Probability in the Engineering and Informational Sciences, **35**, 2021, 708–729.
doi:10.1017/S0269964820000157

BAYESIAN ANALYSIS OF DOUBLY STOCHASTIC MARKOV PROCESSES IN RELIABILITY

Atilla Ay and Refik Soyer
Department of Decision Sciences, The George Washington University, Washington, DC 20052, USA
E-mail: aay@gwu.edu

Joshua Landon
Department of Statistics, The George Washington University, Washington, DC 20052, USA

Süleyman Özekici
Department of Industrial Engineering, Koç University, 34450 İstanbul, Turkey

Markov processes play an important role in reliability analysis and particularly in modeling the stochastic evolution of survival/failure behavior of systems. The probability law of Markov processes is described by its generator or the transition rate matrix. In this paper, we suppose that the process is doubly stochastic in the sense that the generator is also stochastic. In our model, we suppose that the entries in the generator change with respect to the changing states of yet another Markov process. This process represents the random environment that the stochastic model operates in. In fact, we have a Markov modulated Markov process which can be modeled as a bivariate Markov process that can be analyzed probabilistically using Markovian analysis. In this setting, however, we are interested in Bayesian inference on model parameters. We present a computationally tractable approach using Gibbs sampling and demonstrate it by numerical illustrations. We also discuss cases that involve complete and partial data sets on both processes.

Keywords: Bayesian inference, Bayesian reliability analysis, hidden Markov model, Markov modulated Markov process

1. INTRODUCTION

Markov processes are perhaps the most widely used processes in many fields of science, engineering and management to represent random phenomena. There is an abundant literature on Markovian analysis of queueing, inventory and reliability models among many others. These models provide a more realistic framework by allowing additional uncertainty where the system under consideration operates in a stochastic environment. This stochastic environmental process, therefore, modulates our model where its parameters change with respect to the changing states of the environmental process. In this approach, there are two processes: an environmental or modulating process that represents all factors that affect our model, and another process that represents the state of the system that we are interested in analyzing. The main assumption behind our model is that the environmental process is a Markov process, and the modulated process also exhibits Markovian behavior in any

 0269-9648/19 $25.00

given state of the environment. Therefore, we refer to this as a Markov modulated Markov process. The Markov process is doubly stochastic in the sense that the generator at any time is random and it changes with respect to the states of the modulating environmental process.

There are many examples of modulation in the reliability literature. For example, random environments are used to provide a tractable model of the stochastic dependence among the components of a device where the environment is an external process that depicts all physical, structural, operational and other conditions which affect the deterioration, aging and failure of the system. Since all components are subject to the same environmental conditions, their lifetimes are dependent via their common environmental process. Thus, the environmental process is actually a factor of variation in the failure structure of the system. These ideas were introduced by Çınlar and Özekici [11] who propose to construct an intrinsic clock which ticks differently in different environments to measure the intrinsic age of the device. This intrinsic aging model is studied further by Çınlar et al. [12] to determine the conditions that lead to associated component lifetimes. The association of the lifetimes of components subjected to a randomly varying environment is discussed by Lefèvre and Milhaud [17]. Singpurwalla [28] provides a review by discussing hazard potentials in reliability modeling. Applications also include hardware reliability where a device performs a stochastic mission and its failure rate depends on the stage of the mission. Çekyay and Özekici [5] discuss issues related to mean time to failure and availability when the mission or environmental process is semi-Markovian. The reader is referred to Çekyay and Özekici [8,9] for issues related to performance analysis and maintenance of such modulated reliability models. First consideration of modulation in software reliability applications is due to Özekici and Soyer [21] who assume that the failures of the software depend on its operational profile, which is now the environmental process that represents the sequence of operations that the software performs. In a recent article, Landon et al. [16] present a tractable Bayesian approach Markov modulated Poisson model for software reliability. A discrete-time hidden Markov process is considered, to describe software failures, in Pievatolo et al. [23]

The use of Markovian modulation is not limited to applications in reliability. In queueing, Prabhu and Zhu [24] discusses modulated queues and provides a survey of earlier papers, including Eisen and Tainiter [13], Neuts [19] and Purdue [25]. Zhu [30] discusses Markov modulated queueing networks and shows that the steady-state distribution of the queue length has a product form solution. There are other stochastic models where modulation is used. Arifoğlu and Özekici [2] analyze an inventory model operating in a partially observable random environment where the demand process is modulated by a process that represents the stochastic variations in an economy. Finally, modulation also occurs in portfolio optimization problems when the random asset returns are modulated in a so-called regime-switching market, as in Çanakoğlu and Özekici [4].

Although Bayesian analysis of Markov modulated Poisson processes has been considered by Fearnhead and Sherlock [14], Bayesian methods have not been developed for the general class of modulated Markov processes. Our main contribution in this paper is to generalize the methods of Fearnhead and Sherlock [14] for Bayesian analysis of other Markov models, such as modulated birth-death and compound Poisson processes, which arise in reliability applications. In particular, we focus on Bayesian analysis when the states of the environmental or modulating process are unobserved based on observed data on the modulated Markov process. The environmental process is, therefore, a hidden Markov process. Recall that the probabilistic structure of a Markov process includes exponential holding rates in a state couples with transition probabilities between states. The inference will include not only the holding rates and transition probabilities of the observed process but the holding rates and the transition probabilities of the hidden Markov process as well. Our analysis will also focus on the unknown number of states of the environmental process. The details of our model will be presented in Section 2 where the stochastic structures of the modulating and modulated processes are described. This section also includes the modulated compound Poisson process model, which is motivated by the power outage example considered in Section 6. In Section 3, we will assume that the number of states of the hidden process is known and show how we can estimate the holding rates as well as the transition rates of the Markov process. Then, in Section 4, we will assume that we do not know the number of states of the hidden Markov process and will present an approach to obtain the marginal likelihood based on Chib [10] that will enable us to infer the unknown number of states. Our results will be demonstrated using simulated and real data in Section 6.

2. DOUBLY STOCHASTIC MARKOV MODEL

Let $Z = \{Z_t;\ t \geq 0\}$ be a stochastic process such that Z_t depicts the state of the system that we are interested in at time t. In a queueing model, it is the number of customers in the system, while it may be the level of inventory in an inventory model. In a financial model, it may represent the price of a risky asset as it evolves randomly over time. We suppose that Z has a finite state space $F = \{1, 2, \ldots, N\}$, and it has a Markov structure. There is an environmental process $Y = \{Y_t;\ t \geq 0\}$, where Y_t represents the state of the environmental at time t, and it has an effect on the process Z. We assume that Y is a continuous-time Markov process with a finite state space $E = \{1, 2, \ldots, K\}$, where K is the number of states. The generator of the Markov process Y is

$$G_{ij} = \begin{cases} -\rho_i, & \text{if } j = i \\ \rho_i P_{ij}, & \text{if } j \neq i \end{cases} \tag{1}$$

or $G_{ij} = \rho_i(P_{ij} - I_{ij})$, where I is the identity matrix. In other words, the process Y spends an exponential amount of time with holding rate ρ_i in state i and, when it jumps, it randomly goes to state j with transition probability P_{ij}, where $P_{ii} = 0$.

The probabilistic structure of Z is such that while the state of the environment Y_t is $i \in E$, the process Z evolves as a Markov process with generator

$$A_i(x, y) = \begin{cases} -\lambda_i(x), & \text{if } y = x \\ \lambda_i(x) M_i(x, y), & \text{if } y \neq x \end{cases}. \tag{2}$$

This implies that while the process Y is in state i, the process Z spends an exponential amount of time with holding rate $\lambda_i(x)$ in state x and, when it jumps, it randomly goes to state y with transition probability $M_i(x, y)$, where $M_i(x, x) = 0$. We call Z a doubly stochastic Markov process (DSMP) that is modulated by the Markov process Y. The DSMP Z evolves like a Markov process with generator A_{Y_t} at any time t. The parameters of our model involve the exponential holding rates and Markovian transition probabilities given in (1) and (2).

It is clear that the bivariate process $(Y, Z) = \{(Y_t, Z_t);\ t \geq 0\}$ is a Markov process with state space $E \times F$ where the generator Q of (Y, Z) is

$$Q_{(i,x),(j,y)} = \begin{cases} \rho_i P_{ij}, & j \neq i, y = x \\ \lambda_i(x) M_i(x, y), & j = i, y \neq x \\ -(\rho_i + \lambda_i(x)), & j = i, y = x \end{cases} \tag{3}$$

for all $i, j \in E$ and $x, y \in F$.

The stochastic structure of the Markovian model described by (1)–(3) can be used in connection with a variety of applications. Among them, reliability and maintenance applications are perhaps the most widely known. In phased-mission or mission-based models, for example, Y represents the mission process that consists of a random sequence of stages with random durations that a device has to go through. The deterioration or aging of the device is depicted by the process Z such that it depends probabilistically on the stage of the mission performed at any given time. Çekyay and Özekici [6,7] discuss reliability and maintenance issues associated with such a model.

Note that the well-known doubly stochastic Poisson process is a special case when $\lambda_i(x) = \lambda_i$ and $M_i(x, x+1) = 1$. In this setting, Z_t represents the total number of customer arrivals until time t with initial state $Z_0 = 0$. While the environment is in state $Y_t = i$, customers arrive according to a Poisson process with rate λ_i. The reader is referred to Özekici and Soyer [22] for a discussion on modulated Poisson processes.

The Markov modulated birth-death process is another special case of DSMP when $\lambda_i(x) = \gamma_i + \mu_i$ and

$$M_i(x, y) = \begin{cases} \dfrac{\gamma_i}{\gamma_i + \mu_i}, & y = x + 1 \\ \dfrac{\mu_i}{\gamma_i + \mu_i}, & y = x - 1 \end{cases}$$

where γ_i and μ_i are the birth and death rates of the process and $M_i(0, 1) = 1$.

The DSMP model becomes a Markov modulated compound Poisson process when

$$Z_t = \sum_{k=1}^{N_t} W_k \tag{4}$$

where N is a Poisson process with jump rate λ_i in environment i, W_k's are independent and identically distributed random variables representing the jump size, and $M_i(x, y)$ is the probability of having a jump with magnitude $(y - x)$ in environment i. One possibility is to assume that W_k's all have the binomial distribution with parameters n_i and p_i, while the jump occurs in environment i so that

$$M_i(x, x+z) = \binom{n_i}{z} p_i^z (1 - p_i)^{n_i - z} \tag{5}$$

for $z = 0, 1, 2, \ldots, n_i$. We will indeed use this model in the context of a reliability application in Section 6.

Since (Y, Z) is a Markov process with generator matrix Q, its transition function

$$P_{(i,x),(j,y)}(t) = P[Y_t = j, Z_t = y \,|\, Y_0 = i, Z_0 = x] \tag{6}$$

is given by the exponential matrix defined via the Taylor expansion

$$P(t) = \exp(Qt) = \sum_{n=0}^{+\infty} \frac{t^n}{n!} Q^n. \tag{7}$$

There are computationally tractable procedures to use the exponential matrix (7) as discussed in Moler and van Loan [18]. These matrices are very useful in the analysis of stochastic models with Markovian modulation. We refer the reader to Neuts [20] for details and various results on the exponential matrix that we will be using in our analysis. Asmussen [3] provides a survey on Markovian point processes and discusses how they are used in applied probability calculations.

Using the conditional joint distribution (6) of (Y_t, Z_t) given by the exponential matrix (6), one can easily compute the conditional marginal distributions of Y_t and Z_t. If the initial distribution of (Y_0, Z_0) is known, these results can be used in a trivial manner to compute the unconditional distributions. Similarly, if the Markov process (Y, Z) is ergodic, then its steady-state or limiting distribution π be determined using its generator Q by solving the system of linear equations $\pi Q = 0$ with the normalizing equation. In summary, using the matrix exponential form (7) and tools to compute it, one can obtain many important performance measures associated with the transient and ergodic behavior of the DSMP. Our primary objective is to develop statistical inference for the DSMP using a Bayesian framework. It is important to note that in our model, the Y process is latent and, therefore, in addition to the unknown parameters, we also need to make inference about the latent states.

3. BAYESIAN ANALYSIS OF THE DSMP

In this section, we will illustrate how we can estimate all the parameters as well as the latent states in the DSMP. The approach is based on an extension of the Markov Chain Monte Carlo (MCMC) method given in Fearnhead and Sherlock [14] for Poisson processes. This method is based on a Gibbs sampler and requires a three-stage process. We denote the holding rates of the Y process as $\rho = \{\rho_i;\ i \in E\}$ and transition probabilities as $\mathbf{P} = \{P_{ij};\ i, j \in E\}$. Similarly, the holding rates of the DSMP are denoted as $\lambda = \{\lambda_i(x);\ i \in E,\ x \in F\}$ and transition probabilities as $\mathbf{M} = \{M_i(x, y);\ i \in E,\ x, y \in F\}$.

We assume that the system process Z is observed completely until some time t_{obs}, while the process Y is latent. We suppose that the process Z changes its state a total of n times during $[0, t_{\text{obs}}]$, and we let $t^{(1)} < t^{(2)} < \cdots < t^{(n)}$ denote the jump times of the Z process. To simplify the notation, we will let $z^{(k)} = Z_{t^{(k)}}$ denote the observed state of Z after the kth change of state, and $Y^{(k)} = Y_{t^{(k)}}$ denote the unobserved state the hidden Markov process at the kth time of change of the observed process Z. We also set $t^{(0)} = 0, t^{(n+1)} = t_{\text{obs}}$ for completeness, and let $z^{(0)} = Z_0$ denote the initial observed state, while $z^{(n+1)} = z^{(n)}$ is the last state observed. It is clear that $\{t^{(k)}\}$ and $\{z^{(k)}\}$ are all contained in the history or data set $\mathcal{D} = \{z_t;\ 0 \leq t \leq t_{\text{obs}}\}$, where z_t is the state of the DSMP observed at time t.

In Stage 1, we will simulate the state of the hidden Markov process at each of the times $t^{(1)}, t^{(2)}, \ldots, t^{(n)}$ given in our data set $\mathcal{D}$ and conditional on the parameters ρ, $\mathbf{P}, \lambda$ and $\mathbf{M}$. In Stage 2, we will simulate the entire hidden Markov process, and in Stage 3, we will simulate a new set of parameter values using the simulated history of the hidden Markov process and observed data.

3.1. Stage 1: Simulation of the hidden Markov process at event times

At any event time $t^{(k)}$, we also observe the state of the process $z^{(k)}$. Moreover, as long as the state of the Markov process Y is i at an event time $t^{(k)}$, the process Z jumps exponentially with rate $\lambda_i(z^{(k)})$ and, when it jumps at the next event time $t^{(k+1)}$, it jumps to state $z^{(k+1)}$ with probability $M_j(z^{(k)}, z^{(k+1)})$ if the state of the latent process Y is j at time $t^{(k+1)}$. We let $t_k = t^{(k)} - t^{(k-1)}$ denote the length of the kth interval for $k = 1, 2, \ldots, n+1$ where no jump of Z occurs at $t^{(n+1)} = t_{\text{obs}}$. Finally, we represent our data set as $\mathcal{D} = (d_0, d_1, \ldots, d_{n+1})$, where $d_k = (t_k, z^{(k)})$ contains information over the kth interval.

For each interval $k = 1, 2, \ldots n+1$, we define the diagonal matrix

$$\Lambda_{ij}^{(k)} = \begin{cases} \lambda_i(z^{(k-1)}), & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases} \tag{8}$$

https://doi.org/10.1017/S0269964820000157 Published online by Cambridge University Press

and

$$G_{ij}^{(k)} = G_{ij} - \Lambda_{ij}^{(k)}, \tag{9}$$

as well as a likelihood matrix

$$A_{ij}^{(k)} = P[d_k, d_{k+1}, \ldots, d_{n+1}, Y^{(n+1)} = j \,|\, Y^{(k-1)} = i, z^{(k-1)}] \tag{10}$$

for $k = 1, 2, \ldots, n+1$. Then, it is clear from (A.6) in the Appendix that

$$\begin{aligned} A_{ij}^{(n+1)} &= P[d_{n+1} = (t_{n+1}, z^{(n+1)} = z^{(n)}), \\ Y^{(n+1)} &= j \,|\, Y^{(n)} = i, z^{(n)}] = \left[\exp\left(G^{(n+1)} t_{n+1}\right)\right]_{ij}. \end{aligned} \tag{11}$$

Starting with $A_{ij}^{(n+1)}$ in (11) and proceeding backwards, we can determine $A^{(k)}$ for all $k = 1, 2, \ldots, n$ as

$$A_{ij}^{(k)} = \sum_{l \in E} P[Y^{(k)} = l, t_k, z^{(k)}, d_{k+1}, \ldots, d_{n+1}, Y^{(n+1)} = j \,|\, Y^{(k-1)} = i, z^{(k-1)}],$$

which yields

$$\begin{aligned} A_{ij}^{(k)} = &\sum_{l \in E} P[Y^{(k)} = l, t_k, z^{(k)} \,|\, Y^{(k-1)} = i, z^{(k-1)}] \\ &\times P\left[d_{k+1}, \ldots, d_{n+1}, Y^{(n+1)} = j \,|\, Y^{(k)} = l, t_k, z^{(k)}, Y^{(k-1)} = i, z^{(k-1)}\right]. \end{aligned} \tag{12}$$

Now using the Markov property, the second term inside the summation is

$$P\left[d_{k+1}, \ldots, d_{n+1}, Y^{(n+1)} = j \,|\, Y^{(k)} = l, t_k, z^{(k)}, Y^{(k-1)} = i, z^{(k-1)}\right] = A_{lj}^{(k+1)}. \tag{13}$$

Using the main result (A.15) in the Appendix, the first term can be written as follows:

$$\begin{aligned} P[Y^{(k)} &= l, T_\Delta = t_k, Z_{T_\Delta} = z^{(k)} \,|\, Y_0 = i, Z_0 = z^{(k-1)}] \\ &= \left[\exp\left(G^{(k)} t_k\right)\right]_{il} \lambda_l^{(k)} M_l(z^{(k-1)}, z^{(k)}) = \left[T^{(k)} L^{(k)}\right]_{il} \end{aligned}$$

with a slight abuse of notation where

$$T_{ij}^{(k)} = \left[\exp\left(G^{(k)} t_k\right)\right]_{ij} \tag{14}$$

and

$$L_{ij}^{(k)} = \Lambda_{ij}^{(k)} M_j(z^{(k-1)}, z^{(k)}). \tag{15}$$

Finally, putting (12)–(15) together, we obtain the matrix recursion

$$A^{(k)} = T^{(k)} L^{(k)} A^{(k+1)} \tag{16}$$

for $k = 1, 2, \ldots, n$, where the boundary matrix is $A^{(n+1)} = \exp(G^{(n+1)} t_{n+1}) = T^{(n+1)}$ from (11). It follows from this recursion that

$$A^{(k)} = \left[\prod_{m=k}^{n} T^{(m)} L^{(m)}\right] T^{(n+1)} \tag{17}$$

and the final recursion gives us the likelihood

$$A_{ij}^{(1)} = P[d_1, d_2, \ldots, d_{n+1}, Y_{t^{(n+1)}} = j \,|\, Y_0 = i, z^{(0)}] = P[\mathcal{D}, Y^{(n+1)} = j \,|\, Y_0 = i, z^{(0)}]. \quad \textbf{(18)}$$

Suppose that the initial and final states of Y are known as $Y_0 = i$ and $Y^{(n+1)} = j$. Then, the state of the hidden Markov chain Y at time $t^{(k)}$ can be simulated using the conditional distribution

$$P[Y^{(k)} = l \,|\, \mathcal{D}, Y^{(k-1)} = i, Y^{(n+1)} = j] = \frac{T_{il}^{(k)} L_{ll}^{(k)} A_{l,j}^{(k+1)}}{A_{i,j}^{(k)}} \quad \textbf{(19)}$$

recursively by proceeding forwards through the event times $t^{(1)}, t^{(2)}, \ldots, t^{(n)}$ by taking $k = 1, 2, \ldots, n$.

If the initial state Y_0 and the final state $Y^{(n+1)}$ are unknown, as it might be expected in applications, the algorithm can be adjusted trivially as discussed in Landon et al. [16].

3.2. Stage 2: Complete simulation of the hidden Markov process

After completing Stage 1, we will have our simulated states of the hidden Markov process Y at each of our observation times $\{t^{(k)}\}$. We will now use these to simulate the entire hidden Markov process Y. To do this, we first of all simulate it over the interval $(t^{(0)}, t^{(1)})$, then $(t^{(1)}, t^{(2)})$, and so on until $(t^{(n)}, t^{(n+1)})$. The simulation over each interval is done using the uniformization of the Markov process Y, supposing that $\rho = \max_{i \in E} \rho_i$ is finite. It is well-known (see, e.g., Ross [26]) that the Markov process Y can be represented as a Markov chain $\hat{X}$ subordinated to a Poisson process $\hat{N}$ with arrival rate ρ so that $Y_t = \hat{X}_{\hat{N}_t}$ and

$$\begin{aligned} P[Y_t = j \,|\, Y_0 = i] &= P[\hat{X}_{\hat{N}_t} = j \,|\, \hat{X}_{\hat{N}_0} = i] \\ &= \sum_{n=0}^{+\infty} P[\hat{N}_t = n] P[\hat{X}_n = j \,|\, \hat{X}_0 = i] \\ &= \sum_{n=0}^{+\infty} \frac{e^{-\rho t} (\rho t)^n}{n!} R_{i,j}^n, \end{aligned}$$

where

$$R = \frac{1}{\rho} G + I \quad \textbf{(20)}$$

is the transition matrix corresponding to the Markov chain $\hat{X}$. Over any interval $(t^{(k-1)}, t^{(k)})$, we already obtained the simulated states $Y_{t^{(k-1)}} = i_{k-1}$ and $Y_{t^{(k)}} = i_k$ in Stage 1. Therefore, the conditional distribution of the number arrivals of $\hat{N}$ during $(t^{(k-1)}, t^{(k)})$ is

$$\begin{aligned} & P[\hat{N}_{t^{(k)}} - \hat{N}_{t^{(k-1)}} = n \,|\, Y_{t^{(k-1)}} = i_{k-1}, Y_{t^{(k)}} = i_k] \\ & = \left(\frac{e^{-\rho t_k} (\rho t_k)^n}{n!} \right) \left(\frac{R_{i_{k-1}, i_k}^n}{\exp{[G t_k]}_{i_{k-1}, i_k}} \right) \end{aligned} \quad \textbf{(21)}$$

since

$$P[Y_{t^{(k)}} = i_k \,|\, Y_{t^{(k-1)}} = i_{k-1}] = \exp{[G t_k]}_{i_{k-1}, i_k}.$$

Therefore, the number of arrivals $\hat{N}_{t^{(k)}} - \hat{N}_{t^{(k-1)}}$ can be simulated using the distribution (21). If simulation yields $\hat{N}_{t^{(k)}} - \hat{N}_{t^{(k-1)}} = r$, then the r arrival times $\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_r$ of $\hat{N}$ over the interval $(t^{(k-1)}, t^{(k)})$ are simulated by generating r uniform variates over $(t^{(k-1)}, t^{(k)})$ and ordering them. Now, we know that $Y_{t^{(k-1)}} = i_{k-1}$ and $Y_{t^{(k)}} = i_k$ and the states of hidden Markov process at $\hat{t}_1 \leq \hat{t}_2 \leq \cdots \leq \hat{t}_r$ are simulated recursively by using the conditional distributions

$$P[Y_{\hat{t}_j} = s \,|\, Y_{\hat{t}_{j-1}} = \hat{s}_{j-1}, Y_{t^{(k)}} = i_k] = \frac{R_{\hat{s}_{j-1},s} R^{r-j}_{s,i_k}}{R^{r-j+1}_{\hat{s}_{j-1},i_k}} \tag{22}$$

for $j = 1, 2, \ldots, r$. For $j = 1$, one should set $\hat{t}_{j-1} = \hat{t}_0 = t^{(k-1)}$ and $\hat{s}_{j-1} = \hat{s}_0 = i_{k-1}$. It also follows from the conditional distribution (22) that $Y_{\hat{t}_r} = Y_{t^{(k)}} = i_k$ at the last time point when $j = r$ since R^0 is the identity matrix.

3.3. Stage 3: Generation of new parameters using data and the hidden process

Having completed Stages 1 and 2, we should now have the entire simulated hidden Markov process, as well as our data $\mathcal{D}$ on the observed DSMP. Let $\mathcal{F} = \{y_t;\ 0 \leq t \leq t^{(n+1)}\}$ denotes the environmental process generated using the procedure in Stage 2. Thus, we can write out our conditional likelihood function of the parameters and then obtain the full conditionals to generate a new set of values for our parameters at each step of the Gibbs sampler. Let τ_i be the total time that the hidden Markov process Y spends in state i; $\hat{\tau}_i(x)$ be the total amount of time that the modulated process Z spends in state x, while the state of Y is i; r_{ij} be the number of times the hidden process Y makes a transition from state i to state j and $\hat{r}_i(x, y)$ be the number of times that the process Z jumps from state x to y, while the hidden process Y is in state i. It is clear that $\tau_i, r_{ij}, \hat{\tau}_i(x)$ and $\hat{r}_i(x, y)$ are in $\mathcal{F} \cup \mathcal{D}$ for all $i, j \in E$ and $x, y \in F$. Given data $\mathcal{D}$ and the entire history $\mathcal{F}$ of the Markov process, the conditional likelihood function of the parameters ρ, $\mathbf{P}, \lambda$ and $\mathbf{M}$ is given by

$$\mathcal{L}(\rho, \mathbf{P}, \lambda, \mathbf{M}; \mathcal{F}, \mathcal{D}) \propto \prod_{i \in E} \Bigg[\rho_i^{\sum_{j \in E} r_{ij}} \exp(-\rho_i \tau_i) \prod_{j \in E} P_{ij}^{r_{ij}} \Bigg[\prod_{x \in F} \Bigg[\lambda_i(x)^{\sum_{y \in F} \hat{r}_i(x,y)} \times \exp(-\lambda_i(x) \hat{\tau}_i(x)) \prod_{y \in F} M_i(x, y)^{\hat{r}_i(x,y)} \Bigg] \Bigg] \Bigg]. \tag{23}$$

Assuming conjugate independent priors for the unknown parameters, the full conditional distributions can be easily obtained. More specifically, for a given states $i = 1, \ldots, K$ and $x = 1, 2, \ldots, N$, we assume independent gamma priors for ρ_i's denoted as $\rho_i \sim \mathcal{G}(a_i^\rho, b_i^\rho)$, and for $\lambda_i(x)$'s denoted as $\lambda_i(x) \sim \mathcal{G}(a_i^\lambda(x), b_i^\lambda(x))$. For the ith row of the transition matrix P, we assume a Dirichlet prior, independent of the other rows, as $\mathbf{P}_i \sim \mathcal{D}ir(\alpha_{i1}, \ldots, \alpha_{iK})$, where $\mathbf{P}_i = (P_{i1}, \ldots, P_{iK})$. Note that in $\mathbf{P}_i$, we have $P_{ii} = 0$ and the corresponding parameter $\alpha_{ii} = 0$. Similarly, for the xth row of the transition matrix M_i, we assume a Dirichlet prior, independent of the other rows, as $\mathbf{M}_i(x) \sim \mathcal{D}ir(\beta_i(x, 1), \ldots, \beta_i(x, N))$, where $\mathbf{M}_i(x) = (M_i(x, 1), \ldots, M_i(x, N))$. Note that $M_i(x, x) = 0$ and the corresponding parameter is $\beta_i(x, x) = 0$.

Using standard Bayesian results, we can show that given the full history of the hidden Markov process, the full conditional distributions of the parameters can be obtained as

https://doi.org/10.1017/S0269964820000157 Published online by Cambridge University Press

follows:

$$\begin{aligned} \rho_i \,|\, \rho_i^{-i} &\sim \mathcal{G}\left(a_i^{\rho} + \sum_{j \in E} r_{ij}, b_i^{\rho} + \tau_i\right), \\ \lambda_i(x) \,|\, \lambda_i^{-i}(x) &\sim \mathcal{G}\left(a_i^{\lambda} + \sum_{y \in F} \hat{r}_i(x,y), b_i^{\lambda} + \hat{\tau}_i(x)\right) \end{aligned} \tag{24}$$

and

$$\mathbf{P}_i \,|\, \mathbf{P}_i^{-i} \sim \mathcal{D}ir(\alpha_{i1} + r_{i1}, \ldots, \alpha_{iK} + r_{iK}) \tag{25}$$

$$\mathbf{M}_i(x) \,|\, \mathbf{M}_i(x)^{-1} \sim \mathcal{D}ir(\beta_i(x,1) + \hat{r}_i(x,1), \ldots, \beta_i(x,N) + \hat{r}_i(x,N)) \tag{26}$$

where $\alpha_{ii} = r_{ii} = \beta_i(x,x) = \hat{r}_i(x,x) = 0$.

We then generate new values for these parameters from their posterior distribution and then repeat the whole process again, starting with Stage 1.

4. ASSESSMENT OF THE NUMBER OF ENVIRONMENTAL STATES

Our analysis in Section 3 assumed that the number of states K in the hidden Markov process was known. However, in general, the actual number of states may be unknown to us, so it is important to be able to determine how many states there are. The problem of determining K can be considered as a model selection problem in the Bayesian approach where the model choice is made using Bayes factors; see Kass and Raftery [15] for a review. The computation of the Bayes factors requires the evaluation of marginal likelihood for a given model, that is, for given value of K in our case. More specifically, if we let $\mathcal{D}$ denote our observed data, we want to obtain the marginal likelihood $p(\mathcal{D}|K)$. The model with the highest value of $p(\mathcal{D}\,|\,K)$ is the one most supported by the data, and this can be used as the criterion for determining the value of K. Alternatively, assuming a support for K and specifying prior probabilities $P[K=k]$ for different models such that $\sum_k P[K=k]=1$, we can obtain posterior model probabilities $P[K=k\,|\,\mathcal{D}]$ using the marginal likelihood.

Evaluation of the marginal likelihood $p(\mathcal{D}\,|\,K)$ analytically is not possible in many problems since it requires integrating out the unknown parameters. Since draws from prior distributions of the parameters result in unstable estimation, the use of Monte Carlo methods emphasize the use of posterior Monte Carlo samples to evaluate $p(\mathcal{D}\,|\,K)$. Although this is not straightforward in many cases, when the full posterior conditional distributions are known forms, the marginal likelihood terms can be approximated using the approach proposed by Chib [10]. Since the Bayesian analysis of the DSMP in Section 3 is based on known full conditionals, we can adopt Chib's procedure to our problem as will be discussed in the sequel.

In our case, the marginal likelihood for a specific model with dimension K is given by the following equation:

$$p(\mathcal{D}) = \frac{p(\mathcal{D}\,|\,\rho, \mathbf{P}, \lambda, \mathbf{M}, \mathcal{F})\; p(\rho, \mathbf{P}, \lambda, \mathbf{M}, \mathcal{F})}{p(\rho, \mathbf{P}, \lambda, \mathbf{M}, \mathcal{F}\,|\,\mathcal{D})} \tag{27}$$

where ρ is a K-dimensional vectors of ρ_i's and $\mathbf{P}$ is the transition probability matrix of dimension K with zeros on the diagonal. Likewise, λ is a $K \times N$ dimensional vector of

$\lambda_i(x)$'s and $\mathbf{M}$ consists of K matrices each with dimension N and zero entries on the diagonals. We can rewrite (27) as

$$p(\mathcal{D}) = \frac{p(\mathcal{D} \mid \rho, \mathbf{P}, \lambda, \mathbf{M}, \mathcal{F})\ p(\mathcal{F} \mid \rho, \mathbf{P}) p(\rho, \lambda, \mathbf{M}) p(\mathbf{P})}{p(\rho, \mathbf{P}, \lambda, \mathbf{M} \mid \mathcal{F}, \mathcal{D}) p(\mathcal{F} \mid \mathcal{D})}. \tag{28}$$

Equation (28) holds for any values of $(\rho, \mathbf{P}, \lambda, \mathbf{M}, \mathcal{F})$ such as $(\rho^*, \mathbf{P}^*, \lambda^*, \mathbf{M}^*, \mathcal{F}^*)$ which is typically chosen as the mean or mode values of the posterior distributions. We note that all the terms in the numerator are available to us analytically and, therefore, can be evaluated at $(\rho^*, \mathbf{P}^*, \lambda^*, \mathbf{M}^*, \mathcal{F}^*)$. The tricky part to evaluate is the denominator term

$$p(\mathcal{F}^* \mid \mathcal{D}) = \int p(\mathcal{F}^* \mid \mathcal{D}, \rho, \mathbf{P}, \lambda, \mathbf{M}) p(\rho, \mathbf{P}, \lambda, \mathbf{M} \mid \mathcal{D}) d(\rho, \mathbf{P}, \lambda, \mathbf{M})$$

which can be evaluated using G samples from the posterior distribution via

$$p(\mathcal{F}^* \mid \mathcal{D}) = \frac{1}{G} \sum_{g=1}^{G} p(\mathcal{F}^* \mid \rho^{(g)}, \mathbf{P}^{(g)}, \lambda^{(g)}, \mathbf{M}^{(g)}, \mathcal{D}). \tag{29}$$

The first term $p(\rho^*, \mathbf{P}^*, \lambda^*, \mathbf{M}^* \mid \mathcal{F}^*, \mathcal{D})$ can easily be written down as the product of gamma and Dirichlet densities. Thus, for each value of K, we can approximate (28) and determine the model with the highest support of the data. As previously mentioned, using the marginal likelihood, we can also compute posterior model probabilities $P[K = k \mid \mathcal{D}]$ to infer the value of K.

5. TREATMENT OF DIFFERENT TYPES OF DATA SETS

The Bayesian analysis discussed until now is based on the data set $\mathcal{D} = \{z_t;\ 0 \le t \le t_{\text{obs}}\}$ that includes information observed only on the modulated process Z. Since the Markov process Y is unobserved, the procedure described in Stages 1 and 2 of Section 3 is used to simulate it. Then, in Stage 3, the simulated process $\mathcal{F} = \{y_t;\ 0 \le t \le t_{\text{obs}}\}$ is put together with $\mathcal{D} = \{z_t;\ 0 \le t \le t_{\text{obs}}\}$ in order to compute $\tau_i, r_{ij}, \hat{\tau}_i(x)$ and $\hat{r}_i(x, y)$ in $\mathcal{F} \cup \mathcal{D}$. This leads to the posteriors (24)–(26) through the likelihood (23). The case that we considered involves complete data on the modulated process Z only because this is indeed the most interesting situation. In this section, we briefly comment on how this procedure can be used with different data sets.

5.1. Complete data on Y and Z

The application of the procedure when data on both Z ($\mathcal{D} = \{z_t;\ 0 \le t \le t_{\text{obs}}\}$) and Y ($\mathcal{F} = \{y_t;\ 0 \le t \le t_{\text{obs}}\}$) are available is fairly trivial. In this case, we do not need to perform Stages 1 and 2 of the simulation step. Using the given data set $\mathcal{F} \cup \mathcal{D}$, we can directly use Stage 3 to determine the posteriors.

5.2. Complete data on Y

Suppose that the data set $\mathcal{F} = \{y_t;\ 0 \le t \le t_{\text{obs}}\}$ is on the hidden Markov process Y, and the modulated Markov process Z is unobserved. Then, one can easily simulate the process Z on $0 \le t \le t_{\text{obs}}$ to obtain $\mathcal{D} = \{z_t;\ 0 \le t \le t_{\text{obs}}\}$. This can be done fairly easily by noting that, while the hidden process Y is observed to be in some state $i \in E$ during any interval, the modulated Markov process evolves according to a Markov process with generator A_i. Therefore, if the initial state of Z is $x \in F$ at the beginning of the interval, then it remains there for an exponential amount of time with rate $\lambda_i(x)$. If this exponential sojourn ends before the end of the interval, then the next state of Z becomes $y \in F$ with probability $M_i(x,y)$. On the other hand, if the exponential sojourn exceeds the end of the interval, then the new state of Y over the next interval is used in a similar manner. Given $\mathcal{F} = \{y_t;\ 0 \le t \le t_{\mathrm{obs}}\}$, the states of Z can be simulated using time variates generated from the exponential distributions with rates $\{\lambda_i(x)\}$ and state variates generated from the discrete distributions $\{M_i(x,\cdot)\}$. Once $\mathcal{D} = \{z_t;\ 0 \le t \le t_{\mathrm{obs}}\}$ is obtained through this simulation, one can use $\mathcal{F} \cup \mathcal{D}$ as before in Stage 3. Note that the simulation stage in this case is much simpler than Stages 1 and 2 in Section 3 for the case with data on Z.

5.3. Partial data (event times) on Y

So far, we had complete data available either on Z, or Y, or both. It is possible that data is available only partially on the processes involved. In this case, the data set $\mathcal{F}_1 = \{s^{(1)}, s^{(2)}, \ldots, s^{(m)}\}$ is on event times of the Markov process Y, where $s^{(k)}$ denotes the time of kth jump of Y observed during $[0, t_{\mathrm{obs}}]$. The states are unobserved and we need to simulate these states $\{Y^{(1)}, Y^{(2)}, \ldots, Y^{(m)}\}$ at the observed event times where $Y^{k)} = Y_{s^{(k)}}$ to obtain complete data $\mathcal{F} = \{y_t; 0 \le t \le t_{\mathrm{obs}}\}$. Once this is accomplished, we can use the procedure outlines above with complete data on Y to do the Bayesian analysis. We now let $s_k = s^{(k)} - s^{(k-1)}$ denote the length of the kth interval for $k = 1, 2, \ldots, m+1$ where no jump of Y occurs at the end of the $(m+1)$st interval at time $s^{(m+1)} = t_{\mathrm{obs}}$. Our data set can equivalently be represented as $\mathcal{F}_1 = (s_1, s_2, \ldots, s_{m+1})$.

Let

$$B_{ij}^{(k)} = P[s_k, s_{k+1}, \ldots, s_{m+1}, Y^{(m+1)} = j \,|\, Y^{(k-1)} = i] \tag{30}$$

for $k = 1, 2, \ldots, m+1$. Since the process Y makes no jump during $[s^{(m)}, s^{(m+1)} = t_{\mathrm{obs}}]$, we have

$$B_{ij}^{(m+1)} = P[s_{m+1}, Y^{(m+1)} = j \,|\, Y^{(m)} = i] = \exp\left(-\rho_i s_{m+1}\right) I_{ij}. \tag{31}$$

We now proceed backward to determine $B^{(k)}$ for all $k = 1, 2, \ldots, m$. Note that

$$\begin{aligned} B_{ij}^{(k)} &= P[s_k, s_{k+1}, \ldots, s_{m+1}, Y^{(m+1)} = j \,|\, Y^{(k-1)} = i] \\ &= \sum_{l \in E} P[Y^{(k)} = l, s_k, s_{k+1}, \ldots, s_{m+1}, Y^{(m+1)} = j \,|\, Y^{(k-1)} = i], \end{aligned}$$

which can be written as follows:

$$\begin{aligned} B_{ij}^{(k)} &= \sum_{l \in E} P[Y^{(k)} = l, s_k \,|\, Y^{(k-1)} = i] \\ &\quad \times P\left[s_{k+1}, \ldots, s_{m+1}, Y^{(m+1)} = j \,|\, Y^{(k)} = l, s_k, Y^{(k-1)} = i\right]. \end{aligned} \tag{32}$$

Now using the Markov property, the second term inside the summation is

$$P\left[s_{k+1}, \ldots, s_{m+1}, Y^{(m+1)} = j \,|\, Y^{(k)} = l, s_k, Y^{(k-1)} = i\right] = B_{lj}^{(k+1)}.$$

Using the fact that Y is a Markov process, the first term can be written as follows:

$$P[Y^{(k)} = l, s_k \,|\, Y^{(k-1)} = i] = \rho_i \exp\left(-\rho_i s_k\right) P_{il} = T_{il}^{(k)}$$

and (32) leads to the recursion

$$B^{(k)} = T^{(k)} A^{(k+1)}$$

for $k = 1, 2, \ldots, m$ where the boundary matrix $B^{(m+1)}$ is given by (31). It follows from this recursion that

$$B^{(k)} = \left[\prod_{n=k}^{m} T^{(n)} \right] B^{(m+1)}$$

and the final recursion gives us the likelihood

$$B_{ij}^{(1)} = P[s_1, s_2, \ldots, s_{m+1}, Y^{(m+1)} = j \,|\, Y_0 = i] = P[\mathcal{F}_1, Y^{(m+1)} = j \,|\, Y_0 = i].$$

If the initial and final states of Y are known as $Y_0 = i$ and $Y^{(m+1)} = j$, then the state of the hidden Markov chain Y at time $s^{(k)}$ can be simulated using the conditional distribution

$$P[Y^{(k)} = l \,|\, \mathcal{F}_1, Y^{(k-1)} = i, Y^{(m+1)} = j] = \frac{T_{il}^{(k)} B_{l,j}^{(k+1)}}{B_{i,j}^{(k)}} \quad \mathbf{(33)}$$

recursively by proceeding forwards through the event times $s^{(1)}, s^{(2)}, \ldots, s^{(m)}$ by taking $k = 1, 2, \ldots, m$.

Once the states $\mathcal{F}_2 = \{y^{(1)}, y^{(2)}, \ldots, y^{(m)}\}$ at event times $\mathcal{F}_1 = \{s^{(1)}, s^{(2)}, \ldots, s^{(m)}\}$ are generated using (33), we obtain $\mathcal{F}_1 \cup \mathcal{F}_2 = \mathcal{F} = \{y_t;\ 0 \le t \le t_{\text{obs}}\}$, and the procedure described above for the case with complete data on Y can be applied to conduct Bayesian analysis.

The case of partial data on the Z process where only jump times are observed can also be handled in a similar manner.

6. NUMERICAL ILLUSTRATIONS

In this section, we discuss two illustrations of the Markov modulated compound Poisson process by applying it to both real and simulated data. This is motivated by the need to predict the time before the next power outage and the number of people affected by the power outages.

The Markov modulated compound Poisson model is given in (4) and (5) where Z_t represents the total number of people affected by any of the power outages. Here, N_t is a Poisson process that gives the number of outages until t, W_k is the number of people affected by outage k, $n_i = n$ represents the total number of households in an area that potentially could be affected by the outages and p_i is the probability for each household to be affected by an outage that occurs in environment i. We assume p_i as well as the holding rate, λ_i, are affected by the state of the environmental process. Given n and p_i, we assume that W_k follows a binomial distribution. In our development, we assume beta priors for p_i denoted as $p_i \sim \mathcal{B}eta(a_i^p, b_i^p)$, for all i. In Section 6.1, we use simulated data to demonstrate the performance of the compound Poisson process modulated by a three-state Markov process. In Section 6.2, we apply a compound Poisson process modulated by a two-state environmental process to actual power outage data in Northern Virginia.

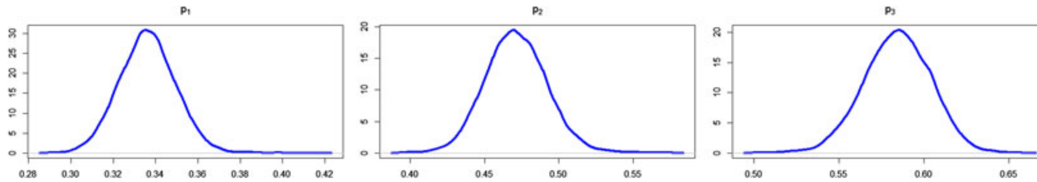

FIGURE 1. Posterior density of p's.

6.1. Compound Poisson process when K = 3

When the number of hidden states is 3, the likelihood function for the compound Poisson process is given by the following equation:

$$\mathcal{L}(\rho, \mathbf{P}, \lambda, \mathbf{M}; \mathcal{F}, \mathcal{D}) \propto \rho_1^{r_{12}+r_{13}} e^{-\rho_1\tau_1} \rho_2^{r_{21}+r_{23}} e^{-\rho_2\tau_2} \rho_3^{r_{31}+r_{32}} e^{-\rho_3\tau_3} \lambda_1^{n_1} e^{-\lambda_1\tau_1} \lambda_2^{n_2} e^{-\lambda_2\tau_2} \lambda_3^{n_3}$$
$$\times e^{-\lambda_3\tau_3} P_{12}^{r_{12}} P_{13}^{r_{13}} P_{21}^{r_{21}} P_{23}^{r_{23}} P_{31}^{r_{31}} P_{32}^{r_{32}} \prod_{y-x\in F} [(p_1^{y-x}(1-p_1)^{n-y+x})^{\hat{r}_1(y-x)}$$
$$\times (p_2^{y-x}(1-p_2)^{n-y+x})^{\hat{r}_2(y-x)} (p_3^{y-x}(1-p_3)^{n-y+x})^{\hat{r}_3(y-x)}]$$

where $\hat{r}_i(y-x)$ represents the number of outages that affect $(y-x)$ people when the system is in environment i and n_i represents the number of outages when the system is in environment i.

We generated our simulated data using the following parameters:

$$\rho_1 = 0.001, \quad \rho_2 = 0.1, \quad \rho_3 = 0.01, \quad \lambda_1 = 0.025, \quad \lambda_2 = 2.5, \quad \lambda_3 = 0.15,$$
$$n = 10, \quad p_1 = 0.33, \quad p_2 = 0.5, \quad p_3 = 0.67$$

In our Bayesian analysis, we use proper but diffused priors for the parameters by setting $a_i^{\lambda} = b_i^{\lambda} = a_i^{\rho} = b_i^{\rho} = a_i^{p} = b_i^{p} = \alpha_{i,j} = 0.01$ for $i,j = 1,2,3$. Running Gibbs sampler, after a small burn-in sample, we obtained the posterior distribution of the parameters. The simulation of 11,000 iterations with a two-state modulating process took 112 min using a computer with 16 GB RAM, i7-8565U CPU, and 64-bit operating system. Whereas, the same simulation took 168 min using the same device when a three-state modulating process is used instead. As the number of hidden states increases, the number of parameters in the likelihood function increase significantly, along with the computation time.

The posterior distributions of the parameters are presented in Figures 1–3. As can be seen from Figure 1, the posterior distributions of p for state 1, 2 and 3 are concentrated around 0.33, 0.47 and 0.59, respectively. We can see from Figure 2 that the posterior densities of λ are concentrated around 0.03, 1.31 and 0.12, respectively. Figure 3 shows the posterior distributions of ρ to be around 0.0005, 0.12 and 0.003 respectively. When the holding rates of different environments are significantly different from each other, as in our simulated data, we do not witness any convergence problems. However, when the holding rates are close to each other, identifying the states correctly can become problematic. This issue is referred to in the literature as the "label switching" problem. An alternative model could be used to overcome this problem by imposing ordering restrictions on the parameters as discussed in [23].

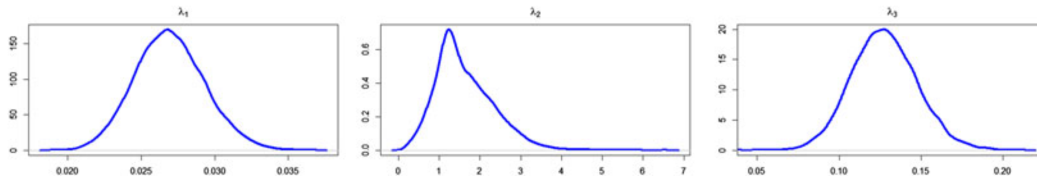

FIGURE 2. Posterior density of λ's.

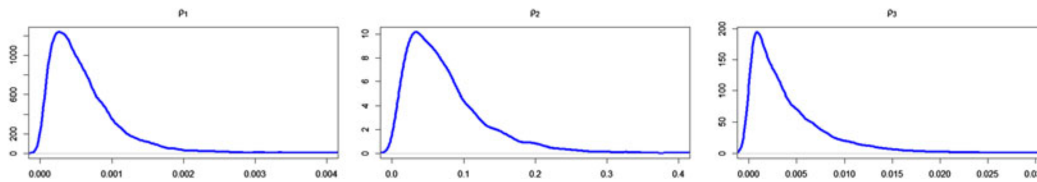

FIGURE 3. Posterior density of ρ's.

For this example, the full conditionals take the following form

$$\mathbf{P}_i \,|\, \mathbf{P}_i^{-1} \sim \mathcal{D}ir(\alpha_{i1} + r_{i1}, \alpha_{i2} + r_{i2}, \alpha_{i3} + r_{i3})$$

$$p_i \,|\, p_i^{-1} \sim \mathcal{B}eta\left(a_i^p + \sum_{y-x\in F} (y-x)\cdot \hat{r}_i(y-x), b_i^p + \sum_{y-x\in F} (n-y+x)\cdot \hat{r}_i(y-x)\right)$$

$$\lambda_i \,|\, \lambda_i^{-1} \sim \mathcal{G}(a_i^\lambda + n_i, b_i^\lambda + \tau_i)$$

$$\rho_i \,|\, \rho_i^{-1} \sim \mathcal{G}\left(a_i^\rho + \sum_{j\in E} r_{ij}, b_i^\rho + \tau_i\right)$$

where $i = 1, 2, 3$ and $\alpha_{ii} = 0$.

Figure 4 shows the actual and the expected time between outages using the holding rate of the environment with the highest posterior probability. The red lines are the expected time between outages when the posterior probability is the highest for state 1, the green lines are the expected time when the posterior probability is the highest for state 2 and the purple line is for state 3. It is visible from the plot that based on the environment, the time between outages vary. State 2 is the most prone to outages with the average time between outages near 0, indicating that outages are nearly constant.

Because we use simulated data, we can identify the state of the environmental process at any time, even though it is typically latent. The confusion matrix for the hidden process shows that out of 253 outages, 132 of them occurred when the environmental process was in state 1, 65 outages occurred when the process was in state 2 and the remaining 56 outages occurred when it was in state 3 (Table 1). The overall hidden state prediction accuracy, based on the highest posterior probability for the latent states, is 0.9723. When the process is in state 2, the model identifies the environmental state correctly for all 65 occurrences. When the process is in state 1, the model correctly predicts the state for 129 out of 132 outages, whereas when the environment is in state 3, the model predicts it correctly for 51 out of 56 occurrences.

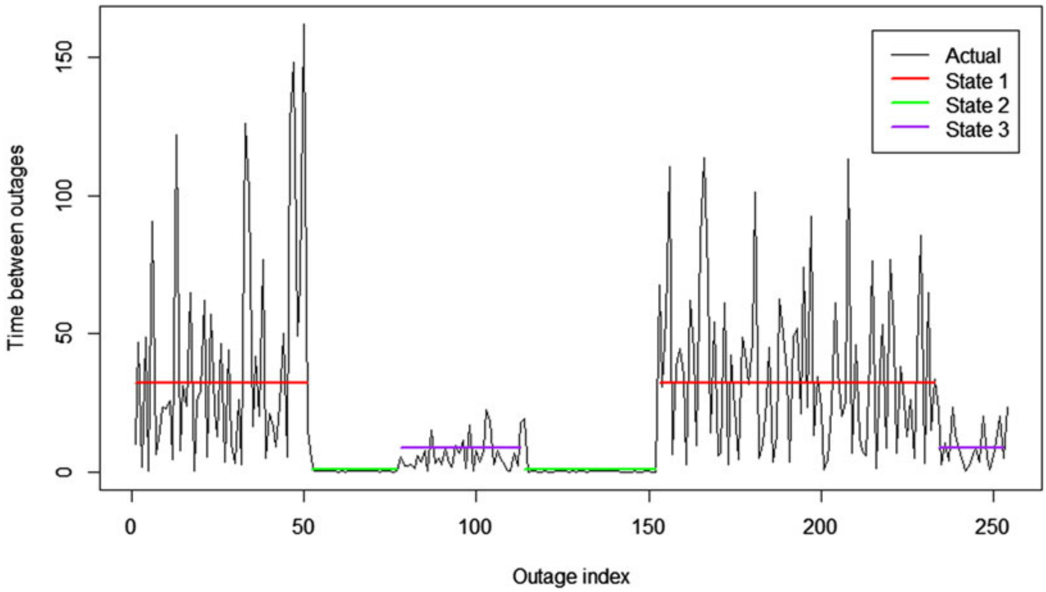

FIGURE 4. Actual and expected time between failures based on the environmental state with the highest posterior probability.

TABLE 1. Confusion matrix for hidden states

		Predicted			
		State 1	State 2	State 3	Total
Actual	State 1	129	0	3	132
	State 2	0	65	0	65
	State 3	5	0	51	56
	Total	134	65	54	253

6.2. Compound Poisson process on power outages data

The power outage data we have include the time of the power outages in Stafford County, Virginia, from September 10, 2012 to December 11, 2013, as well as the number of customers affected by each outage. The data consist of 316 power outages. The number of customers affected by each outage ranges from 0 to 918. The number of customers in Stafford County that potentially could be affected by the outages is 5037.

Our results show that the power system we observe operates in two different environments with different outage rates and probability characteristics. We can see from Figure 5 that the posterior distribution of p_1 is more diffuse than p_2, and it is concentrated around 0.011, whereas the posterior distribution of p_2 is concentrated around 0.015. Figure 6 suggests that state 2 is more prone to outages because the posterior distributions of λ are concentrated around 0.012 and 0.1, for state 1 and state 2, respectively. We observe from Figure 7 that the posterior distribution of ρ_2 is concentrated around 0.0125 which means that the expected time for the environment to stay in state 2 is around 80 h. In Figure 8, the black line is the actual time between outages (in hours), red lines are the expected time between outages when the posterior probability favors state 1 and the green lines are the expected time between outages when the posterior probability favors state 2. It is also

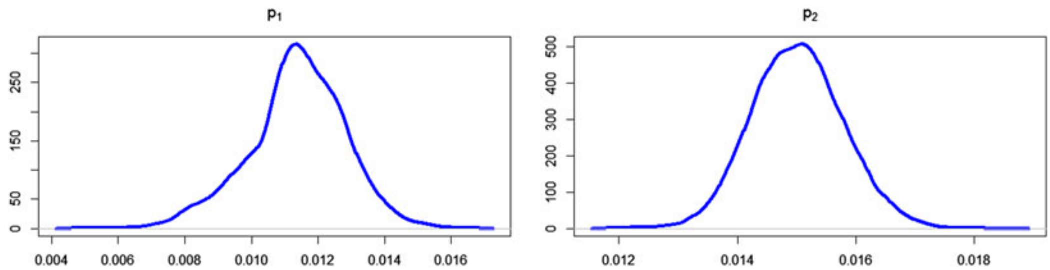

FIGURE 5. Posterior density of p's.

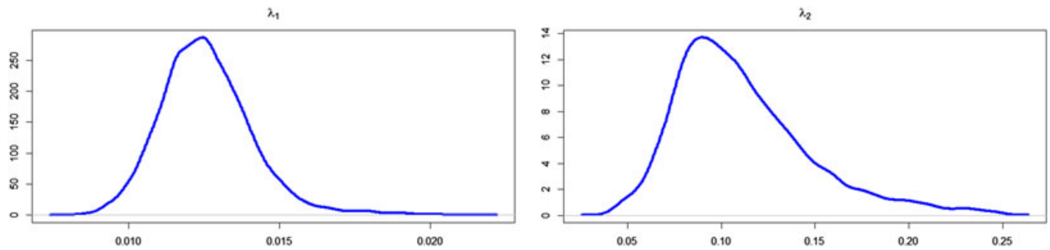

FIGURE 6. Posterior density of λ's.

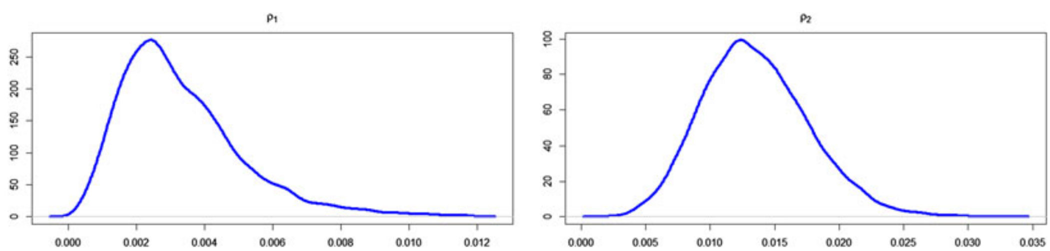

FIGURE 7. Posterior density of ρ's.

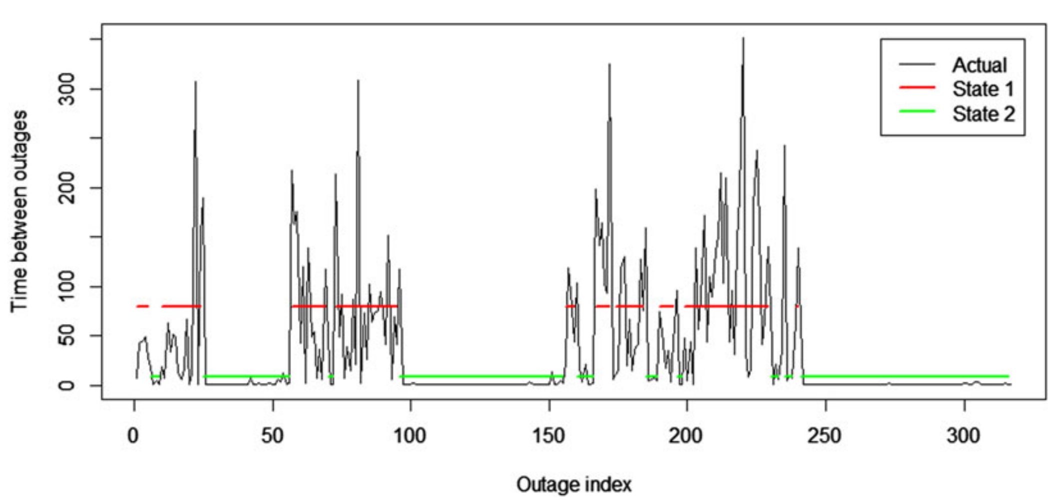

FIGURE 8. Actual and expected time between outages based on the environmental state with the highest posterior probability.

https://doi.org/10.1017/S0269964820000157 Published online by Cambridge University Press

TABLE 2. Outage index and date intervals where the model predicts the environment is in state 2

Outage index	Date of occurrence
25–56	October 29–31, 2012
96–155	March 6–8, 2013
241–316	December 8–11, 2013

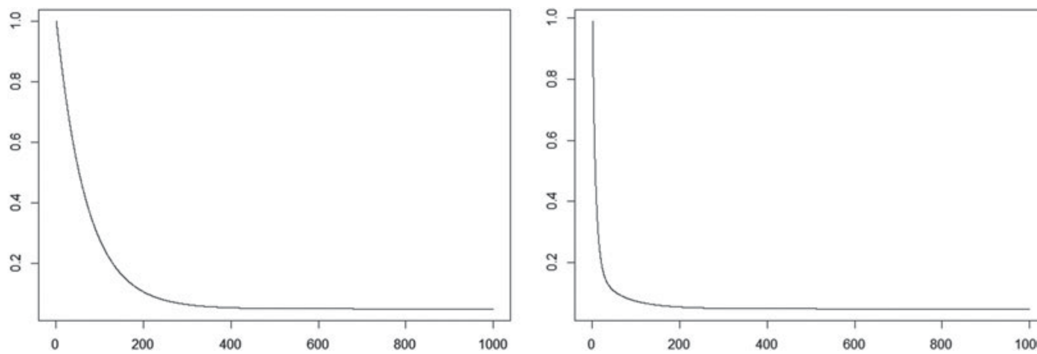

FIGURE 9. Posterior predictive reliability.

important to mention that although the environmental process switches nine times during the observation period, the posterior probabilities for environmental states show that for some of those state changes, the model only slightly favors one state over the other, with the state probabilities close to 0.5. Table 2 gives the outage intervals and dates where we see the model is certain the environment is in state 2.

After using a modulated model, obtaining posterior probabilities for the environmental states, and comparing the environment change times with the National Centers for Environmental Information (NCEI) storm events database, we see that the environmental process captures the effects of extreme weather conditions. During the time we observe the power system, NCEI database reports three major weather events in Stafford County that could have had an impact on the power system. Those reported events are "high wind" on October 29, 2012 and "winter storms" on March 5, 2013 and on December 8, 2013. Our model shows that during, or immediately after the time of those weather events, the environment was in state 2.

It is possible to obtain the posterior probability of time to the next power outage, that is, the posterior system reliability using (A.6) in the appendix and the Monte Carlo average approximation using the posterior sample of parameters ρ and λ. Posterior predictive reliability gives us the probability that there will not be an outage in the next t period of time, conditioned on the current state of the environment, Y_0, written as follows:

$$P(T_1 > t \,|\, Y_0{=}i, \mathcal{D}) \approx \frac{1}{S}\sum_{s=1}^{S}\sum_{j\in E}\exp((G^S - \Lambda^S)t)_{ij}.$$

Figure 9 presents the posterior reliability function of the power system for $Y_0 = 1$ on the left and $Y_0 = 2$ on the right. Because the failure rate is higher for the second environment, the posterior predictive reliability goes to 0 much faster for $Y_0 = 2$ compared to $Y_0 = 1$.

Using the posterior probabilities for the environmental states and the posterior sample of the parameter p, one can obtain the posterior probability distribution of the number of

https://doi.org/10.1017/S0269964820000157 Published online by Cambridge University Press

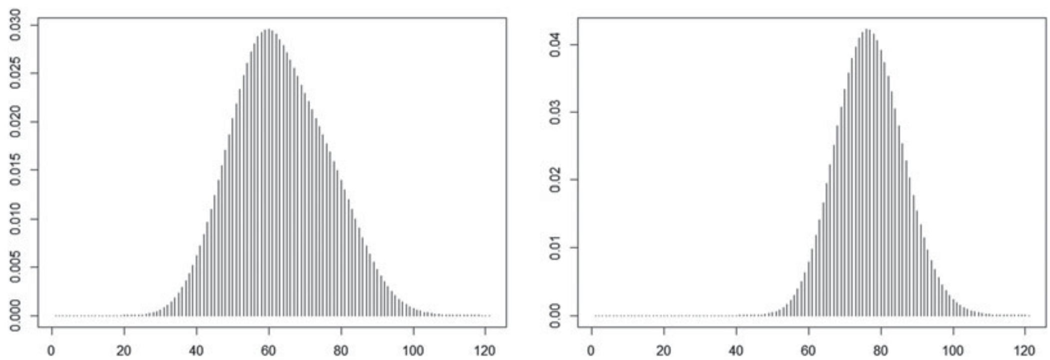

FIGURE **10.** Posterior probabilities for the number of people affected by the outages 17 and 294.

people affected by a specific outage using a mixture of binomial distributions. We selected two outages, index 17 and index 294, to present how the posterior probability distribution for the number of people affected by the outages varies based on if the model predicts the environment to be in state 1 versus state 2. Potentially, this can also be used to make predictions for a future outage. The actual number of people affected by the outages 17 and 294 are 58 and 85, respectively. The model is confident that outage 17 occurred in the first environment and the outage 294 occurred in the second environment. As a result, the expected number of people that will be affected by outage 294 is higher than the number for outage 17 (Figure 10).

7. CONCLUDING REMARKS

In this paper, we have considered Markov modulated Markov processes and developed their Bayesian analysis. We illustrated the implementation of our approach using a Markov modulated compound Poisson process model and applying it to simulated, as well as actual, power outage data to describe the reliability of the power system and the number of people that were affected by the outages.

As the number of parameters in the model increases, the computation time increases dramatically. Therefore, it is important to select the number of hidden states carefully in order not to make the model computationally inefficient. We presented an approach to obtain the marginal likelihood of the data for different dimensions of the hidden process that enables us to infer the unknown number of states. This approach allows us to restrict the number of hidden states and maintain computational efficiency.

This work can be extended in many different ways. In this paper, we only considered the power outages in Stafford County, Virginia. A possible extension would be to widen the geographical range, which would allow us to consider different characteristics of the random environment. An alternative extension can be introducing a spatial component to the model to capture the interaction of different neighborhoods under a changing environment. The proposed approach for analysis of the compound Poisson process has potential use in the software reliability as discussed in Şahinoğlu [27]. Furthermore, our Bayesian approach for modulated Markov processes can be used to develop Bayesian maintenance policies in problems considered in Ahmadi and Fouladirad [1] and van der Weide and Pandley [29].

References

1. Ahmed, R. & Fouladirad, M. (2017). Maintenance planning for a deteriorating production process. *Reliability Engineering & System Safety* 159: 108–118.
2. Arifoğlu, K. & Özekici, S. (2010). Optimal policies for inventory systems with finite capacity and partially observed Markov-modulated demand and supply processes. *European Journal of Operational Research* 204: 421–483.
3. Asmussen, S. (2000). Matrix-analytic models and their analysis. *Scandinavian Journal of Statistics* 27: 193–226.
4. Çanakoğlu, E. & Özekici, S. (2010). Portfolio selection in stochastic markets with HARA utility functions. *European Journal of Operational Research* 201: 520–536.
5. Çekyay, B. & Özekici, S. (2010). Mean time to failure and availability of semi-Markov missions with maximal repair. *European Journal of Operational Research* 207: 1442–1454.
6. Çekyay, B. & Özekici, S. (2012). Optimal maintenance of systems with Markovian mission and deterioration. *European Journal of Operational Research* 219: 123–133.
7. Çekyay, B. & Özekici, S. (2012). Performance measures for systems with Markovian missions and aging. *IEEE Transactions on Reliability* 61: 769–778.
8. Çekyay, B. & Özekici, S. (2015). Optimal maintenance of semi-Markov missions. *Probability in the Engineering and Informational Sciences* 29: 77–98.
9. Çekyay, B. & Özekici, S. (2015). Reliability, MTTF and steady-state availability analysis of systems with exponential lifetimes. *Applied Mathematical Modelling* 39: 284–296.
10. Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 90: 1313–1321.
11. Çınlar, E. & Özekici, S. (1987). Reliability of complex devices in random environments. *Probability in the Engineering and Informational Sciences* 1: 97–115.
12. Çınlar, E., Shaked, M., & Shanthikumar, J.G. (1989). On lifetimes influenced by a common environment. *Stochastic Processes and Their Applications* 33: 347–359.
13. Eisen, M. & Tainiter, M. (1963). Stochastic variations in queuing processes. *Operations Research* 11: 922–927.
14. Fearnhead, P. & Sherlock, C. (2006). An exact Gibbs sampler for the Markov-modulated Poisson process. *Journal of the Royal Statistical Society: Series B* 68: 767–784.
15. Kass, R.E. & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association* 90: 773–795.
16. Landon, J., Özekici, S., & Soyer, R. (2013). A Markov modulated Poisson model for software reliability. *European Journal of Operational Research* 229: 404–410.
17. Lefèvre, C. & Milhaud, X. (1990). On the association of the lifelenghts of components subjected to a stochastic environment. *Advances in Applied Probability* 22: 961–964.
18. Moler, C. & van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review* 45: 3–49.
19. Neuts, M.F. (1974). A queue subject to extraneous phase changes. *Advances in Applied Probability* 3: 78–119.
20. Neuts, M.F. (1994). *Matrix-geometric solutions in stochastic models, an algorithic approach*. New York: Dover Publications, Inc.
21. Özekici, S. & Soyer, R. (2003). Reliability of software with an operational profile. *European Journal of Operational Research* 149: 459–474.
22. Özekici, S. & Soyer, R. (2006). Semi-Markov modulated Poisson process: probabilistic and statistical analysis. *Mathematical Methods of Operations Research* 64: 125–144.
23. Pievatolo, A., Ruggeri, F., & Soyer, R. (2012). A Bayesian hidden Markov model for imperfect debugging. *Reliability Engineering & System Safety* 103: 11–21.
24. Prabhu, N.U. & Zhu, Y. (1989). Markov-modulated queueing systems. *Queueing Systems* 5: 215–246.
25. Purdue, P. (1974). The M/M/1 queue in a Markovian environment. *Operations Research* 22: 562–569.
26. Ross, S. (1996). *Stochastic processes*, 2nd ed. New York: Wiley.
27. Şahinoğlu, M. (1992). Compound-poisson software reliability model. *IEEE Transactions on Software Engineering* 18: 624–630.
28. Singpurwalla, N.D. (2006). The hazard potential: introduction and overview. *Journal of the American Statistical Association* 101: 1705–1717.
29. van der Weide, J.A.M. & Pandley, M.D. (2011). Stochastic analysis of shock process and modeling of condition-based maintenance. *Reliability Engineering & System Safety* 96: 619–626.
30. Zhu, Y. (1994). Markovian queueing networks in a random environment. *Operations Research Letters* 15: 11–17.

APPENDIX A

This Appendix presents derivations and explanations of some of the results used throughout this paper.

APPENDIX A.1. THE MODEL

Suppose that Y is a continuous-time Markov process with a finite state space $E = \{1, 2, \ldots, K\}$, where K is the number of states and generator

$$G_{ij} = \begin{cases} -\rho_i, & \text{if } j = i \\ \rho_i P_{ij}, & \text{if } j \neq i \end{cases}$$

or $G_{ij} = \rho_i(P_{ij} - I_{ij})$, where I is the identity matrix. In other words, the process Y spends an exponential amount of time with holding rate ρ_i in state i and, when it jumps, it randomly goes to state j with transition probability P_{ij}, where $P_{ii} = 0$.

Suppose that while Y is in state i, an event occurs exponentially with some rate λ_i. Letting T denotes the time at which this event occurs, the probabilistic structure of T, conditional on Y, satisfies

$$P[T > t \,|\, Y_u; u \leq t] = \exp\left(-\int_0^t \lambda_{Y_s} ds\right).$$

We are interested in determining probability distributions involving T and Y_T. The interpretation of the event time depends on the specific application. It may be the time of arrival of a customer, time of failure of a machine or time of jump of another process that is modulated by the Markov process Y .

We define the diagonal matrix

$$\Lambda_{ij} = \begin{cases} \lambda_i, & \text{if } j = i \\ 0, & \text{if } j \neq i. \end{cases} \tag{A.1}$$

Note that the diagonal entries of the matrix Λ can also be written as the vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_K)$ using a compact notation.

We now define a new process

$$\hat{Y}_t = \begin{cases} Y_t, & \text{if } t < T \\ \Delta, & \text{if } t \geq T. \end{cases} \tag{A.2}$$

so that the process Y is dumped to state Δ as soon as the event happens at time T. It is clear that the process $\hat{Y}$ is also a Markov process with extended state space $E \cup \{\Delta\}$ and generator

$$\hat{G} = \begin{bmatrix} G - \Lambda & \lambda \\ 0 & 0 \end{bmatrix} \tag{A.3}$$

where the last row and column correspond to state Δ. It also follows that the first passage time

$$T = \inf\{t \geq 0;\ \hat{Y}_t = \Delta\} \tag{A.4}$$

has a phase-type distribution. Moreover, the transition function of $\hat{Y}$ is given by the matrix exponential

$$\exp\left(\hat{G}t\right) = \begin{bmatrix} \exp\left(\left(G - \Lambda\right) t\right) & \left[G - \Lambda\right]^{-1} \left[\exp\left(\left(G - \Lambda\right) t\right) - I\right] \lambda \\ 0 & 1 \end{bmatrix} \tag{A.5}$$

We refer our interested readers to Neuts [20] and Asmussen [3] for details and various results on the exponential matrix and phase-type distributions.

Our construction of the $\hat{Y}$ process also implies that

$$P[Y_t = j, T > t \,|\, Y_0{=}i] {=} P[\hat{Y}_t{=}j \,|\, \hat{Y}_0{=}i] {=} \left[\exp\left(\left(G - \Lambda\right) t\right)\right]_{ij} \tag{A.6}$$

from (A.5).

APPENDIX A.2. THE STATE AT ABSORPTION

We are firstly interested in the state Y_T at which the event happens or when $\hat{Y}$ is absorbed. Letting

$$S = \inf\{t \geq 0;\ Y_t \neq Y_0\}$$

denotes the first time that the Markov process jumps, we can write

$$\begin{aligned} F_{ij} = P_i[Y_T = j] &= P_i[Y_T = j, S > T] + P_i[Y_T{=}j,\ S \leq T] \\ &= \left(\frac{\lambda_i}{\lambda_i + \rho_i}\right) I_{ij} + \int_0^{+\infty} \rho_i e^{-\rho_i s} e^{-\lambda_i s} ds \sum_{k\in E} P_{ik} F_{kj} \\ &= \left(\frac{\lambda_i}{\lambda_i + \rho_i}\right) I_{ij} + \left(\frac{\rho_i}{\lambda_i + \rho_i}\right) \sum_{k\in E} P_{ik} F_{kj} \end{aligned} \quad \textbf{(A.7)}$$

using the Markov property and the fact that S has an exponential distribution with rate ρ_i if the initial state is i. But (A.7) can be rewritten as follows:

$$\sum_{k\in E} \rho_i P_{ik} F_{kj} - (\lambda_i + \rho_i)\, F_{ij} = -\lambda_i I_{ij}$$

or, equivalently,

$$(G - \Lambda)\, F = -\Lambda \quad \textbf{(A.8)}$$

using the compact matrix notation and the solution is

$$F = -\left(G - \Lambda\right)^{-1} \Lambda. \quad \textbf{(A.9)}$$

APPENDIX A.3. THE TIME AT ABSORPTION

The distribution of T can be found easily from

$$\begin{aligned} P_i[T \leq t] = 1 - P_i[T > t] &= 1 - P_i[\hat{Y}_t \in E] \\ &= 1 - \sum_{k\in E} \left[\exp\left((G - \Lambda)t\right)\right]_{ik} \\ &= 1 - \left[\exp\left((G - \Lambda)t\right) 1\right]_i \end{aligned} \quad \textbf{(A.10)}$$

where 1 denotes a column vector of 1's. One can easily show that $d\exp(At)/dt = \exp(At)A = A\exp(At)$ for any generator A and this gives the density function

$$\begin{aligned} \frac{dP_i[T \leq t]}{dt} &= -\left[\exp\left((G - \Lambda)t\right)(G - \Lambda)1\right]_i \\ &= \left[\exp\left((G - \Lambda)t\right)\lambda\right]_i \end{aligned} \quad \textbf{(A.11)}$$

since $(G - \Lambda)1 = -\lambda$.

APPENDIX A.4. THE STATE AND TIME AT ABSORPTION

Our interest lies more in the joint distribution of Y_T and T. Note that, using the Markov property at S, we have

$$
\begin{aligned}
F_{ij}(t) = P_i[Y_T = j,\ T > t] = P_i[Y_T{=}j,\ T > t,\ S > t] + P_i[Y_T{=}j,\ T > t,\ S \le t] \\
= I_{ij} P_i[T > t,\ S > t] + \int_0^t \rho_i e^{-\rho_i s} e^{-\lambda_i s} ds \sum_{k\in E} P_{ik} F_{kj}(t-s) \\
= e^{-(\lambda_i+\rho_i)t} I_{ij} + \int_0^t \rho_i e^{-\rho_i s} e^{-\lambda_i s} ds \sum_{k\in E} P_{ik} F_{kj}(t-s) \\
= e^{-(\lambda_i+\rho_i)t} \left[I_{ij} + \int_0^t e^{(\lambda_i+\rho_i)u} du \sum_{k\in E} \rho_i P_{ik} F_{kj}(u) \right]
\end{aligned}
$$

after the change of variables $u = t - s$. This implies that the derivative satisfies

$$
\begin{aligned}
\frac{dF_{ij}(t)}{dt} &= -\left(\lambda_i + \rho_i\right) F_{ij}(t) + \sum_{k\in E} \rho_i P_{ik} F_{kj}(t) \\
&= \sum_{k\in E} \left(G - \Lambda\right)_{ik} F_{kj}(t)
\end{aligned}
$$

and the solution is the matrix exponential

$$
F(t) = \exp\left((G-\Lambda)t\right) F(0)
$$

where $F(0)$ gives the boundary condition. But note from (A.9) that $F(0) = -(G-\Lambda)^{-1}\Lambda$ and

$$
F(t) = -\exp\left((G-\Lambda)t\right)\left(G-\Lambda\right)^{-1}\Lambda. \tag{A.12}
$$

Moreover, since

$$
\bar{F}_{ij}(t) = P_i[Y_T = j,\ T \le t] = P_i[Y_T = j] - P_i[Y_T = j,\ T > t]
$$

it follows from (A.9) and (A.12) that

$$
\begin{aligned}
\bar{F}(t) &= -\left(G-\Lambda\right)^{-1}\Lambda + \exp\left((G-\Lambda)t\right)\left(G-\Lambda\right)^{-1}\Lambda \\
&= \left[\exp\left((G-\Lambda)t\right) - I\right]\left(G-\Lambda\right)^{-1}\Lambda.
\end{aligned} \tag{A.13}
$$

This leads to the density

$$
\begin{aligned}
\frac{d\bar{F}(t)}{dt} &= (G-\Lambda)\exp\left((G-\Lambda)t\right)\left(G-\Lambda\right)^{-1}\Lambda \\
&= \exp\left((G-\Lambda)t\right)\Lambda
\end{aligned} \tag{A.14}
$$

since $A\exp(At) = \exp(At)A$. In open form, we can, therefore, write

$$
\frac{dP_i[Y_T = j, T \le t]}{dt} = \left[\exp\left((G-\Lambda)t\right)\Lambda\right]_{ij} = \left[\exp\left((G-\Lambda)t\right)\right]_{ij}\lambda_j. \tag{A.15}
$$

https://doi.org/10.1017/S0269964820000157 Published online by Cambridge University Press