2$, namely $f_{N_1,N_2}(x_1,x_2) := \mathbb{P}(N_1=x_1,\ N_2=x_2)$, is given by*

$$f_{N_1,N_2}(x_1,x_2) = f_{\tau_1,\tau_2}\left(\frac{x_1-k_1}{c_1}+2,\ \frac{x_2-k_2}{c_2}+2\right),$$

where $f_{\tau_1,\tau_2}(n_1,n_2) = \mathbb{P}(\tau_1=n_1,\ \tau_2=n_2)$.

2.3 Relationships with other bivariate DPH distributions

Let us first note that the (bivariate version of) the so-called mDPH class considered in Bladt and Yslas (2023b) is a subclass of our proposed CDPH. Indeed, if we let $\boldsymbol{\beta}$ be a probability row vector, then choosing the parameters $\boldsymbol{\alpha}=(1)$, $\boldsymbol{P}=(0)$ and $\boldsymbol{U}=\boldsymbol{\beta}$ for a CDPH over the lattice $(\mathbb{N}+1)\times(\mathbb{N}+1)$ (i.e., $c_1=c_2=1$ and $k_1=k_2=-1$) yields an mDPH distribution. This distribution arises from starting two terminating processes in the same state according to the initial probability vector $\boldsymbol{\beta}$ then and evolving them independently from there onward according to the subprobability matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, respectively.

Remark 2.2. At first glance, one may be mistaken that the proposed additive shock structure would impose a positive dependence structure upfront between the two random variables in a bivariate CDPH distribution. However, our CDPH class is indeed more flexible because it contains, as a special case, the bivariate version of Bladt and Yslas (2023b)'s mDPH class, which is known to be dense in the space of bivariate discrete distributions on $(\mathbb{N}+1)\times(\mathbb{N}+1)$ (see Theorem 3.10 therein). Therefore, in principle, it can approximate any dependence structure, including both positive and negative correlation of different strength. In particular, once the CDPH model parameters have been estimated from a given dataset via the EM algorithm in Section 4, the correlation can be calculated using Remark 2.1 with the help of Corollary 2.3. Interested readers are also referred to Remark 2.1 in Albrecher et al. (2023), which discusses how negative dependence may arise between $\tau_1-\tau_{1,2}$ and $\tau_2-\tau_{1,2}$ (albeit in the setting of continuous random vectors).

In turn, the CDPH with support in $(\mathbb{N}+2)\times(\mathbb{N}+2)$ forms a subclass of the MDPH* class introduced in Chapter 5.2 of Navarro (2019). These distributions accumulate integer-valued rewards across each occupation epoch within a shared Markov chain, rather than bifurcating into independent processes after a common shock component concludes. Specifically, let us consider a partitioned state space $\mathcal{E}\cup(\mathcal{S}\times\mathcal{S})\cup(\{\partial_1\}\times\mathcal{S})\cup(\mathcal{S}\times\{\partial_2\})$ and a Markov chain $\gamma=\{\gamma(n)\}_{n\in\mathbb{N}}$ defined over it. The chain

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

evolves within $\mathcal{E}$ while the common shock is ongoing, then jumps to $\mathcal{S}\times\mathcal{S}$ (here we take the lexicographic order of $\mathcal{S}\times\mathcal{S}$) while the independent components are active, and finally either terminates, jumps to $\{\partial_1\}\times\mathcal{S}$, or jumps to $\mathcal{S}\times\{\partial_2\}$. These transitions occur depending on whether the independent components terminate at the same time, the first coordinate terminates before the second, or the second coordinate terminates before the first. Thus, we define the bivariate random vector (τ_1,τ_2) as

$$\tau_1 := \sum_{\ell=0}^{\sigma-1} \mathbb{1}_{\gamma(\ell)\in\mathcal{E}\cup(\mathcal{S}\times\mathcal{S})\cup(\mathcal{S}\times\{\partial_2\})}, \qquad \tau_2 := \sum_{\ell=0}^{\sigma-1} \mathbb{1}_{\gamma(\ell)\in\mathcal{E}\cup(\mathcal{S}\times\mathcal{S})\cup(\{\partial_1\}\times\mathcal{S})}, \tag{2.11}$$

where σ corresponds to the termination time of γ so that $\sigma := \inf\{n\in\mathbb{N}_+ : \gamma(n)\notin\mathcal{E}\cup(\mathcal{S}\times\mathcal{S})\cup(\{\partial_1\}\times\mathcal{S})\cup(\mathcal{S}\times\{\partial_2\})\}$. The above representation of the bivariate vector (τ_1,τ_2) is already in the form of Equation (5.5) of Navarro (2019). It remains to specify the transition probability of the terminating Markov chain γ.

To mimic the distributional properties of the original processes M_1 and M_2 embedded in the Markov chain γ, we take the block-partitioned parameters $(\boldsymbol{\alpha},\mathbf{0},\mathbf{0},\mathbf{0})$ and the subprobability matrix

$$\boldsymbol{P}_{\max} := \begin{pmatrix} \boldsymbol{P} & \boldsymbol{U}^* & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}_1\otimes\boldsymbol{Q}_2 & \boldsymbol{q}_1\otimes\boldsymbol{Q}_2 & \boldsymbol{Q}_1\otimes\boldsymbol{q}_2 \\ \mathbf{0} & \mathbf{0} & \boldsymbol{Q}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{Q}_1 \end{pmatrix}, \tag{2.12}$$

where $\boldsymbol{U}^* = \{u^*_{i,(j_1,j_2)}\}_{i\in\mathcal{E},(j_1,j_2)\in\mathcal{S}\times\mathcal{S}}$ is defined by $u^*_{i,(j_1,j_2)} := u_{ij_1}\delta_{j_1,j_2}$, with δ_{j_1,j_2} denoting the Kronecker delta and u_{ij} the (i,j)-th element of $\boldsymbol{U}$, and $\otimes$ represents the Kronecker product between matrices. Then, σ is DPH $((\boldsymbol{\alpha},\mathbf{0},\mathbf{0},\mathbf{0}),\boldsymbol{P}_{\max})$. Note that the resulting parameters are similar to those used to compute the distribution of the maximum between two discrete phase-type distributions in Theorem 1.2.67 of Bladt and Nielsen (2017). The derivation of these parameters for the process γ is as follows:

- The process evolves in $\mathcal{E}$ according to $\boldsymbol{P}$, mimicking the common shock component $\tau_{1,2}$, eventually jumping from a state $i\in\mathcal{E}$ to a state $(j,j)\in\mathcal{S}\times\mathcal{S}$ with probability u_{ij}.
- It then evolves in $\mathcal{S}\times\mathcal{S}$ according to $\boldsymbol{Q}_1\otimes\boldsymbol{Q}_2$ such that a transition from $(i_1,i_2)\in\mathcal{S}\times\mathcal{S}$ to $(j_1,j_2)\in\mathcal{S}\times\mathcal{S}$ occurs with probability $q_{1,i_1j_1}\,q_{2,i_2j_2}$ (where, for $k=1,2$, we use $q_{k,ij}$ to denote the (i,j)-th element of $\boldsymbol{Q}_k$). Then, either both components terminate simultaneously according to the exit column vector $\boldsymbol{q}_1\otimes\boldsymbol{q}_2$; the first terminates while the second remains active, moving to $\{\partial_1\}\times\mathcal{S}$; or the second terminates while the first remains active, moving to $\mathcal{S}\times\{\partial_2\}$.
- If the process jumps to $\{\partial_1\}\times\mathcal{S}$ or $\mathcal{S}\times\{\partial_2\}$, it continues evolving according to the matrix $\boldsymbol{Q}_2$ or $\boldsymbol{Q}_1$, respectively.

The above results are summarized in the following proposition.

Proposition 2.5. *Concerning bivariate* DPH *distributions, our proposed* CDPH *class satisfies the relationships*

$$\text{mDPH} \subset \text{CDPH (with support } (\mathbb{N}+1)\times(\mathbb{N}+1)),$$
$$\text{CDPH (with support } (\mathbb{N}+2)\times(\mathbb{N}+2)) \subset \text{MDPH}^*,$$

where mDPH *is the class introduced by Bladt and Yslas (2023b) and* MDPH* *is the class introduced by Navarro (2019).*

It is important to emphasize the practical advantages of our CDPH class relative to the more general MDPH* framework from Navarro (2019). While CDPH is a subclass of MDPH*, its tailored construction for common shocks yields an explicit and closed-form joint pmf (see Proposition 2.1). This analytical

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

tractability, which is not a feature of the general MDPH* class, is the key that enables direct likelihood-based inference and allows for the data-driven estimation of the underlying common shock distribution, as implemented via the EM algorithm in Section 4.

Remark 2.3. Since the general CDPH distributed random vector (N_1, N_2) in (2.10) with support on $(c_1\mathbb{N}+k_1)\times(c_2\mathbb{N}+k_2)$ is merely linear transformation of its basic version (τ_1,τ_2) which has support on $(\mathbb{N}+2)\times(\mathbb{N}+2)$, from now we shall simply use CDPH to denote the class with support on $(\mathbb{N}+2)\times(\mathbb{N}+2)$, and say that (τ_1,τ_2) follows a CDPH $(\boldsymbol{\alpha},\boldsymbol{P},\boldsymbol{U},\boldsymbol{Q}_1,\boldsymbol{Q}_2)$ distribution.

2.4 Further distributional and closure properties

Below, we explore further properties of the CDPH class. In particular, we investigate certain operations on the vector (τ_1,τ_2) that belong to the class of DPH or CDPH distributions. Apart from theoretical significance, such closure properties can also be of practical importance. For example, closure of the CDPH class under mixtures and sums means that the resulting bivariate random vectors still follow CDPH distributions but with updated parameters, and the closed-form joint pmf of the CDPH class is still conveniently applicable for computation purposes.

Minimum and maximum between τ_1 and τ_2

Consider again the Markov chain γ driven by (2.12), with initial vector $(\boldsymbol{\alpha},\mathbf{0},\mathbf{0},\mathbf{0})$, state-space $\mathcal{E}\cup(\mathcal{S}\times\mathcal{S})\cup(\{\partial_1\}\times\mathcal{S})\cup(\mathcal{S}\times\{\partial_2\})$, and termination time σ. One can easily see that σ indeed corresponds to the maximum $\tau_{\max}:=\max\{\tau_1,\tau_2\}$ of τ_1 and τ_2. Moreover, the minimum $\tau_{\min}:=\min\{\tau_1,\tau_2\}$ is simply the exit time of the Markov chain γ from the set $\mathcal{E}\cup(\mathcal{S}\times\mathcal{S})$. Defining the matrix

$$\boldsymbol{P}_{\min}:=\begin{pmatrix}\boldsymbol{P} & \boldsymbol{U}^*\\ \mathbf{0} & \boldsymbol{Q}_1\otimes\boldsymbol{Q}_2\end{pmatrix}, \tag{2.13}$$

the next proposition follows as a direct consequence.

Proposition 2.6. *Given a* CDPH $(\boldsymbol{\alpha},\boldsymbol{P},\boldsymbol{U},\boldsymbol{Q}_1,\boldsymbol{Q}_2)$ *random vector* (τ_1,τ_2)*, the maximum* $\tau_{\max}$ *follows a* DPH $((\boldsymbol{\alpha},\mathbf{0},\mathbf{0},\mathbf{0}),\boldsymbol{P}_{\max})$ *distribution and the minimum* $\tau_{\min}$ *follows a* DPH $((\boldsymbol{\alpha},\mathbf{0}),\boldsymbol{P}_{\min})$ *distribution, where* $\boldsymbol{P}_{\max}$ *and* $\boldsymbol{P}_{\min}$ *are given by (2.12) and (2.13), respectively.*

Aggregation of coordinates $\tau_1+\tau_2$

When considering the sum of the coordinates of (τ_1,τ_2), it is helpful to represent it as $\tau_1+\tau_2=2\min\{\tau_1,\tau_2\}+(\max\{\tau_1,\tau_2\}-\min\{\tau_1,\tau_2\})$. That is, $\tau_1+\tau_2$ counts twice the steps in which both τ_1 and τ_2 have not terminated and then adds the steps during which the larger value exceeds the smaller one. Consequently, we have

$$\tau_1+\tau_2=\sum_{\ell=0}^{\sigma-1}\left(2\cdot\mathbb{1}_{\gamma(\ell)\in\mathcal{E}\cup(\mathcal{S}\times\mathcal{S})}+\mathbb{1}_{\gamma(\ell)\in(\{\partial_1\}\times\mathcal{S})\cup(\mathcal{S}\times\{\partial_2\})}\right). \tag{2.14}$$

We have the following proposition.

Proposition 2.7. *Given a* CDPH $(\boldsymbol{\alpha},\boldsymbol{P},\boldsymbol{U},\boldsymbol{Q}_1,\boldsymbol{Q}_2)$ *random vector* (τ_1,τ_2)*, the sum* $\tau_1+\tau_2$ *follows a* DPH *distribution.*

Mixture of CDPH random vectors

For each $i\in\{1,2,\ldots,m\}$, suppose that the random vector $(\tau_1(i),\tau_2(i))$ is CDPH distributed. Let a mixture be formed where the component $(\tau_1(i),\tau_2(i))$ is chosen with probability $w_i>0$. The following proposition shows that the class of CDPH distributions is closed under finite mixture.

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

Proposition 2.8. *Suppose that a mixture is formed such that a bivariate random vector* $(\tau_{1,\text{mix}}, \tau_{2,\text{mix}})$ *is chosen to be* $(\tau_1(i), \tau_2(i))$ *with probability* $w_i > 0$, *such that* $(\tau_1(i), \tau_2(i))$ *is* $\text{CDPH}(\boldsymbol{\alpha}(i), \boldsymbol{P}(i), \boldsymbol{U}(i), \boldsymbol{Q}_1(i), \boldsymbol{Q}_2(i))$ *distributed and* $\sum_{i=1}^m w_i = 1$. *Then,* $(\tau_{1,\text{mix}}, \tau_{2,\text{mix}})$ *follows a* $\text{CDPH}(\boldsymbol{\alpha}_{\text{mix}}, \boldsymbol{P}_{\text{mix}}, \boldsymbol{U}_{\text{mix}}, \boldsymbol{Q}_{1,\text{mix}}, \boldsymbol{Q}_{2,\text{mix}})$ *distribution, where the parameters are given by*

$$\boldsymbol{\alpha}_{\text{mix}} := \begin{pmatrix} w_1\boldsymbol{\alpha}(1) & w_2\boldsymbol{\alpha}(2) & \cdots & w_m\boldsymbol{\alpha}(m) \end{pmatrix},$$

$$\boldsymbol{P}_{\text{mix}} := \begin{pmatrix} \boldsymbol{P}(1) & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}(2) & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{P}(m) \end{pmatrix}, \quad \boldsymbol{U}_{\text{mix}} := \begin{pmatrix} \boldsymbol{U}(1) & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}(2) & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{U}(m) \end{pmatrix},$$

$$\boldsymbol{Q}_{1,\text{mix}} := \begin{pmatrix} \boldsymbol{Q}_1(1) & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_1(2) & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{Q}_1(m) \end{pmatrix}, \quad \boldsymbol{Q}_{2,\text{mix}} := \begin{pmatrix} \boldsymbol{Q}_2(1) & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2(2) & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{Q}_2(m) \end{pmatrix}.$$

Sum of CDPH random vectors

Suppose that the random vector $(\tau_1(i), \tau_2(i))$ is CDPH distributed for each $i \in \{1, 2, \ldots, m\}$, and the random vectors $\{(\tau_1(i), \tau_2(i))\}_{i=1}^m$ are independent. Then one could be interested in the joint distribution of the sum $(\tau_{1,\text{sum}}, \tau_{2,\text{sum}}) := (\sum_{i=1}^m \tau_1(i), \sum_{i=1}^m \tau_2(i))$. Indeed, this follows a CDPH distribution. In the next proposition, we consider the case $m = 2$ and explicitly state the resulting parameters.

Proposition 2.9. *Consider the bivariate random vector* $(\tau_{1,\text{sum}}, \tau_{2,\text{sum}}) = (\tau_1(1) + \tau_1(2), \tau_2(1) + \tau_2(2))$, *where* $(\tau_1(i), \tau_2(i))$ *is* $\text{CDPH}(\boldsymbol{\alpha}(i), \boldsymbol{P}(i), \boldsymbol{U}(i), \boldsymbol{Q}_1(i), \boldsymbol{Q}_2(i))$ *distributed for* $i \in \{1, 2\}$, *and the random vectors* $(\tau_1(1), \tau_2(1))$ *and* $(\tau_1(2), \tau_2(2))$ *are independent. Then,* $(\tau_{1,\text{sum}}, \tau_{2,\text{sum}})$ *follows a* $\text{CDPH}(\boldsymbol{\alpha}_{\text{sum}}, \boldsymbol{P}_{\text{sum}}, \boldsymbol{U}_{\text{sum}}, \boldsymbol{Q}_{1,\text{sum}}, \boldsymbol{Q}_{2,\text{sum}})$ *distribution, where the parameters are given by*

$$\boldsymbol{\alpha}_{\text{sum}} := (\boldsymbol{\alpha}(1), \boldsymbol{0}), \tag{2.15}$$

$$\boldsymbol{P}_{\text{sum}} := \begin{pmatrix} \boldsymbol{P}(1) & \boldsymbol{U}(1) \otimes \boldsymbol{\alpha}(2) \\ \boldsymbol{0} & \boldsymbol{I} \otimes \boldsymbol{P}(2) \end{pmatrix}, \quad \boldsymbol{U}_{\text{sum}} := \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{I} \otimes \boldsymbol{U}(2) & \boldsymbol{0} \end{pmatrix}, \tag{2.16}$$

$$\boldsymbol{Q}_{1,\text{sum}} := \begin{pmatrix} \boldsymbol{Q}_1(1) \otimes \boldsymbol{I} & \boldsymbol{q}_1(1) \otimes \boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{Q}_1(2) \end{pmatrix}, \quad \boldsymbol{Q}_{2,\text{sum}} := \begin{pmatrix} \boldsymbol{I} \otimes \boldsymbol{Q}_2(1) & \boldsymbol{I} \otimes \boldsymbol{q}_2(1) \\ \boldsymbol{0} & \boldsymbol{Q}_2(2) \end{pmatrix}, \tag{2.17}$$

with $\boldsymbol{q}_1(1) = \boldsymbol{1} - \boldsymbol{Q}_1(1)\boldsymbol{1}$ *and* $\boldsymbol{q}_2(1) = \boldsymbol{1} - \boldsymbol{Q}_2(1)\boldsymbol{1}$. *The state space corresponding to the transitions specified by* $\boldsymbol{P}_{\text{sum}}$ *is* $\mathcal{E}_{\text{sum}} := \mathcal{E}(1) \cup (\mathcal{S}(1) \times \mathcal{E}(2))$, *and that for the transitions specified by* $\boldsymbol{Q}_{1,\text{sum}}$ *and* $\boldsymbol{Q}_{2,\text{sum}}$ *is* $\mathcal{S}_{\text{sum}} := (\mathcal{S}(1) \times \mathcal{S}(2)) \cup \mathcal{S}(2)$. *Here, with the obvious notation,* $\mathcal{E}(i) \cup \mathcal{S}(i)$ *is the state space of the underlying terminating Markov chains that define the vector* $(\tau_1(i), \tau_2(i))$.

It is clear that $(\tau_{1,\text{sum}}, \tau_{2,\text{sum}}) = (\sum_{i=1}^m \tau_1(i), \sum_{i=1}^m \tau_2(i))$ for values of m greater than two still follows a CDPH distribution by summing the CDPH bivariate random vectors one at a time and using Proposition 2.9 repeatedly.

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

3 Compound sums under bivariate CDPH claim count

Having defined the class of bivariate CDPH distributions, we now turn our attention to a class of bivariate random variables (Y_1, Y_2) with a random sum representation

$$Y_1 := \sum_{\ell=1}^{\tau_1} X_{1,\ell}, \qquad Y_2 := \sum_{\ell=1}^{\tau_2} X_{2,\ell}, \tag{3.1}$$

where, for each coordinate $k \in \{1, 2\}$, we aggregate a random number of positive variables $X_{k,1}, X_{k,2}, X_{k,3}, \ldots$, with the total number $\tau_k \in \mathbb{N}$ drawn from a bivariate CDPH distribution. In actuarial science, (τ_1, τ_2) can be interpreted as bivariate claim count for correlated business lines. For $k \in \{1, 2\}$, if $X_{k,1}, X_{k,2}, X_{k,3}, \ldots$ are the individual claim amounts specific to line k, then the compound sum Y_k corresponds to the aggregate loss for line k. Before analyzing the distribution of (Y_1, Y_2) in detail, we briefly outline the assumptions we make regarding the sequence $\{(X_{1,\ell}, X_{2,\ell})\}_{\ell\in\mathbb{N}_+}$. First, we assume that $\{(X_{1,\ell}, X_{2,\ell})\}_{\ell\in\mathbb{N}_+}$ is a sequence of independent and identically distributed bivariate random vectors following the same distribution as a reference random vector (X_1, X_2). Second, the sequence $\{(X_{1,\ell}, X_{2,\ell})\}_{\ell\in\mathbb{N}_+}$ is independent of (τ_1, τ_2). While the CDPH distributed (τ_1, τ_2) has support on $(\mathbb{N} + 2) \times (\mathbb{N} + 2)$, the cases where the counting variables are replaced by more general CDPH distributed (N_1, N_2) with support on $(\mathbb{N} + k_1) \times (\mathbb{N} + k_2)$ for $k_1, k_2 \in \mathbb{N}$ can be addressed by modifying our arguments with minimal technical effort (see Remark 2.3).

We shall consider two scenarios for X_1 and X_2: the case where they are independent, discussed in Subsection 3.1, and the case where they are dependent, addressed in Subsection 3.2.

3.1 Independent summands

Suppose that X_1 is independent of X_2. Below, we provide information about the distribution of (Y_1, Y_2) under this assumption.

Proposition 3.1. *For $\theta_1, \theta_2 \geq 0$, the joint Laplace transform $\mathcal{L}_{Y_1,Y_2}(\theta_1, \theta_2) := \mathbb{E}\left(\exp\left(-\theta_1 Y_1 - \theta_2 Y_2\right)\right)$ of the bivariate random sum in (3.1) is given by*

$$\mathcal{L}_{Y_1,Y_2}(\theta_1, \theta_2) = \boldsymbol{\alpha}\left(\boldsymbol{I} a_{1,\theta_1}^{-1} a_{2,\theta_2}^{-1} - \boldsymbol{P}\right)^{-1} \boldsymbol{U}\left(\left(\left(\boldsymbol{I} a_{1,\theta_1}^{-1} - \boldsymbol{Q}_1\right)^{-1} \boldsymbol{q}_1\right) \odot \left(\left(\boldsymbol{I} a_{2,\theta_2}^{-1} - \boldsymbol{Q}_2\right)^{-1} \boldsymbol{q}_2\right)\right),$$

where $a_{k,\theta_k} := \mathbb{E}\left(e^{-\theta_k X_k}\right)$ is the Laplace transform of X_k for $k \in \{1, 2\}$.

Proposition 3.1 provides a characterization of the distribution of (Y_1, Y_2) in terms of the joint Laplace transform while leaving the distributions of X_1 and X_2 unspecified. It is easy to see that if a_{1,θ_1} and a_{2,θ_2} are rational in θ_1 and θ_2, respectively, then $\mathcal{L}_{Y_1,Y_2}(\theta_1, \theta_2)$ is rational in θ_1 and θ_2. Since univariate rational Laplace transform characterizes matrix-exponential distributions (see, e.g., Chapter 4 in Bladt and Nielsen, 2017), while multivariate rational Laplace transform characterizes the class of MME distributions proposed by Bladt and Nielsen (2010), we have the following corollary as an immediate consequence.

Corollary 3.2. *If X_1 and X_2 follow independent matrix-exponential distributions, then the bivariate random vector (Y_1, Y_2) defined in (3.1) belongs to the class of* MME *distributions introduced in Bladt and Nielsen (2010).*

In the following subsection, we establish a related result by showing that if (X_1, X_2) belongs to the subclass of multivariate phase-type distributions (MPH*) proposed in Kulkarni (1989) then so does (Y_1, Y_2). The main advantage of MPH* distributions is that, unlike the MME class, they possess a simple probabilistic interpretation.

3.2 Dependent summands

Here we assume that (X_1, X_2) follows a bivariate version of Kulkarni's MPH* class, which we briefly describe next. Such an MPH* distribution is constructed using a single terminating continuous-time

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

Markov jump process $J = \{J(t)\}_{t\geq 0}$ with finite state space $\mathcal{C}$ and a system of state-dependent rewards. Specifically, let $(\boldsymbol{\pi}, \boldsymbol{S}, \boldsymbol{R})$ be the parameters, where

- $\boldsymbol{\pi}$ is the initial probability vector;
- $\boldsymbol{S}$ is the subintensity matrix governing the continuous-time transitions between transient states; and
- $\boldsymbol{R} = \{r_{kj}\}$ is the reward matrix, where r_{kj} represents the instantaneous nonnegative reward associated to the j-th state in $\mathcal{C}$ and k-th coordinate for $k \in \{1, 2\}$.

With these elements, J evolves according to $(\boldsymbol{\pi}, \boldsymbol{S})$ and the coordinates X_1 and X_2 are defined via

$$X_k := \int_0^{\delta} r_{k,J(s)}\, ds, \qquad k \in \{1, 2\},$$

where $\delta := \inf\{t \geq 0 : J(t) \notin \mathcal{C}\}$ is the termination time of J. The random vector (X_1, X_2) obtained through this construction is said to follow an MPH* $(\boldsymbol{\pi}, \boldsymbol{S}, \boldsymbol{R})$ distribution.

This construction allows for capturing dependencies between X_1 and X_2 through the shared underlying Markov process J. Apart from including independence as a special case, the class of MPH* distributions is known to be dense within the class of bivariate distributions with support on the positive quadrant. Unfortunately, the MPH* class is not particularly tractable in the general case as the joint distribution function cannot be expressed in closed form. However, its joint Laplace transform is known to be

$$\mathbb{E}\left(e^{-\theta_1 X_1 - \theta_2 X_2}\right) = \boldsymbol{\pi}\left(\boldsymbol{\Delta}_{\theta_1,\theta_2,\boldsymbol{R}} - \boldsymbol{S}\right)^{-1}(-\boldsymbol{S}\mathbf{1}),$$

for $\theta_1, \theta_2 \geq 0$, where

$$\boldsymbol{\Delta}_{\theta_1,\theta_2,\boldsymbol{R}} = \operatorname{diag}\{\theta_1 r_{1j} + \theta_2 r_{2j} : j \in \mathcal{C}\}.$$

Our aim is to prove that, under this setup, the bivariate random vector (Y_1, Y_2) defined in (3.1) with each $(X_{1,\ell}, X_{2,\ell})$ distributed as the generic bivariate random vector (X_1, X_2) follows an MPH* distribution as well.

Proposition 3.3. *Suppose that each bivariate random vector in the sequence* $\{(X_{1,\ell}, X_{2,\ell})\}_{\ell\in\mathbb{N}_+}$ *follows an* MPH* *distribution with parameters* $(\boldsymbol{\pi}, \boldsymbol{S}, \boldsymbol{R})$. *Then the pair* (Y_1, Y_2) *defined in (3.1) follows an* MPH* *distribution with an underlying continuous-time Markov jump process on the state-space* $\big(\mathcal{E} \cup (\mathcal{S} \times \mathcal{S}) \cup (\{\partial_1\} \times \mathcal{S}) \cup (\mathcal{S} \times \{\partial_2\})\big) \times \mathcal{C}$, *initial probability vector*

$$\boldsymbol{\pi}_{\text{sum}} := (\boldsymbol{\alpha}, \mathbf{0}, \mathbf{0}, \mathbf{0}) \otimes \boldsymbol{\pi},$$

subintensity matrix

$$\boldsymbol{S}_{\text{sum}} := \begin{pmatrix} \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I} \end{pmatrix} \otimes \boldsymbol{S} + \begin{pmatrix} \boldsymbol{P} & \boldsymbol{U}^* & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}_1 \otimes \boldsymbol{Q}_2 & \boldsymbol{q}_1 \otimes \boldsymbol{Q}_2 & \boldsymbol{Q}_1 \otimes \boldsymbol{q}_2 \\ \mathbf{0} & \mathbf{0} & \boldsymbol{Q}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{Q}_1 \end{pmatrix} \otimes (-\boldsymbol{S}\mathbf{1}\boldsymbol{\pi}),$$

and reward matrix

$$\boldsymbol{R}_{\text{sum}} := \begin{pmatrix} \mathbf{1}^\intercal \otimes \boldsymbol{r}_1 & \mathbf{1}^\intercal \otimes \boldsymbol{r}_1 & \mathbf{0} & \mathbf{1}^\intercal \otimes \boldsymbol{r}_1 \\ \mathbf{1}^\intercal \otimes \boldsymbol{r}_2 & \mathbf{1}^\intercal \otimes \boldsymbol{r}_2 & \mathbf{1}^\intercal \otimes \boldsymbol{r}_2 & \mathbf{0} \end{pmatrix}.$$

Here, for $k \in \{1, 2\}$, the quantity $\boldsymbol{r}_k$ represents the row vector $(r_{k1}, r_{k2}, \ldots, r_{k|\mathcal{C}|})$ such that

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \end{pmatrix}.$$

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

With the random vector (Y_1, Y_2) fully characterized as an MPH* distribution, the joint moments of the pair follow directly from Theorem 8.1.5 in Bladt and Nielsen (2017). For completeness, we explicitly state the formulas for the first- and second-order moments below, which are essential for calculating dependence measures like covariance and correlation.

Corollary 3.4. *Suppose that each bivariate random vector in the sequence* $\{(X_{1,\ell}, X_{2,\ell})\}_{\ell\in\mathbb{N}_+}$ *follows an* MPH* *distribution with parameters* $(\boldsymbol{\pi}, \boldsymbol{S}, \boldsymbol{R})$*, so that* (Y_1, Y_2) *defined in (3.1) is* MPH* *distributed with parameters* $(\boldsymbol{\pi}_{\text{sum}}, \boldsymbol{S}_{\text{sum}}, \boldsymbol{R}_{\text{sum}})$ *given in Proposition 3.3. For* $k \in \{1, 2\}$*, let* $\boldsymbol{D}_k = \text{diag}(\boldsymbol{r}_{k,\text{sum}})$ *be the diagonal matrix with the elements of* $\boldsymbol{r}_{k,\text{sum}}$ *in its diagonal, where* $\boldsymbol{r}_{k,\text{sum}}$ *is the k-th row of the reward matrix* $\boldsymbol{R}_{\text{sum}}$*. Then, for each* $k \in \{1, 2\}$*, the first moment of the compound sum is given by*

$$\mathbb{E}(Y_k) = \boldsymbol{\pi}_{\text{sum}}(-\boldsymbol{S}_{\text{sum}}^{-1})\boldsymbol{r}_{k,\text{sum}}^{\intercal},$$

and the second moment is

$$\mathbb{E}(Y_k^2) = 2\boldsymbol{\pi}_{\text{sum}}(-\boldsymbol{S}_{\text{sum}}^{-1})\boldsymbol{D}_k(-\boldsymbol{S}_{\text{sum}}^{-1})\boldsymbol{r}_{k,\text{sum}}^{\intercal}.$$

Furthermore, the second-order cross-moment is given by

$$\mathbb{E}(Y_1Y_2) = \boldsymbol{\pi}_{\text{sum}}(-\boldsymbol{S}_{\text{sum}}^{-1})\boldsymbol{D}_1(-\boldsymbol{S}_{\text{sum}}^{-1})\boldsymbol{r}_{2,\text{sum}}^{\intercal} + \boldsymbol{\pi}_{\text{sum}}(-\boldsymbol{S}_{\text{sum}}^{-1})\boldsymbol{D}_2(-\boldsymbol{S}_{\text{sum}}^{-1})\boldsymbol{r}_{1,\text{sum}}^{\intercal}.$$

4 Estimation for bivariate CDPH random variables

We shall perform parameter estimation given the data $\{(n_1^{(m)}, n_2^{(m)})\}_{m=1}^n$, where each $(n_1^{(m)}, n_2^{(m)})$ for $m = 1, 2, \ldots, n$ is an independent realization from our proposed bivariate CDPH distribution. For later use, we denote $(M_1^{(m)}, M_2^{(m)})$ as the joint process underlying the m-th CDPH random vector with realization $(n_1^{(m)}, n_2^{(m)})$. We first review the fully observed case and then proceed to provide maximum likelihood estimation through the use of the EM algorithm.

4.1 The fully observed case

Let the parameters be collected in $\boldsymbol{\Theta} := (\boldsymbol{\alpha}, \boldsymbol{P}, \boldsymbol{U}, \boldsymbol{Q}_1, \boldsymbol{Q}_2)$. The likelihood is given by

$$\mathcal{L}^O(\boldsymbol{\Theta} \mid \{(n_1^{(m)}, n_2^{(m)})\}_{m=1}^n) = \prod_{m=1}^n f_{\tau_1,\tau_2}(n_1^{(m)}, n_2^{(m)}; \boldsymbol{\Theta}),$$

where we place emphasis on the dependence of the pmf on $\boldsymbol{\Theta}$ by writing $f_{\tau_1,\tau_2}(\cdot,\cdot;\boldsymbol{\Theta})$ instead of $f_{\tau_1,\tau_2}(\cdot,\cdot)$ (see Proposition 2.1).

While direct optimization via gradient methods is possible and relatively quick for very small dimensions, the EM algorithm is more precise and efficient for moderate matrix dimensions $|\mathcal{E}|$ and $|\mathcal{S}|$. Thus, we introduce the following latent variables:

1. A_i: the number of times the processes $\{M_1^{(m)}\}_{m=1}^n$ start in state $i \in \mathcal{E}$.
2. N_{ij}^A: the total number of jumps of $\{M_1^{(m)}\}_{m=1}^n$ from state $i \in \mathcal{E}$ to $j \in \mathcal{E}$.
3. N_{ij}^T: the total number of jumps of $\{M_1^{(m)}\}_{m=1}^n$ from state $i \in \mathcal{E}$ to $j \in \mathcal{S}$.
4. $N_{ij}^{B,k}$: the total number of jumps of $\{M_k^{(m)}\}_{m=1}^n$ from state $i \in \mathcal{S}$ to $j \in \mathcal{S}$, where $k \in \{1, 2\}$.
5. $N_i^{B,k}$: the total number of jumps of $\{M_k^{(m)}\}_{m=1}^n$ from state $i \in \mathcal{S}$ to the absorbing state, where $k \in \{1, 2\}$.

Having specified the above sufficient statistics, we are able to reconstruct the paths of each of the Markov jump processes. Recall that α_i is the i-th element of the row vector $\boldsymbol{\alpha}$ for $i \in \mathcal{E}$ as well as the notation $\boldsymbol{P} = \{p_{ij}\}_{i,j\in\mathcal{E}}$, $\boldsymbol{Q}_k = \{q_{k,ij}\}_{i,j\in\mathcal{S}}$ and $\boldsymbol{U} = \{u_{ij}\}_{i\in\mathcal{E},j\in\mathcal{S}}$. For $k \in \{1, 2\}$, we further define $q_{k,i}$ to be the

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

i-th element of the column vector $\boldsymbol{q}_k$ for $i \in \mathcal{S}$. Then, we can write the complete likelihood as

$$\begin{aligned}&\mathcal{L}^C(\boldsymbol{\Theta} \mid \{M_1^{(m)}\}_{m=1}^n, \{M_2^{(m)}\}_{m=1}^n)\\&= \left(\prod_{i\in\mathcal{E}} \alpha_i^{A_i}\right)\left(\prod_{i,j\in\mathcal{E}} p_{ij}^{N_{ij}^A}\right)\left(\prod_{i\in\mathcal{E}}\prod_{j\in\mathcal{S}} u_{ij}^{N_{ij}^T}\right)\left(\prod_{k=1}^{2}\prod_{i,j\in\mathcal{S}} q_{k,ij}^{N_{ij}^{B,k}}\right)\left(\prod_{k=1}^{2}\prod_{i\in\mathcal{S}} q_{k,i}^{N_{i}^{B,k}}\right).\end{aligned}$$

The above specification is particularly convenient since it belongs to the exponential dispersion family of distributions, and thus has fully explicit maximum likelihood estimators (with a '^' placed on top of the corresponding parameter) given by

$$\hat{\alpha}_i = \frac{A_i}{n}, \quad \hat{p}_{ij} = \frac{N_{ij}^A}{\sum_{\ell\in\mathcal{E}} N_{i\ell}^A + \sum_{\ell\in\mathcal{S}} N_{i\ell}^T}, \quad \hat{u}_{ij} = \frac{N_{ij}^T}{\sum_{\ell\in\mathcal{E}} N_{i\ell}^A + \sum_{\ell\in\mathcal{S}} N_{i\ell}^T},$$
$$\hat{q}_{k,ij} = \frac{N_{ij}^{B,k}}{\sum_{\ell\in\mathcal{S}} N_{i\ell}^{B,k} + N_i^{B,k}}, \quad \hat{q}_{k,i} = \frac{N_i^{B,k}}{\sum_{\ell\in\mathcal{S}} N_{i\ell}^{B,k} + N_i^{B,k}}.$$

4.2 The EM algorithm

Since the full-trajectory data are not observed, we employ the EM algorithm to find the maximum likelihood estimate iteratively. This implies that at each iteration, the conditional expectations of the sufficient statistics, given the absorption times, are computed, corresponding to the E-step. Subsequently, such conditional expectations are replaced into $\mathcal{L}^C$ instead of the statistics themselves, and so upon maximization, we obtain updated parameters ($\boldsymbol{\Theta}$), commonly referred to as the M-step.

The conditional expectations required in the E-step are given as follows:

Lemma 4.1. *Let $\xi := \{(N_1^{(m)}, N_2^{(m)}) = (n_1^{(m)}, n_2^{(m)})\}_{m=1}^n$ be the observed data. The conditional expectations required in the E-step are given as follows:*

$$\mathbb{E}[A_i \mid \xi] = \sum_{m=1}^{n} \frac{\alpha_i f_{\tau_1,\tau_2|M_1,M_2}(n_1^{(m)}, n_2^{(m)}|i)}{f_{\tau_1,\tau_2}(n_1^{(m)}, n_2^{(m)})}.$$

$$\mathbb{E}[N_{ij}^A \mid \xi] = \sum_{m=1}^{n}\sum_{t\geq 1} \frac{(\boldsymbol{\alpha}\boldsymbol{P}^{t-1})_i\, p_{ij} f_{\tau_1,\tau_2|M_1,M_2}(n_1^{(m)} - t, n_2^{(m)} - t|j)}{f_{\tau_1,\tau_2}(n_1^{(m)}, n_2^{(m)})}.$$

$$\mathbb{E}[N_{ij}^T \mid \xi] = \sum_{m=1}^{n}\sum_{t\geq 1} \frac{(\boldsymbol{\alpha}\boldsymbol{P}^{t-1})_i\, u_{ij} (\boldsymbol{Q}_1^{n_1^{(m)}-t-1}\boldsymbol{q}_1)_j (\boldsymbol{Q}_2^{n_2^{(m)}-t-1}\boldsymbol{q}_2)_j}{f_{\tau_1,\tau_2}(n_1^{(m)}, n_2^{(m)})}.$$

$$\mathbb{E}[N_{ij}^{B,1} \mid \xi] = \sum_{m=1}^{n}\sum_{t\geq 2}\sum_{\ell=1}^{t-1}\sum_{r\in\mathcal{S}} \frac{(\boldsymbol{\alpha}\boldsymbol{P}^{\ell-1}\boldsymbol{U})_r (\boldsymbol{Q}_1^{t-\ell-1})_{ri}\, q_{1,ij} (\boldsymbol{Q}_1^{n_1^{(m)}-t-1}\boldsymbol{q}_1)_j (\boldsymbol{Q}_2^{n_2^{(m)}-\ell-1}\boldsymbol{q}_2)_r}{f_{\tau_1,\tau_2}(n_1^{(m)}, n_2^{(m)})}.$$

$$\mathbb{E}[N_{ij}^{B,2} \mid \xi] = \sum_{m=1}^{n}\sum_{t\geq 2}\sum_{\ell=1}^{t-1}\sum_{r\in\mathcal{S}} \frac{(\boldsymbol{\alpha}\boldsymbol{P}^{\ell-1}\boldsymbol{U})_r (\boldsymbol{Q}_2^{t-\ell-1})_{ri}\, q_{2,ij} (\boldsymbol{Q}_2^{n_2^{(m)}-t-1}\boldsymbol{q}_2)_j (\boldsymbol{Q}_1^{n_1^{(m)}-\ell-1}\boldsymbol{q}_1)_r}{f_{\tau_1,\tau_2}(n_1^{(m)}, n_2^{(m)})}.$$

$$\mathbb{E}[N_{i}^{B,1} \mid \xi] = \sum_{m=1}^{n}\sum_{t\geq 2}\sum_{\ell=1}^{t-1}\sum_{r\in\mathcal{S}} \frac{(\boldsymbol{\alpha}\boldsymbol{P}^{\ell-1}\boldsymbol{U})_r (\boldsymbol{Q}_1^{t-\ell-1})_{ri}\, q_{1,i} (\boldsymbol{Q}_2^{n_2^{(m)}-\ell-1}\boldsymbol{q}_2)_r}{f_{\tau_1,\tau_2}(n_1^{(m)}, n_2^{(m)})}.$$

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

$$\mathbb{E}[N_i^{B,2} \mid \xi] = \sum_{m=1}^{n} \sum_{t\geq 2} \sum_{\ell=1}^{t-1} \sum_{r\in\mathcal{S}} \frac{(\boldsymbol{\alpha}\boldsymbol{P}^{\ell-1}\boldsymbol{U})_r (\boldsymbol{Q}_2^{t-\ell-1})_{ri}\, q_{2,i} (\boldsymbol{Q}_1^{n_1^{(m)}-\ell-1}\boldsymbol{q}_1)_r}{f_{\tau_1,\tau_2}(n_1^{(m)}, n_2^{(m)})}.$$

Here f_{τ_1,τ_2} and $f_{\tau_1,\tau_2|M_1,M_2}$ are given by (2.2) and (2.3) with the help of (2.4) and (2.5).

5 Simulations and real data analysis

In this section, we investigate the flexibility and applicability of our proposed CDPH model on count data. Specifically, we conduct simulations involving bivariate Poisson and Poisson-Lindley distributions, followed by an application to real-world insurance claims data. Our focus is on assessing how well CDPH can capture or mimic these distributions. The values of $|\mathcal{E}|$ and $|\mathcal{S}|$ are chosen iteratively, starting with a simple model and gradually increasing complexity until further improvements in log-likelihood and visual fit become negligible.

5.1 Bivariate Poisson distribution with common shocks

We begin by demonstrating that our proposed CDPH model can effectively fit a bivariate Poisson distribution. Although the Poisson distribution itself is not a special case of the DPH distribution, we aim to show that the CDPH model provides a good approximation.

To define the bivariate Poisson distribution, we let (N_1^*, N_2^*) be a bivariate Poisson distributed pair such that $N_1^* := V_1 + Z$ and $N_2^* := V_2 + Z$, where V_1, V_2, and Z are independent Poisson random variables with parameters λ_{V_1}, λ_{V_2}, and λ_Z, respectively. Under this construction, the marginal distributions are again Poisson with intensities $\lambda_{N_1^*} := \lambda_{V_1} + \lambda_Z$ and $\lambda_{N_2^*} := \lambda_{V_2} + \lambda_Z$. The joint pmf of (N_1^*, N_2^*) is provided in Equation (2) of Holgate (1964).

To evaluate the effectiveness of our model, we simulate data from (N_1^*, N_2^*) and then fit $(N_1^* + 2, N_2^* + 2)$ using our CDPH approach.[1], The EM algorithm guarantees convergence to local minima, so trying different starting points is generally recommended. An interesting aspect of the study is to observe how the proposed model performs as the value of the common shock rate λ_Z increases, while keeping $\lambda_{N_1^*}$ and $\lambda_{N_2^*}$ fixed. Specifically, we fix $\lambda_{N_1^*} = 5$ and $\lambda_{N_2^*} = 4$ and consider three different values of λ_Z, namely 1, 2, and 3. The results for the case with $\lambda_Z = 1$ are presented in Figure 1, while the other two cases are collected in the Appendix. For each value of λ_Z, we explore three different model configurations with dimensions $(|\mathcal{E}|, |\mathcal{S}|) = (2, 1)$, $(|\mathcal{E}|, |\mathcal{S}|) = (3, 2)$, and $(|\mathcal{E}|, |\mathcal{S}|) = (4, 3)$. Since matrix representations are not unique and hence not identifiable, the use of Akaike information criterion (AIC) and Bayesian information criterion (BIC) is not appropriate in this context. Instead, we track the evolution of the likelihood as a function of the EM iteration step to assess model performance.

The simulations suggest that as the common shock component becomes more pronounced, our model captures it more effectively. That is, the fitted pmf of the common shock component retrieved by our CDPH model offers a more accurate representation of the empirical pmf of the actual common shock Z. Nevertheless, in all cases, the model provides a satisfactory fit, which is quantified through heatmap plots, showing the best-fitted pmf among the three considered models (which is the most complex one with dimensions $(|\mathcal{E}|, |\mathcal{S}|) = (4, 3)$), as well as the absolute difference between the empirical pmf and the best-fitted one.

5.2 Bivariate Poisson–Lindley distribution

In addition to the bivariate Poisson case, we now investigate the ability of our model to capture overdispersion in claim counts, where the variance is not necessarily equal to the mean. The Poisson–Lindley

[1] In this paper, the initial parameters for estimation are randomly generated: each matrix entry comprises an independent uniform random variable, suitably normalized to comply with the matrix constraints.

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

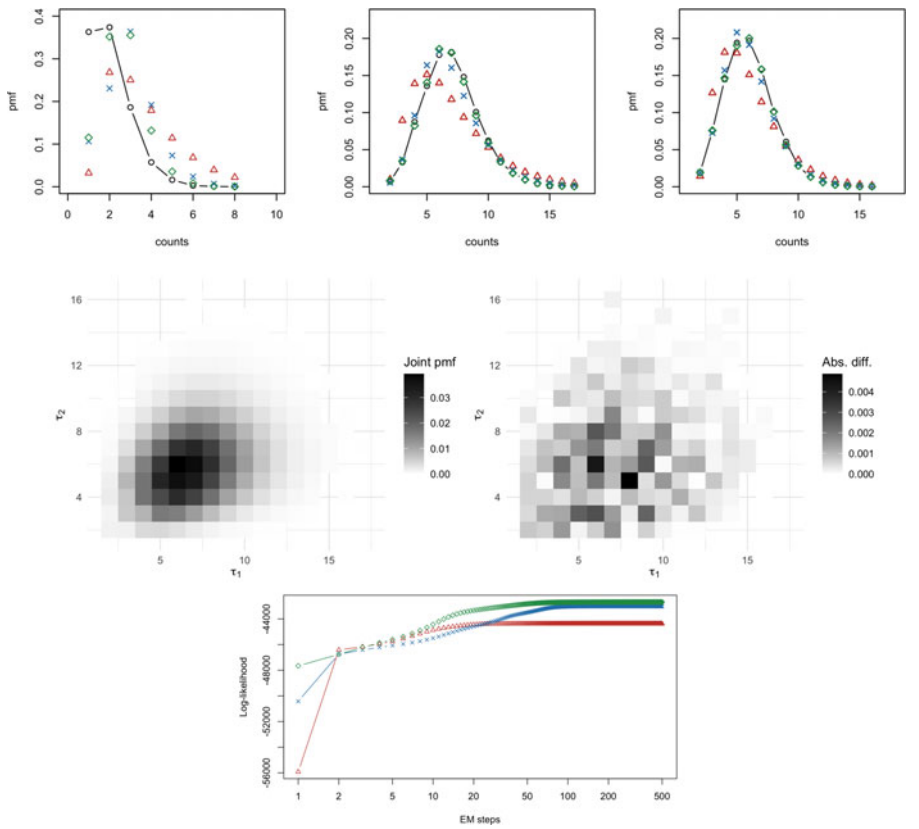


Figure 1. *Bivariate Poisson distribution with common shock intensity $\lambda_Z = 1$. In the top panels, we show the empirical pmf (black circles) versus the fitted pmf (by increasing matrix dimension: red triangles, blue crosses, and green diamonds), for the distribution of the common shock (left), and the two marginals (center and right). In the middle panel, we have two heatmaps, corresponding to the best-fitted bivariate pmf (left), and the absolute difference between the best fitted and empirical bivariate pmf (right). The bottom panel shows the evolution of the likelihood for the three models as the number of EM steps increases to 500.*

is a mixture distribution that serves as a natural candidate for this setting. Specifically, the bivariate Poisson–Lindley distribution is constructed by considering conditionally independent Poisson random variables, where the Poisson parameter is mixed over a Lindley distribution that depends on a parameter θ (see Definition 2 in Gómez-Déniz *et al*., 2012). The joint pmf in the bivariate case is also provided in Equation (17) of Gómez-Déniz *et al*. (2012).

To examine the performance of our model, we simulate (N_1^*, N_2^*) from this distribution and apply our CDPH model to the transformed data $(N_1^* + 2, N_2^* + 2)$. Concretely, we simulate a Lindley random variable L with $\theta = 2$ and subsequently construct N_1^* and N_2^* as independent Poisson with parameters $2L$ and $3L$, respectively. This is then repeated 10,000 times. We again consider three model configurations with dimensions $(|\mathcal{E}|, |\mathcal{S}|) = (2, 1)$, $(|\mathcal{E}|, |\mathcal{S}|) = (3, 2)$, and $(|\mathcal{E}|, |\mathcal{S}|) = (4, 3)$. The evolution of the likelihood across EM iterations is tracked, as in the previous study, to assess convergence and fitting accuracy. The results are provided in Figure 2, and they further illustrate the adaptability of our CDPH approach in modeling overdispersed count data. Unlike the bivariate Poisson setting, the Poisson–Lindley mixture does not explicitly possess a common shock component. Remarkably, however, our model is still able to effectively capture and mimic dependencies between the two components, demonstrating its practical

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

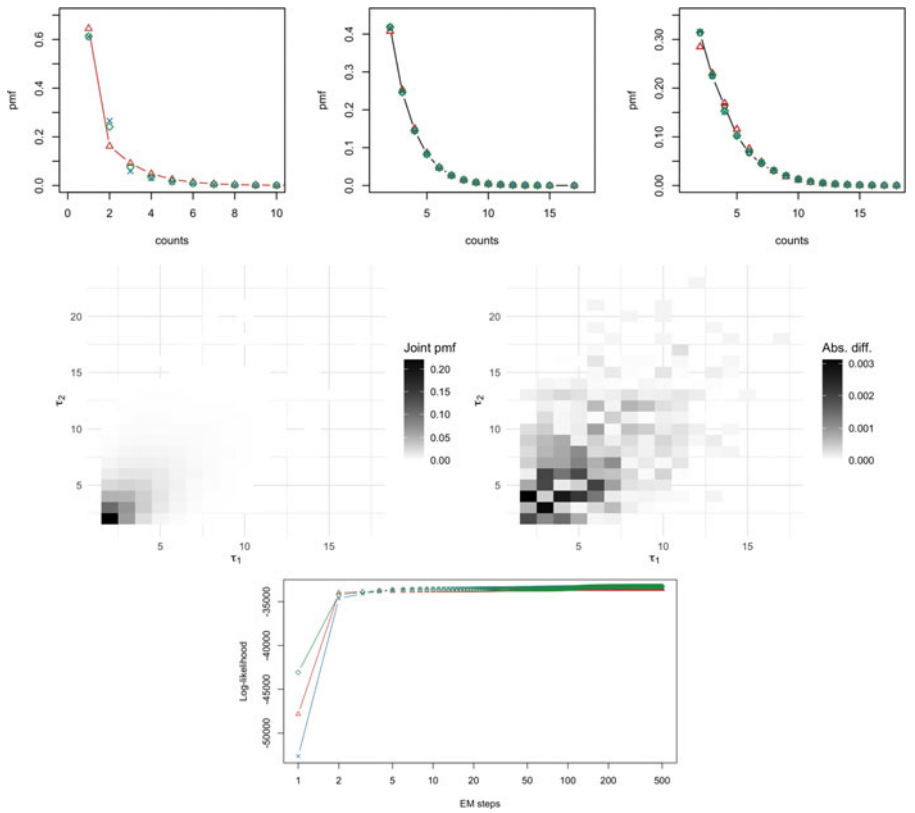


Figure 2. *Poisson–Lindley distribution with parameter $\theta = 2$. In the top panels, we show the empirical pmf (black circles) versus the fitted pmf (by increasing matrix dimension: red triangles, blue crosses, and green diamonds) for the two marginals (center and right), and the fitted values for a suggested common shock (left). In the middle panel, we have two heatmaps, corresponding to the best-fitted bivariate pmf (left), and the absolute difference between the best fitted and empirical bivariate pmf (right). The bottom panel shows the evolution of the likelihood for the three models as the number of EM steps increases to 500.*

modeling flexibility. In particular, our methodology is able to detect the presence of the inherent (or hidden) common shock component and recover its distribution, which can be useful for risk management purposes.

5.3 Application to insurance data

We now apply our model to real-world insurance data, previously analyzed in Vernic (2000). The dataset can be found in Table 1 of Vernic (2000) and was originally modeled using the bivariate generalized Poisson distribution (BGPD). It consists of claim frequencies from two related insurance categories, with a total of 708 claims.

We fit this dataset to our proposed CDPH distribution, applying the same shift transformation as before. The plots in Figure 3 illustrate the results for the same three different model configurations: $(|\mathcal{E}|, |\mathcal{S}|) = (2, 1)$, $(|\mathcal{E}|, |\mathcal{S}|) = (3, 2)$, and $(|\mathcal{E}|, |\mathcal{S}|) = (4, 3)$. As with the previous cases, we track the likelihood evolution over EM iterations and evaluate the fit using probability plots. The CDPH model in general provides a good fit to the insurance dataset, even in the case of low dimensions. It is interesting to note that a non-negligible common shock component is estimated quite robustly by the three models, which by the previous simulation studies suggests that there could either be a true underlying common

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

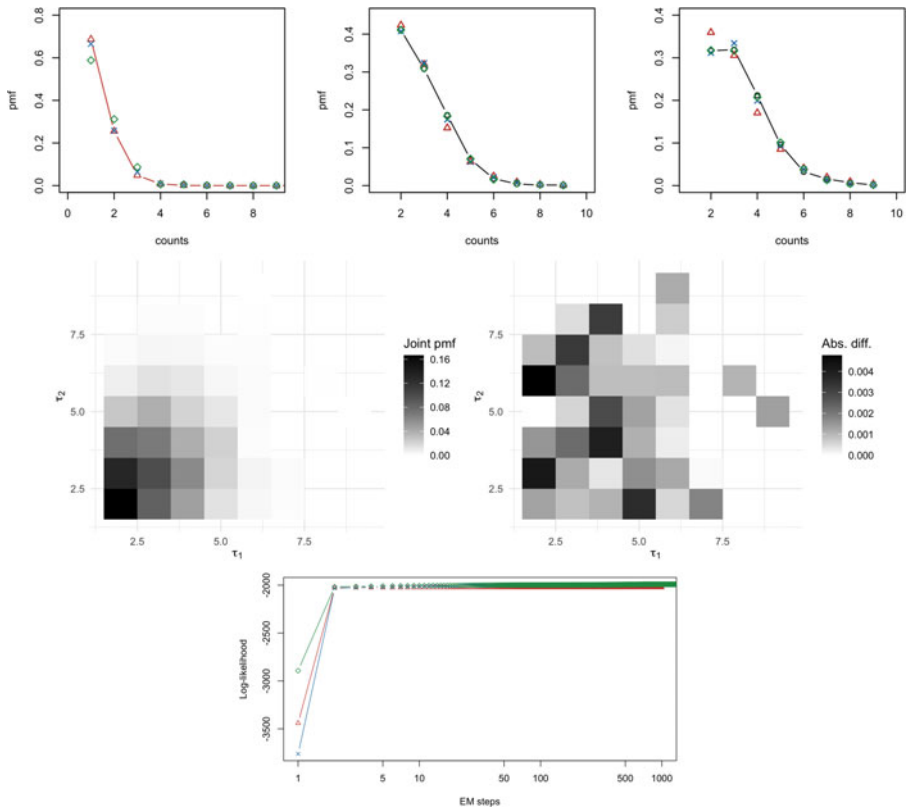


Figure 3. *Bivariate claim frequency insurance data. In the top panels, we show the fitted pmf (by increasing matrix dimension: red triangles, blue crosses, and green diamonds), for the distribution of the suggested common shock (left), and the two marginals (center and right). In the middle panel, we have two heatmaps, corresponding to the best-fitted bivariate pmf (left), and the absolute difference between the best fitted and empirical bivariate pmf (right). The bottom panel shows the evolution of the likelihood for the three models as the number of EM steps increases to 500.*

shock or that the bivariate distribution is well approximated by the CDPH class. Despite the structural differences between BGPD and CDPH, our approach is also able to capture dependencies effectively.

In fact, we also compare our proposed models directly to the BGPD models of Vernic (2000), using the χ^2 measure of goodness-of-fit as in that reference. The benchmark BGPD models yield χ^2 values of 25.94 and 23.61, respectively. Among our three matrix-based models, the first two, while simpler, produce higher χ^2 values of 58.13 and 29.08, indicating poorer fit. However, the third and most complex matrix model outperforms both BGPD benchmarks, with a lower χ^2 value of 18.93. We expect that models with even larger matrix dimensions (not shown here) could improve fit further, though this comes at the cost of increased parameter complexity. Addressing this trade-off – ideally through predictive performance rather than classical criteria like AIC or BIC, which are known to be inadequate for matrix distributions – is left for future research.

6 Conclusion

In this paper, we have introduced a novel class of bivariate common-shock discrete phase-type distributions, abbreviated as CDPH, to model dependencies in risk models induced by common shocks. The key idea is that the two random variables (τ_1, τ_2) of the CDPH distribution are defined as the termination times of two Markov chains that first evolve jointly (resembling common shocks influenced by shared

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

risk factors) and then proceed independently (following individual-specific dynamics). In addition to deriving explicit and tractable formulas for the joint pmf and the joint pgf, we have also established clear relationships between our CDPH class and other classes of bivariate DPH distributions proposed by Bladt and Yslas (2023b) and Navarro (2019). We have proved a number of closure properties of the CDPH class as well. In particular, $\min\{\tau_1, \tau_2\}$, $\max\{\tau_1, \tau_2\}$, and $\tau_1 + \tau_2$ are shown to be DPH, whereas mixtures and sums of independent and identically distributed copies of (τ_1, τ_2) belong to the CDPH class. Further properties of compound sums with CDPH claim counts are also discussed, and an EM algorithm for the CDPH class is developed for parameter estimation and subsequently implemented in our numerical studies, which demonstrate that the distribution of the inherent common shock component can be estimated from bivariate datasets.

Our proposed class of bivariate CDPH distributions opens up new research directions for modeling dependencies in multivariate risks, particularly in situations where common shocks play a significant role. First, one obvious direction is to further generalize the model to the multivariate case. In this paper, we have chosen a bivariate setting to allow for a clearer exposition of the core ideas so as to articulate our new construction and to thoroughly investigate its fundamental properties without the overwhelming notational and structural complexities of the multivariate case. Indeed, a multivariate extension in our context may not be straightforward, as it would require modeling multiple common shocks that can possibly contribute to different subsets of the random variables contained in the multivariate random vector. To ensure that the multivariate random vector admits an explicit joint pmf that facilitates both computations and the development of EM fitting algorithm, it is likely necessary to impose a block feed-forward type of structure on the probability matrices (resembling the block upper diagonal matrix in (2.1)). Nevertheless, a multivariate extension will undoubtedly involve a large number of parameters and handling of matrices with higher dimensions. An important research direction will be to study how the construction and the resulting computations can be done efficiently. Second, our proposed CDPH distributions may be potentially applied to credit risk modeling, particularly in describing the dependent default times of multiple entities (see e.g. Bielecki and Rutkowski, 2004 for a review of credit risk). Since default times are predominantly modeled as continuous random variables, a continuous analog of our discrete setting would be required. In credit risk, one is often interested in the pricing of contingent cashflow payable, for example, (1) at the i-th default time among n entities if it has occurred before a fixed time T and (2) at the fixed time T if the i-th default has not occurred. Pricing in this context will also require the specification of (1) a stochastic interest rate model for discounting payments; (2) the flow of information over time (e.g. credit ratings); and (3) a risk-neutral measure. This will be a challenging research problem but worthy of investigation. Finally, future research could explore the incorporation of covariate information or consider the tail behavior through the introduction of inhomogeneity functions. We leave all these as open questions.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/asb.2025.10062.

Acknowledgement. The authors would like to thank the editor and the reviewers for their helpful comments and suggestions, which have greatly improved the initially submitted manuscript. MB would like to acknowledge the financial support from the Swiss National Science Foundation Project 200021_191984, as well as from the Carlsberg Foundation, grant CF23-1096. EC and JKW acknowledge the support from the Australian Research Council's Discovery Project DP200100615.

References

Abramowitz, M. and Stegun, I.A. (1972) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. *Applied Mathematics Series*, Vol. **55**. Dover Publications.

Ahn, S., Kim, J.H.T. and Ramaswami, V. (2012) A new class of models for heavy tailed distributions in finance and insurance risk. *Insurance: Mathematics and Economics*, **51**(1), 43–52.

Albrecher, H., Bladt, M. and Müller, A.J. (2023) Joint lifetime modeling with matrix distributions. *Dependence Modeling*, **11**(1), 20220153.

Albrecher, H., Cheung, E.C.K., Liu, H. and Woo, J.-K. (2022) A bivariate Laguerre expansions approach for joint ruin probabilities in a two-dimensional insurance risk process. *Insurance: Mathematics and Economics*, **103**, 96–118.

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

Asmussen, S. and Albrecher, H. (2010) *Ruin Probabilities*, 2nd edn. World Scientific.
Assaf, D., Langberg, N.A., Savits, T.H. and Shaked, M. (1984) Multivariate phase-type distributions. *Operations Research*, **32**(3), 688–702.
Badescu, A.L. and Landriault, D. (2009) Applications of fluid flow matrix analytic methods in ruin theory - a review. *RACSAM - Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, **103**(2), 353–372.
Badila, E.S., Boxma, O.J. and Resing, J.A.C. (2015) Two parallel insurance lines with simultaneous arrivals and risks correlated with inter-arrival times. *Insurance: Mathematics and Economics*, **61**, 48–61.
Bielecki, T.R. and Rutkowski, M. (2004) *Credit Risk: Modeling, Valuation and Hedging*. Springer.
Bladt, M. (2023) A tractable class of multivariate phase-type distributions for loss modeling. *North American Actuarial Journal*, **27**(4), 710–730.
Bladt, M. and Nielsen, B.F. (2010) Multivariate matrix-exponential distributions. *Stochastic Models*, **26**(1), 1–26.
Bladt, M. and Nielsen, B.F. (2017) *Matrix-Exponential Distributions in Applied Probability*. Springer.
Bladt, M. and Yslas, J. (2023a) Phase-type mixture-of-experts regression for loss severities. *Scandinavian Actuarial Journal*, **2023**(4), 303–329.
Bladt, M. and Yslas, J. (2023b) Robust claim frequency modeling through phase-type mixture-of-experts regression. *Insurance: Mathematics and Economics*, **111**, 1–22.
Bolancé, C. and Vernic, R. (2019) Multivariate count data generalized linear models: Three approaches based on the Sarmanov distribution. *Insurance: Mathematics and Economics*, **85**, 89–103.
Cheung, E.C.K. and Landriault, D. (2010) A generalized penalty function with the maximum surplus prior to ruin in a MAP risk model. *Insurance: Mathematics and Economics*, **46**(1), 127–134.
Cheung, E.C.K., Peralta, O. and Woo, J.-K. (2022) Multivariate matrix-exponential affine mixtures and their applications in risk theory. *Insurance: Mathematics and Economics*, **106**, 364–389.
Cossette, H., Mailhot, M., Marceau, E. and Mesfioui, M. (2016) Vector-valued Tail Value-at-Risk and capital allocation. *Methodology and Computing in Applied Probability*, **18**(3), 653–674.
Drekic, S., Dickson, D.C.M., Stanford, D.A. and Willmot, G.E. (2004) On the distribution of the deficit at ruin when claims are phase-type. *Scandinavian Actuarial Journal*, **2004**(2), 105–120.
Eisele, K.T. (2006) Recursions for compound phase distributions. *Insurance: Mathematics and Economics*, **38**(1), 149–156.
Fackrell, M. (2009) Modelling healthcare systems with phase-type distributions. *Health Care Management Science*, **12**(1), 11–26.
Frees, E.W. and Valdez, E.A. (1998) Understanding relationships using copulas. *North American Actuarial Journal*, **2**(1), 1–25.
Frostig, E., Pitts, S.M. and Politis, K. (2012) The time to ruin and the number of claims until ruin for phase-type claims. *Insurance: Mathematics and Economics*, **51**(1), 19–25.
Fung, T.C., Badescu, A.L. and Lin, X.S. (2019) A class of mixture of experts models for general insurance: Application to correlated claim frequencies. *ASTIN Bulletin*, **49**(3), 647–688.
Gómez-Déniz, P., Sarabia, J.M. and Balakrishnan, N. (2012) A multivariate discrete Poisson-Lindley distribution: Extensions and actuarial applications. *ASTIN Bulletin*, **42**(2), 655–678.
Hassan Zadeh, A. and Stanford, D. (2016) Bayesian and Bühlmann credibility for phase-type distributions with a univariate risk parameter. *Scandinavian Actuarial Journal*, **2016**(4), 338–355.
He, Q.-M. and Ren, J. (2016a) Analysis of a multivariate claim process. *Methodology and Computing in Applied Probability*, **18**, 257–273.
He, Q.-M. and Ren, J. (2016b) Parameter estimation of discrete multivariate phase-type distributions. *Methodology and Computing in Applied Probability*, **18**(3), 629–651.
Holgate, P. (1964) Estimation for the bivariate Poisson distribution. *Biometrika*, **51**(1/2), 241–245.
Jensen, A. (1954) *A distribution model applicable to economics*. Thesis Dissertation, Munksgaard, Copenhagen.
Kulkarni, V.G. (1989) A new class of multivariate phase type distributions. *Operations Research*, **37**(1), 151–158.
Landsman, Z., Makov, U. and Shushi, T. (2016) Multivariate tail conditional expectation for elliptical distributions. *Insurance: Mathematics and Economics*, **70**, 216–223.
Lee, S. and Lin, X.S. (2012) Modeling dependent risks with multivariate Erlang mixtures. *ASTIN Bulletin*, **42**(1), 153–180.
Lin, X.S. and Liu, X. (2007) Markov aging process and phase-type law of mortality. *North American Actuarial Journal*, **11**(4), 92–109.
Navarro, A.C. (2019) *Order statistics and multivariate discrete phase-type distributions*. *Thesis Dissertation*, Technical University of Denmark.
Pechon, F., Trufin, J. and Denuit, M. (2018) Multivariate modelling of household claim frequencies in motor third-party liability insurance. *ASTIN Bulletin*, **48**(3), 969–993.
Ren, J. (2010) Recursive formulas for compound phase distributions–univariate and bivariate cases. *ASTIN Bulletin*, **40**(2), 615–629.
Shi, P. and Valdez, E.A. (2014) Multivariate negative binomial models for insurance claim counts. *Insurance: Mathematics and Economics*, **55**, 18–29.
Vernic, R. (2000) A multivariate generalization of the generalized Poisson distribution. *ASTIN Bulletin*, **30**(1), 57–67.

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

Wang, Y.F., Garrido, J. and Léveillé, G. (2018) The distribution of discounted compound PH-renewal processes. *Methodology and Computing in Applied Probability*, **20**(1), 69–96.
Willmot, G.E. and Woo, J.-K. (2015) On some properties of a class of multivariate Erlang mixtures with insurance applications. *ASTIN Bulletin*, **45**(1), 151–173.
Wu, X. and Li, S. (2010) Matrix-form recursions for a family of compound distributions. *ASTIN Bulletin*, **40**(1), 351–368.

Appendix

Proofs

This section provides the detailed proofs for selected propositions, corollaries, and lemmas from the main text. The numbering of these results and their corresponding equations follows that of the main paper for ease of reference.

Proof of Proposition 2.1

Proof. Equations (2.2) and (2.3) follow by straightforward application of the law of total probability. Next,

$$\begin{aligned}
&\mathbb{P}(\tau_{1,2}=m,\tau_1-\tau_{1,2}=z_1,\tau_2-\tau_{1,2}=z_2)\\
&\quad=\sum_{j\in\mathcal{S}}\mathbb{P}(\tau_{1,2}=m,M_1(m)=M_2(m)=j)\\
&\qquad\times\mathbb{P}(\tau_1-\tau_{1,2}=z_1,\tau_2-\tau_{1,2}=z_2\mid\tau_{1,2}=m,M_1(m)=M_2(m)=j)\\
&\quad=\sum_{j\in\mathcal{S}}\mathbb{P}(\tau_{1,2}=m,M_1(m)=M_2(m)=j)\\
&\qquad\times\mathbb{P}(\tau_1-\tau_{1,2}=z_1\mid\tau_{1,2}=m,M_1(m)=j)\mathbb{P}(\tau_2-\tau_{1,2}=z_2\mid\tau_{1,2}=m,M_2(m)=j)\\
&\quad=\sum_{j\in\mathcal{S}}(\boldsymbol{\alpha}\boldsymbol{P}^{m-1}\boldsymbol{U})_j(\boldsymbol{Q}_1^{z_1-1}\boldsymbol{q}_1)_j(\boldsymbol{Q}_2^{z_2-1}\boldsymbol{q}_2)_j\\
&\quad=\boldsymbol{\alpha}\boldsymbol{P}^{m-1}\boldsymbol{U}\left((\boldsymbol{Q}_1^{z_1-1}\boldsymbol{q}_1)\odot(\boldsymbol{Q}_2^{z_2-1}\boldsymbol{q}_2)\right),
\end{aligned}$$

which establishes (2.4). Finally, (2.5) follows by replacing the initial distribution $\boldsymbol{\alpha}$ with $\boldsymbol{e}_i^{\mathsf{T}}$. □

Proof of Corollary 2.2

Proof. As ψ is the joint pmf of $(\tau_{1,2},\tau_1-\tau_{1,2},\tau_2-\tau_{1,2})$, we have

$$\begin{aligned}
\mathbb{E}\left(\zeta_0^{\tau_{1,2}}\zeta_1^{\tau_1-\tau_{1,2}}\zeta_2^{\tau_2-\tau_{1,2}}\right)&=\sum_{z_0,z_1,z_2\geq 1}\psi(z_0,z_1,z_2)\zeta_0^{z_0}\zeta_1^{z_1}\zeta_2^{z_2}\\
&=\sum_{z_0,z_1,z_2\geq 1}\sum_{j\in\mathcal{S}}(\boldsymbol{\alpha}\boldsymbol{P}^{z_0-1}\boldsymbol{U})_j(\boldsymbol{Q}_1^{z_1-1}\boldsymbol{q}_1)_j(\boldsymbol{Q}_2^{z_2-1}\boldsymbol{q}_2)_j\zeta_0^{z_0}\zeta_1^{z_1}\zeta_2^{z_2}\\
&=\zeta_0\zeta_1\zeta_2\sum_{j\in\mathcal{S}}\sum_{z_0,z_1,z_2\geq 1}(\boldsymbol{\alpha}(\zeta_0\boldsymbol{P})^{z_0-1}\boldsymbol{U})_j((\zeta_1\boldsymbol{Q}_1)^{z_1-1}\boldsymbol{q}_1)_j((\zeta_2\boldsymbol{Q}_2)^{z_2-1}\boldsymbol{q}_2)_j\\
&=\zeta_0\zeta_1\zeta_2\sum_{j\in\mathcal{S}}(\boldsymbol{\alpha}(\boldsymbol{I}-\zeta_0\boldsymbol{P})^{-1}\boldsymbol{U})_j((\boldsymbol{I}-\zeta_1\boldsymbol{Q}_1)^{-1}\boldsymbol{q}_1)_j((\boldsymbol{I}-\zeta_2\boldsymbol{Q}_2)^{-1}\boldsymbol{q}_2)_j,
\end{aligned}$$

which yields (2.6). Substituting $\zeta_0=\zeta_1\zeta_2$ into (2.6) gives (2.7). □

Proof of Corollary 2.3

Proof. It is standard that the (n_0,n_1,n_2)-th falling factorial moment is obtained from the joint pgf (2.6) by taking n_i derivatives with respect to ζ_i for all $i=0,1,2$ and then evaluating at $\zeta_0=\zeta_1=\zeta_2=1$

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

so that

$$\mathbb{E}\big(\tau_{1,2}^{[n_0]}(\tau_1-\tau_{1,2})^{[n_1]}(\tau_2-\tau_{1,2})^{[n_2]}\big)=\frac{\partial^{n_0+n_1+n_2}}{\partial\zeta_0^{n_0}\partial\zeta_1^{n_1}\partial\zeta_2^{n_2}}\mathbb{E}(\zeta_0^{\tau_{1,2}}\zeta_1^{\tau_1-\tau_{1,2}}\zeta_2^{\tau_2-\tau_{1,2}})\Big|_{\zeta_0=1,\zeta_1=1,\zeta_2=1}.$$

Exploiting the linearity of differentiation yields (2.8) as long as one can prove the formula

$$\frac{\partial^r}{\partial\zeta_0^r}\boldsymbol{\zeta}_0(\boldsymbol{I}-\boldsymbol{T}\zeta_0)^{-1}=r!\boldsymbol{T}^{r-1}(\boldsymbol{I}-\boldsymbol{T}\zeta_0)^{-r-1},\qquad r\in\mathbb{N}_+,$$

where $\boldsymbol{T}$ can be $\boldsymbol{P}$, $\boldsymbol{Q}_1$ or $\boldsymbol{Q}_2$. It can be seen from the proof of Theorem 1.2.69 of Bladt and Nielsen (2017) that the above equation is indeed valid, concluding our proof. □

Proof of Proposition 2.7

Proof. With the representation (2.14) for the random variable $\tau_1+\tau_2$, it follows from Theorem 5.2 in Navarro (2019) that the distribution of $\tau_1+\tau_2$ is a mixture of a probability mass at zero and a DPH distribution. However, both τ_1 and τ_2 are at least one from their definitions, implying that the aforementioned probability mass at zero must be non-existent and hence $\tau_1+\tau_2$ follows a DPH distribution. The construction of the parameters of such a DPH distribution can be found in Chapter 5.1.1.2 in Navarro (2019) and is omitted for brevity. □

Proof of Proposition 2.8

Proof. Indeed, the initial distribution $\boldsymbol{\alpha}_{\text{mix}}$ is simply a weighted combination of the initial probabilities $\{\boldsymbol{\alpha}(i)\}_{i=1}^m$ of the component CDPH distributions, with weight w_i assigned to $\boldsymbol{\alpha}_i$. The transition matrix $\boldsymbol{P}_{\text{mix}}$ is a block-diagonal matrix, where each block corresponds to the transition matrix $\boldsymbol{P}(i)$ of the i-th component. Similarly, $\boldsymbol{U}_{\text{mix}}$, $\boldsymbol{Q}_{1,\text{mix}}$, and $\boldsymbol{Q}_{2,\text{mix}}$ are block-diagonal matrices, where each block corresponds to $\boldsymbol{U}(i)$, $\boldsymbol{Q}_1(i)$, and $\boldsymbol{Q}_2(i)$, respectively. Such block-diagonal structure ensures that there is no interaction among the components (i.e. $(\tau_1(i),\tau_2(i))$ across different i's) once a specific component is chosen according to the mixing weights $\{w_i\}_{i=1}^m$. This construction guarantees that the resulting parameters describe a valid CDPH distribution for $(\tau_{1,\text{mix}},\tau_{2,\text{mix}})$. □

Proof of Proposition 2.9

Proof. In order to explain the key ideas behind the proof, for $i,k\in\{1,2\}$ it will be helpful to write $\tau_k(i)=\tau_{1,2}(i)+(\tau_k(i)-\tau_{1,2}(i))$. Here the random variable $\tau_{1,2}(i)$ is the common shock component for the pair $(\tau_1(i),\tau_2(i))$. We can thus decompose each coordinate of the random vector $(\tau_{1,\text{sum}},\tau_{2,\text{sum}})=(\tau_1(1)+\tau_1(2),\tau_2(1)+\tau_2(2))$ as

$$\begin{aligned}\tau_{1,\text{sum}}&=(\tau_{1,2}(1)+\tau_{1,2}(2))+(\tau_1(1)-\tau_{1,2}(1))+(\tau_1(2)-\tau_{1,2}(2)),\\ \tau_{2,\text{sum}}&=(\tau_{1,2}(1)+\tau_{1,2}(2))+(\tau_2(1)-\tau_{1,2}(1))+(\tau_2(2)-\tau_{1,2}(2)).\end{aligned}$$

To show that $(\tau_{1,\text{sum}},\tau_{2,\text{sum}})$ follows a CDPH distribution, we note from the above representation that $\tau_{1,2}(1)+\tau_{1,2}(2)$ can be regarded as the common shock component from which we can construct $\boldsymbol{P}_{\text{sum}}$. After $\tau_{1,2}(1)$ and $\tau_{1,2}(2)$ are realized and given the states in $\mathcal{S}(1)$ and $\mathcal{S}(2)$ entered upon exiting $\mathcal{E}(1)$ and $\mathcal{E}(2)$, the variables $(\tau_1(1)-\tau_{1,2}(1))+(\tau_1(2)-\tau_{1,2}(2))$ and $(\tau_2(1)-\tau_{1,2}(1))+(\tau_2(2)-\tau_{1,2}(2))$ are conditionally independent: such observation can be used to construct $\boldsymbol{Q}_{1,\text{sum}}$ and $\boldsymbol{Q}_{2,\text{sum}}$. The construction of the matrix parameters of the CDPH representation of $(\tau_{1,\text{sum}},\tau_{2,\text{sum}})$ is described as follows:

- We start by considering the part $\tau_{1,2}(1)$ that belongs to the common shock $\tau_{1,2}(1)+\tau_{1,2}(2)$. The process underlying $(\tau_{1,\text{sum}},\tau_{2,\text{sum}})$ starts in $\mathcal{E}(1)$ with the initial probability $\boldsymbol{\alpha}(1)$ (which explains (2.15)) and then transitions within $\mathcal{E}(1)$ according to $\boldsymbol{P}(1)$. Upon exiting $\mathcal{E}(1)$, the variable $\tau_{1,2}(1)$ is realized and the process underlying $(\tau_1(1),\tau_2(1))$ enters $\mathcal{S}(1)$ according to $\boldsymbol{U}(1)$. While one

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

can start counting the remaining common shock component $\tau_{1,2}(2)$ using the initial probability $\boldsymbol{\alpha}(2)$ and transition probability matrix $\boldsymbol{P}(2)$ within $\mathcal{E}(2)$, it is important to also keep track of the state in $\mathcal{S}(1)$ because this will have an impact on the distributions of $\tau_1(1) - \tau_{1,2}(1)$ and $\tau_2(1) - \tau_{1,2}(1)$. These explain the block matrix representation of $\boldsymbol{P}_{\text{sum}}$ in (2.16). The submatrix $\boldsymbol{I} \otimes \boldsymbol{U}(2)$ of $\boldsymbol{U}_{\text{sum}}$ in (2.16) subsequently takes care of the termination of $\tau_{1,2}(2)$ as the process exits $\mathcal{S}(1) \times \mathcal{E}(2)$ and enters $\mathcal{S}(1) \times \mathcal{S}(2)$.

- Once the underlying process of $(\tau_{1,\text{sum}}, \tau_{2,\text{sum}})$ has entered $\mathcal{S}(1) \times \mathcal{S}(2)$, the clocks for $\tau_1(1) - \tau_{1,2}(1)$ and $\tau_2(1) - \tau_{1,2}(1)$ start ticking independently within the set $\mathcal{S}(1)$ according to the transition matrices $\boldsymbol{Q}_1(1)$ and $\boldsymbol{Q}_2(1)$, respectively, while keeping the state in $\mathcal{S}(2)$ fixed. When $\tau_1(1) - \tau_{1,2}(1)$ (resp. $\tau_2(1) - \tau_{1,2}(1)$) terminates according to the probability vector $\boldsymbol{q}_1(1)$ (resp. $\boldsymbol{q}_2(1)$), one no longer needs to keep track of the state space $\mathcal{S}(1)$, and then we move on to consider $\tau_1(2) - \tau_{1,2}(2)$ (resp. $\tau_2(2) - \tau_{1,2}(2)$) as the process evolves within $\mathcal{S}(2)$ according to $\boldsymbol{Q}_1(2)$ (resp. $\boldsymbol{Q}_2(2)$) until it gets absorbed. Such arguments lead to the specification of the matrices $\boldsymbol{Q}_{1,\text{sum}}$ and $\boldsymbol{Q}_{2,\text{sum}}$ in (2.17).

The above construction ensures that $(\tau_{1,\text{sum}}, \tau_{2,\text{sum}})$ is CDPH distributed, and the proof is complete. □

Proof of Proposition 3.1

Proof. By conditioning on (τ_1, τ_2), we have

$$\begin{aligned}
\mathcal{L}_{Y_1,Y_2}(\theta_1, \theta_2) &= \mathbb{E}\left(\exp\left(-\theta_1 \sum_{\ell=1}^{\tau_1} X_{1,\ell} - \theta_2 \sum_{\ell=1}^{\tau_2} X_{2,\ell}\right)\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\exp\left(-\theta_1 \sum_{\ell=1}^{\tau_1} X_{1,\ell} - \theta_2 \sum_{\ell=1}^{\tau_2} X_{2,\ell}\right)\middle| \tau_1, \tau_2\right)\right) \\
&= \mathbb{E}\left(\left[\prod_{\ell=1}^{\tau_1} \mathbb{E}\left(e^{-\theta_1 X_{1,\ell}}\right)\right]\left[\prod_{\ell=1}^{\tau_2} \mathbb{E}\left(e^{-\theta_2 X_{2,\ell}}\right)\right]\right) \\
&= \mathbb{E}\left(a_{1,\theta_1}^{\tau_1}\, a_{2,\theta_2}^{\tau_2}\right).
\end{aligned}$$

The above expression is simply the joint pgf in (2.7) evaluated at the arguments $(\zeta_1, \zeta_2) = (a_{1,\theta_1}, a_{2,\theta_2})$. Since a_{1,θ_1} and a_{2,θ_2} are the Laplace transforms of positive random variables, one must have $a_{1,\theta_1}, a_{2,\theta_2} \in [0, 1]$ for $\theta_1, \theta_2 \geq 0$. This yields the desired result. □

Proof of Proposition 3.3

Proof. We follow a probabilistic approach similar to that in Theorem 3.1.28 of Bladt and Nielsen (2017). The key idea is to construct a continuous-time Markov jump process whose accumulated rewards correspond to the random variables Y_1 and Y_2.

Recall γ is the discrete-time Markov chain that leads to the MDPH* representation (2.11). That is, γ is a Markov chain on the state space $\mathcal{E} \cup (\mathcal{S} \times \mathcal{S}) \cup (\{\partial_1\} \times \mathcal{S}) \cup (\mathcal{S} \times \{\partial_2\})$. For each $\ell \in \mathbb{N}_+$, suppose that J_ℓ is the process underlying the MPH* pair $(X_{1,\ell}, X_{2,\ell})$. Then, we concatenate independent realizations of J_ℓ's to match the number of steps of γ prior to termination. Thus, we construct a process that keeps track of the state of γ while simultaneously evolving within $\mathcal{C}$. After starting $\gamma \times J_1$ according to the initial distribution $(\boldsymbol{\alpha}, \mathbf{0}, \mathbf{0}, \mathbf{0}) \otimes \boldsymbol{\pi}$, the constructed process evolves as follows:

- While in $\mathcal{E} \times \mathcal{C}$ (representing the common-shock phase), γ evolves according to the transition matrix $\boldsymbol{P}$ and J_ℓ evolves independently according to the subintensity matrix $\boldsymbol{S}$. In this phase, both Y_1 and Y_2 accumulate rewards, reflecting that both components are active.
- When γ leaves $\mathcal{E}$, it transitions to $\mathcal{S} \times \mathcal{S}$ with transition probabilities given by $\boldsymbol{U}^*$, and the constructed process moves into $(\mathcal{S} \times \mathcal{S}) \times \mathcal{C}$ representing the phase after the common shocks.

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

Here, γ evolves according to the Kronecker product $\boldsymbol{Q}_1 \otimes \boldsymbol{Q}_2$, while J_ℓ continues to evolve according to $\boldsymbol{S}$. Both Y_1 and Y_2 continue to accumulate rewards.

- If one of the components reaches absorption (i.e., transitions to ∂_1 or ∂_2), the process J_ℓ moves to $\{\partial_1\} \times \mathcal{S}$ or $\mathcal{S} \times \{\partial_2\}$, respectively, representing single-coordinate continuation in $(\{\partial_1\} \times \mathcal{S}) \times \mathcal{C}$ or $(\mathcal{S} \times \{\partial_2\}) \times \mathcal{C}$ for the constructed process. In $(\{\partial_1\} \times \mathcal{S}) \times \mathcal{C}$, only Y_2 continues to accumulate rewards, while the reward accumulation for Y_1 ceases. Conversely, in $(\mathcal{S} \times \{\partial_2\}) \times \mathcal{C}$, only Y_1 continues to accumulate rewards.
- The process terminates when both components have reached absorption.

By constructing the Markov jump process as described and defining the rewards accordingly, we ensure that the accumulated rewards Y_1 and Y_2 correspond to the total occupation times of their respective components before absorption. This setup mirrors the dynamics of the original MPH* distribution for the pair (Y_1, Y_2), thus completing the proof. □

Proof of Lemma 4.1

Proof. The key in all expressions is to apply the disintegration formula by conditioning on the common shock. All the remaining calculations then follow along similar lines as in the univariate case, since the paths evolve independently thereafter. □

Further simulation analysis

Further examples of bivariate Poisson distribution with common shocks

This section presents additional numerical results to complement the simulation studies in Section 5 of the main paper. Specifically, we provide the detailed fitting results for the bivariate Poisson distribution with common shock intensities of $\lambda_Z = 2$ (Figure 4) and $\lambda_Z = 3$ (Figure 5). These examples further illustrate the performance and flexibility of the CDPH model, particularly its ability to capture dependence as the common shock component becomes more pronounced.

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

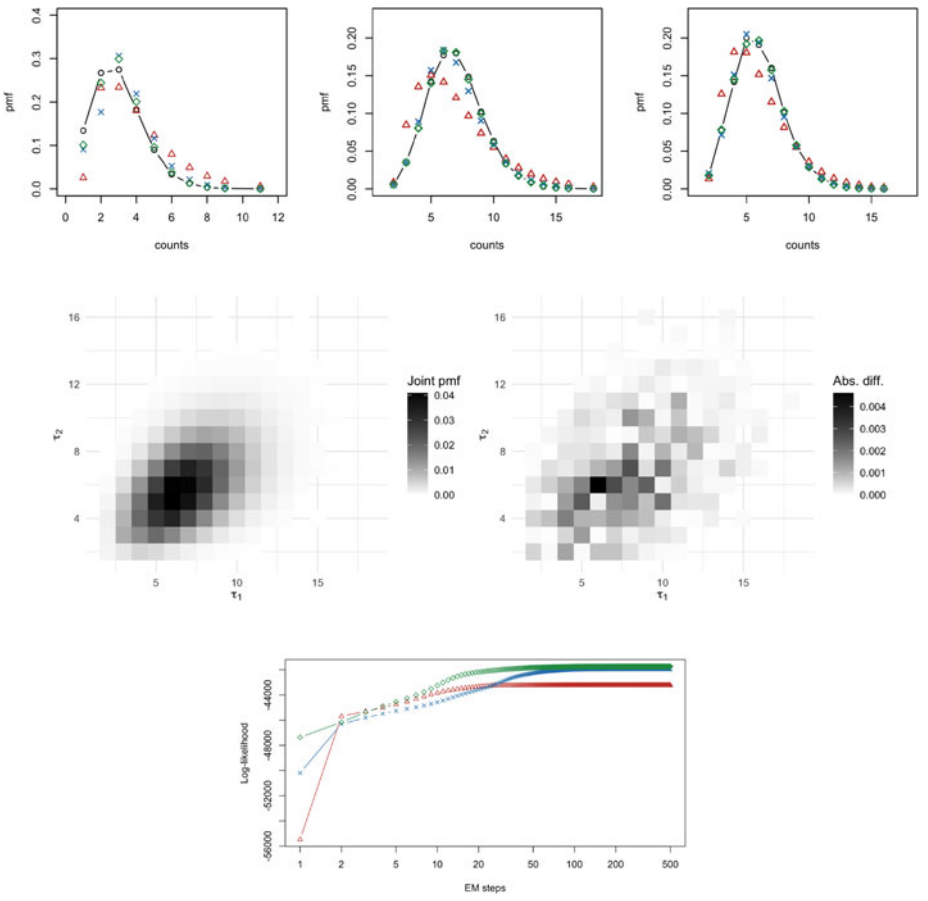


Figure 4. *Bivariate Poisson distribution with common shock intensity $\lambda_Z = 2$. In the top panels we show the empirical pmf (black circles) versus the fitted pmf (by increasing matrix dimension: red triangles, blue crosses, green diamonds), for the distribution of the common shock (left), and the two marginals (center and right). In the middle panel we have two heatmaps, corresponding to the best fitted bivariate pmf (left), and the absolute difference between the best fitted and empirical bivariate pmf (right). The bottom panel shows the evolution of the likelihood for the three models as the number of EM steps increases to 500.*

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press

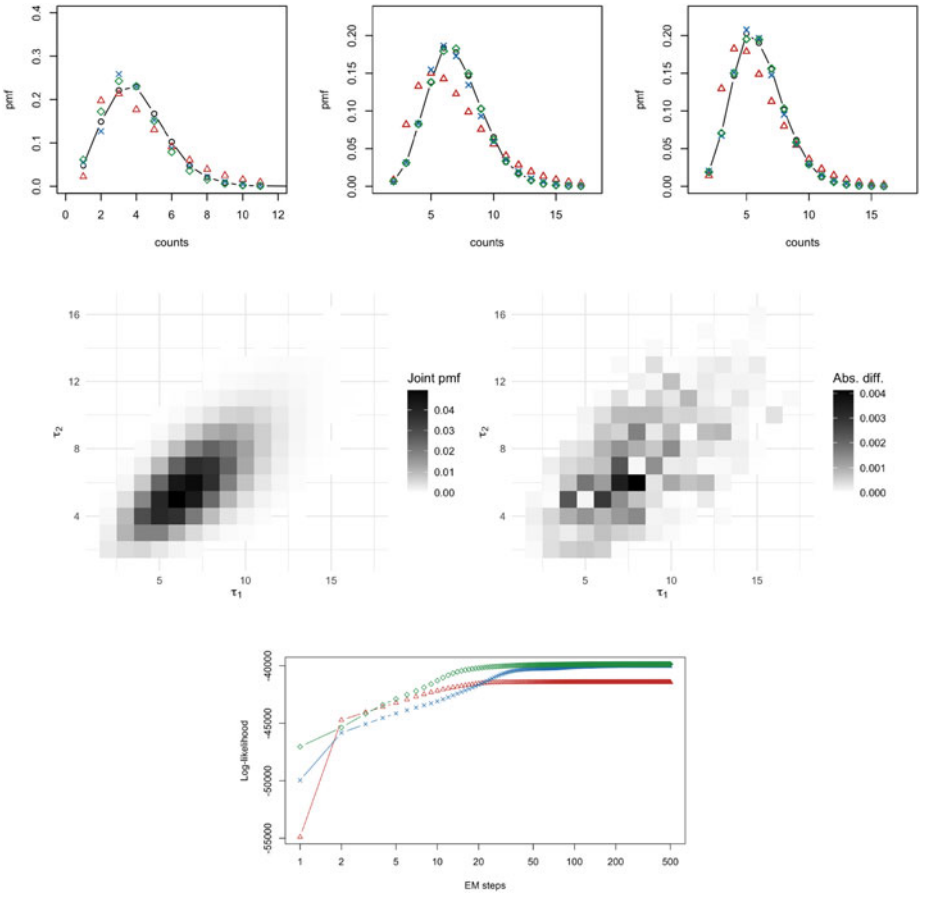


Figure 5. *Bivariate Poisson distribution with common shock intensity $\lambda_Z = 3$. In the top panels we show the empirical pmf (black circles) versus the fitted pmf (by increasing matrix dimension: red triangles, blue crosses, green diamonds), for the distribution of the common shock (left), and the two marginals (center and right). In the middle panel we have two heatmaps, corresponding to the best fitted bivariate pmf (left), and the absolute difference between the best fitted and empirical bivariate pmf (right). The bottom panel shows the evolution of the likelihood for the three models as the number of EM steps increases to 500.*

https://doi.org/10.1017/asb.2025.10062 Published online by Cambridge University Press