Scandinavian Journal of Statistics

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Bayesian Estimation of Transition Rates in Two-State Nonhomogeneous Markov Jump Processes With Intermittent Observations: An Honest-Time Data-Augmentation Approach

Dario Gasbarra[1] | Sangita Kulathinal[2] | Etienne Sebag[2]

[1]Department of Mathematics and Statistics, University of Vaasa, Vaasa, Finland | [2]Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

Correspondence: Etienne Sebag (etienne.sebag@helsinki.fi)

Received: 25 April 2025 | **Revised:** 22 January 2026 | **Accepted:** 3 July 2026

Keywords: Bayesian data augmentation | Bayesian estimation | Gibbs sampling | honest random time | intermittent observations | Metropolis-Hastings algorithm | nonhomogeneous Markov processes | Poisson point processes | transition intensity and probability

ABSTRACT

A possibly time-dependent transition intensity matrix or generator ($Q(t)$) characterizes the law of a Markov jump process (MP). For a time-homogeneous MP, the transition probability matrix (TPM) can be expressed as a matrix exponential of Q. However, when dealing with a time nonhomogeneous MP, there is often no simple analytical form of the TPM in terms of $Q(t)$, unless all the $Q(t)$ commute. This poses a challenge because when a continuous MP is observed intermittently, a TPM is required to build a likelihood. In this paper, we show that the Bayesian estimation of the transition intensities of a two-state nonhomogeneous Markov model can be carried out by augmenting the intermittent observations with honest random times associated with two independent driving Poisson point processes, and that sampling the full path is not required. We thus propose a Bayesian data augmentation algorithm wherein the observations are augmented to include honest times, thus facilitating the sampling of the model parameters. Finally, we illustrate our approach by simulating a continuous MP and by using observed (intermittent) time grids extracted from real clinical visits data.

1 | Introduction

Multistate models are commonly employed in many medical applications to study diseases within a cohort, whereby different states describe specific statuses, or markers, of a disease. In this framework, researchers are interested in the state occupancy times (how long individuals remain in each state) and the exact transition times (when individuals move between states). This information is essential to better understand and track disease progression or remission, as well as to identify how individuals respond to a certain treatment, for example. Nonetheless, continuous-time state observations are rarely available in practice because the monitored subjects with the disease are only observed at fixed points in time corresponding to clinical visits. The setup naturally lends itself to that of a nonhomogeneous Markov process, characterized by movements between states whose properties depend on time. The process, $\{X(t), t \geq 0\}$, is observed at intermittent, discrete time-points for each individual. Therefore, clinicians/practitioners only have access to snapshots in time concerning the underlying process instead of having the full trajectory.

The underlying dynamics of the process are often captured in terms of the transition intensities, and the transition probabilities are then derived from the intensities by solving Kolmogorov forward equations. From a statistical point of view, when considering such a process, the interest is in the estimation of these parameters, as well as in the prediction of the latent path sequence at unobserved times. When we know the exact state transition times, then the likelihood is written in terms of the transition intensities, and is tractable and easy to work with. Yet in the more common cases that the occupied states are only known at specific and intermittent time points, the likelihood must be constructed by considering the transition probabilities over intervals. Since the states are only known at specific times, for example, clinical visits, say t_1 and t_2, and what happened between t_1 and t_2 is unknown, one must rely exclusively on the transition probabilities $P(X(t_2) = j|X(t_1) = i)$ to formulate the likelihood of transitioning between states over the unobserved time interval. The overarching problem of using these transition probabilities in complex multistate models is that, often, the Kolmogorov forward equations do not have nice closed-form analytical solutions, which can consequently impose a substantial numerical and computational burden under practical implementations. Various workarounds exist in the statistical literature: approximating the transition intensities by piecewise constant intensities and then applying the matrix exponential, using a smooth function of time, or using a time transform on which the process becomes homogeneous (Hubbard et al. 2007; Kendall et al. 2024; Titman 2011). Some of the methods are made available in the `nhm` R-package (Titman 2019). In general, these approaches have limited applicability.

In this paper, we propose a Markov Chain Monte Carlo (MCMC) approach which consists of a novel Bayesian data augmentation procedure using Poisson point processes (PPP) to eliminate the need altogether to handle these computations. In order to illustrate this methodology, we will focus on a simple two-state reversible model to capture the dynamics of disease in which subjects transition back and forth between two distinct health states. This is illustrated in Figure 1 and examples can be found from (Cook and Lawless 2018; Mehtäälä et al. 2015). It is of statistical interest to estimate the transition intensities $\alpha_0(t)$ and $\alpha_1(t)$, which will be formally defined in Section 2.2, for this model. Two-state disease models often deal with binary clinical hard outcomes, which are directly measurable at each clinical visit. For example, a hard clinical outcome at each visit might be the presence (state 1) or absence (state 0) of a disease marker, which subsequently can be used to allow for broader categorizations such as "infected" vs "not-infected", or "disease onset" vs "no disease". This disease marker may fluctuate at any point in time. Moreover, we assume that each subject's disease status is fully captured by one of these two states at any given point in time. Treatment for diseases may or may not be administered at each clinical visit that aims to influence these transitions. For instance, if a subject stays in disease state 1 for an extended period of time, then we can speculate that this subject is exhibiting a poor response to the treatment.

It is important to highlight that our developed framework avoids the need to simulate the full trajectory of the process. The key idea is intuitive, and it relies on the fact that we have constructed both the last state and the time at which the process occupies

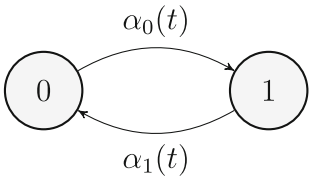


FIGURE 1 | A two-state model describing a typical process of a disease marker with transition intensities $\alpha_0(t)$ and $\alpha_1(t)$.

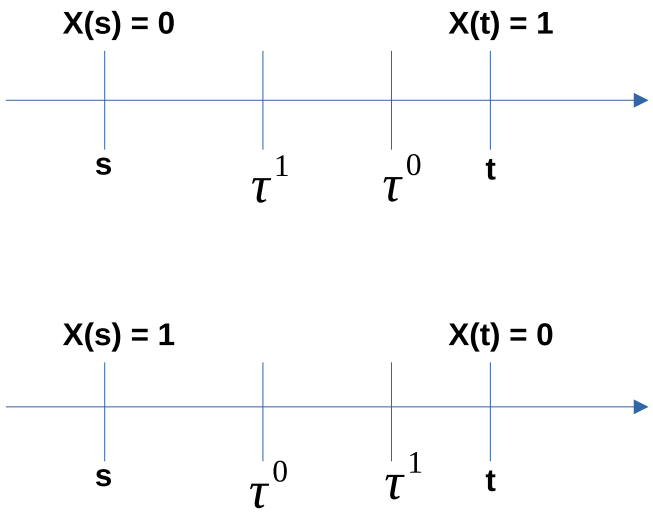


FIGURE 2 | Illustration of honest times of the two independent Poisson point processes over the interval $[s, t)$ for a two-state Markov process. Two other cases: $(X(s) = 1, X(t) = 1)$ and $(X(s) = 0, X(t) = 0)$ are similar to the upper and lower paths, respectively, with the exception that the two honest times can be equal. See Appendices B and C for further explanation.

that final state within the interval $[s, t)$, given the observed states $(X(s), X(t))$ at times $s < t$. To aid in the understanding and visualization, we can refer to Figure 2: along the interval $[s, t)$, the random times τ^0 and τ^1 mark the last possible transition opportunities generated by two independent PPPs with intensity $\alpha_0(t)$ and $\alpha_1(t)$ respectively. For concreteness, we define τ^k to be the last event time in $(s, t]$ of an associated independent PPP with intensity $\alpha_k(t)$, corresponding to a potential transition $k \to 1 - k$, for $k \in \{0, 1\}$. The last state, as discussed above, then depends only on which of these occurs later. This interpretation naturally introduces the concept of honest times, which serve as the foundation of our methodology.

- If $\tau^0 > \tau^1$, the last transition is $0 \to 1$, then the process ends in state 1 (upper panel of Figure 2).
- If $\tau^1 > \tau^0$, the last transition is $1 \to 0$, then the process ends up in state 0 (lower panel of Figure 2).

Nowadays, there exist several statistical packages in the `R` programming language to assist with statistical inference of multistate models, namely the `msm` package (Jackson 2011), `nhm`, and `mstate` packages. The `nhm` package, which builds on `msm`, is tailored to estimating transition rates from panel data. Nevertheless, to our knowledge, the computations fail under `nhm` when the observed time grids are irregular; that is, when there is

variation in the time grids across subjects. Hence, our proposed methodology aims to address this gap by providing a versatile framework that handles estimation in nonhomogeneous cases for two-state models. The rest of the paper is organized as follows: in Sections 2 and 3 we present the theoretical background in order to explain how the intractable computations in this setting arise, and we motivate the use of PPPs and honest random times. We highlight why the data augmentation procedure involving a PPP is a contextually and mathematically justifiable framework. In Section 4, we develop the specific details of the data augmentation algorithm. Next, in Section 5, we carefully illustrate the steps and computations related to this algorithm by using as an example Weibull (time-dependent) rates. Section 6 depicts the results from our algorithm when applied to a simulated process and when using real-life subject time grids. Both scenarios employ Weibull-distributed rates. We end with a discussion in Section 7.

2 | Background Information and Theoretical Framework

2.1 | Ongoing Research

It is well documented in the existing statistical literature that multistate models offer a very suitable framework for analyzing the progression of biological life processes using a discrete state space, and for additionally modeling the stochastic movement of individuals across these health states. Their use has thus become ubiquitous in the analysis of public health data. Nonetheless, the complex structure of the longitudinal intermittent subject-level trajectories poses significant challenges since observations are sparse and unevenly spread across time. Much of the ongoing research on this topic has actually focused on continuous-time Markov models because these can handle multistate processes that are composed of irregular and sparse time grids when the exact transition time is unknown. There have also been several attempts to formulate the problem as a semi-Markov model, meaning that the state transitions also depend on the sojourn time in a particular state. To illustrate these aforementioned points, Cheung et al. (2022), for instance, used continuous multistate processes applied to cancer data to model the progression of the disease, and they further assessed the impacts of (incorrectly) assuming time-homogeneous processes when, in fact, the transition intensities are time-dependent and also follow the semi-Markov assumption. Throughout their work, they indeed acknowledge that using time-homogeneous Markov models eases the computational issues associated with interval-censored panel data, but that often making this assumption is not contextually appropriate for cancer or disease modelling in general (Cheung et al. 2022). As mentioned in Section 1, approximating the transition intensities (and any time-dependent covariates in the model) by piecewise constant functions is a simplifying assumption that has been proposed in the past to handle intractable likelihoods that arise from panel data. This approximation is implemented by holding covariates constant between successive observation times/visits, and by modelling the baseline intensity functions as piecewise-constant over prespecified intervals (Tom and Farewell 2011). More recently, Machado et al. (2021) proposed nonparametric transition intensity specifications that relied on spline functions. Then, penalized maximum likelihood estimation was used in a setting to estimate the parameters of the multistate model via a novel way to estimate the penalty parameters (Machado et al. 2021). Aralis and Brookmeyer (2019) developed an iterative stochastic expectation-maximization algorithm applied to the semi-Markov case in which the likelihood function could be approximated via rejection-sampling (Aralis and Brookmeyer 2019). Further reading about the mathematical and rigorous background of continuous-time multistate models and how to fit them in `R` using the packages `hesim` and `flexsurv` can be found in Chapter 11 of (Thom et al. 2025).

However, we point out that the majority of the work dedicated to the analysis of continuous-time multistate models has been done under the frequentist paradigm, with Bayesian methods slightly more limited due to computational challenges (Luo et al. 2021; Zingoni et al. 2021). Yet Bayesian approaches can handle the structure of longitudinal data very well by having the flexibility to handle missing data and the ability to incorporate prior information, which is very useful in a public health setting when medical practitioners can incorporate existing beliefs about a disease, for example (Taha et al. 2024). Bayesian approaches to continuous-time multistate models with specific applications can be found in (Williams et al. 2020), (Klausch et al. 2023), (Nevala et al. 2024), and (Luo and Sherlock 2025).

Using a Bayesian framework to address intractable likelihood computations in this setting, Tancredi (2019) employed Approximate Bayesian Computation (ABC) methods to estimate the posterior distributions of the model parameters by comparing and approximately matching summary statistics from simulated and observed datasets under Markov and semi-Markov set-ups across different observation times. The limitation of this approach as an MCMC method is that it fails to yield asymptotically exact inference when the number of iterations tends to infinity because the ABC relies on summary statistics that are not sufficient (Tancredi 2019). Additionally, Barone and Tancredi (2022) later adopted the uniformization algorithm in order to simulate continuous-time, nonhomogeneous Markovian trajectories that are conditional on the end-points and then integrated this approach with a Metropolis-Hastings algorithm for a fully Bayesian approach. Their work also applies to conditioned semi-Markov trajectories (Barone and Tancredi 2022).

It is also worthwhile to mention the work by Luo et al. (2021). They developed a continuous-time hidden Markov model in a fully Bayesian setting with a Metropolis-Hastings-within-Gibbs algorithm. Their method allows the rate matrix of the underlying Markov jump process to vary across sub-intervals defined by the observation endpoints (Luo et al. 2021). To sample latent trajectories, they employ the idea developed by Hobolth and Stone (2009), whereby a complete reconstruction of the sample path can be achieved (recovering the jump-time points) by using modified rejection sampling when it is known that at least one transition occurred, conditioned on the known end state (Hobolth and Stone 2009).

Our approach is similar in flavor to the aforementioned points, except that, to the best of our knowledge, the novelty in our approach lies in the fact that we can exploit a data augmentation procedure on the honest times to precisely avoid the

explicit simulation of full sample paths. Existing methods based on Hobolth and Stone (2009) appear to use conditional path simulation to reconstruct all intermediate jump times, whereas our intuitive framework, as mentioned in Section 1 and shown in Figure 2, directly provides us with an unbiased estimator of the transition probabilities by simulating only the occurrence and time of the last transition in each of the observed time intervals.

2.2 | Transition Intensities and Probabilities

When we only have access to the intermittent observations of the occupied state at clinical visits, the likelihood formulation is in terms of the transition probabilities. We denote by $\{X(t), t \geq 0\}$ the observed process. Furthermore, let $\mathcal{H}(t) = \{X(s), 0 \leq s \leq t\}$ denote the corresponding history (collection of available information) of the process over the time interval $[0, t]$.

Both the transition intensities and the transition probabilities govern the stochastic movement between states, and they exhibit an explicit dependence on the history $\mathcal{H}(t)$. The latter take the general form, for $0 \leq s < t$,

$$P_{ij}(s,t) = P_{ij}(s,t\,|\mathcal{H}(s)) = P(X(t) = j|X(s) = i, \mathcal{H}(s)).$$

Recall also that the transition intensities $\alpha_0(t\,|\mathcal{H}(t-))$ and $\alpha_1(t\,|\mathcal{H}(t-))$ need to be estimated for the model. With $X(t)$ representing the occupied state at time t for an arbitrary individual, we define the transition intensity function $\alpha_0(t\,|\mathcal{H}(t-))$ between states 0 and 1 as follows:

$$\alpha_0(t\,|\mathcal{H}(t-)) = \lim_{\Delta t \downarrow 0} \frac{P(X(t+\Delta t-) = 1\,|X(t-) = 0, \mathcal{H}(t-))}{\Delta t}$$

Interchanging the positions of 0 and 1 in the above equation gives the transition intensity $\alpha_1(t\,|\mathcal{H}(t-))$. Moving forward, for the sake of convenience, we will suppress the dependency on $\mathcal{H}(t)$ in the notation.

The generator matrix $\boldsymbol{Q}(t)$, also called the transition intensity matrix, for a two-state model (Figure 1) is given by

$$\boldsymbol{Q}(t) = \begin{bmatrix} -\alpha_0(t) & \alpha_0(t) \\ \alpha_1(t) & -\alpha_1(t) \end{bmatrix}. \tag{1}$$

The corresponding transition probability matrix is denoted by $\boldsymbol{P}(s,t)$ and has the form

$$\boldsymbol{P}(s,t) = \begin{bmatrix} P_{00}(s,t) & P_{01}(s,t) \\ P_{10}(s,t) & P_{11}(s,t) \end{bmatrix}. \tag{2}$$

Note that the rows of $\boldsymbol{Q}(t)$ sum to zero while the rows of $\boldsymbol{P}(s,t)$ must sum to one. The transition probabilities satisfy the linear ordinary differential equations (ODE) obtained using the Kolmogorov forward equations as

$$\frac{\partial}{\partial t}\boldsymbol{P}(s,t) = \boldsymbol{P}(s,t)\boldsymbol{Q}(t). \tag{3}$$

The general solution to (3) is the noncommutative product integral (Gill and Johansen 1990)

$$\boldsymbol{P}(s,t) = \prod_{(s,t]}(\boldsymbol{I} + \boldsymbol{Q}(r)dr),$$

where $\boldsymbol{I}$ is the 2×2 identity matrix. Only when all $\boldsymbol{Q}(r)$ commute with each other, $\boldsymbol{P}(s,t)$ can be given as the matrix exponential $\boldsymbol{Exp}$ since (3) has a matrix product solution:

$$\boldsymbol{P}(s,t) = \boldsymbol{Exp}\left(\int_s^t \boldsymbol{Q}(r)dr\right).$$

Now, performing the matrix multiplication expansion on the right-hand side of (3), then the (1, 2) and (2, 1) matrix entries, for example, respectively yield

$$\frac{\partial}{\partial t}P_{01}(s,t) = P_{00}(s,t)\alpha_0(t) - P_{01}(s,t)\alpha_1(t),$$
$$\frac{\partial}{\partial t}P_{10}(s,t) = -P_{10}(s,t)\alpha_0(t) + P_{11}(s,t)\alpha_1(t).$$

Solutions to the above ODE's, and use of the fact that the transition probabilities out of a state sum to 1, allows expressing the transition probabilities fully in terms of the transition intensities, with the result (see Appendix A):

$$\begin{aligned}
P_{01}(s,t) &= \int_s^t \exp\left(-\int_v^t \big(\alpha_0(r) + \alpha_1(r)\big)dr\right)\alpha_0(v)dv, \\
P_{00}(s,t) &= 1 - P_{01}(s,t), \\
P_{10}(s,t) &= \int_s^t \exp\left(-\int_v^t \big(\alpha_0(r) + \alpha_1(r)\big)dr\right)\alpha_1(v)dv \\
P_{11}(s,t) &= 1 - P_{10}(s,t) = 1 - \int_s^t \exp\left(-\int_v^t \big(\alpha_0(r) + \alpha_1(r)\big)dr\right) \\
&\quad \times \alpha_1(v)dv \\
&= \exp\left(-\int_s^t \big(\alpha_0(r) + \alpha_1(r)\big)dr\right) \\
&\quad + \int_s^t \exp\left(-\int_v^t \big(\alpha_0(r) + \alpha_1(r)\big)dr\right)\alpha_0(v)dv.
\end{aligned} \tag{4}$$

The last equality follows by making the following observation and then integrating by parts

$$\exp\left(-\int_v^t \alpha_1(r)\,dr\right)\alpha_1(v)\,dv = d\exp\left(-\int_v^t \alpha_1(r)\,dr\right). \tag{5}$$

It also follows by observing that (as pointed out by a reviewer):

$$\begin{aligned}
&\int_s^t \exp\left(-\int_v^t \big(\alpha_0(r) + \alpha_1(r)\big)dr\right)(\alpha_0(v) + \alpha_1(v))dv \\
&= -\left[\exp\left(-\int_v^t \big(\alpha_0(r) + \alpha_1(r)\big)dr\right)\right]_s^t \\
&= -1 + \exp\left(-\int_s^t \big(\alpha_0(r) + \alpha_1(r)\big)dr\right).
\end{aligned}$$

The expression for $P_{01}(s,t)$ can be simplified to some extent as shown below.

$$\begin{aligned}
P_{01}(s,t) &= \int_s^t \exp\left(-\int_v^t \big(\alpha_0(r) + \alpha_1(r)\big)dr\right)\alpha_0(v)dv \\
&= \int_s^t \exp\left(-\int_v^t \alpha_1(r)\,dr\right)\,d\exp\left(-\int_v^t \alpha_0(r)\,dr\right).
\end{aligned} \tag{6}$$

The second equality follows because of the identity similar to equation (5) when $\alpha_1(t)$ is replaced by $\alpha_0(t)$.

As we will see in Section 5, the equations for the transition probabilities simplify nicely in the case of constant transition intensities $\alpha_0(t) = \lambda_0$ and $\alpha_1(t) = \lambda_1$. However, the overarching problem is that, when we deal with time-dependent (nonhomogeneous) transition rates, such as those that are of Weibull-type, then we still cannot compute analytically the integral in (6). Although there may be a possibility of finding some special function to represent the integral, this paper introduces and describes an alternative framework using Poisson point processes (PPP) for Bayesian data augmentation. This approach entirely avoids the need to compute the transition probabilities analytically or by way of splitting the time window.

2.3 | PPP, Honest Times and Data Augmentation

The central idea of this distinct framework is to obtain an unbiased estimator of the transition probabilities, and thereby of the observed data likelihood, under this two-state setting. This is achieved by performing a data augmentation procedure by using a simpler PPP, instead of attempting to numerically solve the integrals for the transition probabilities. Consider an independent PPP $\Pi^1(dt)$ with intensity $\alpha_1(t)$. Then, by noting that the integrand in Equation (6) is the survival probability, we can write integral (6):

$$\begin{aligned} P_{01}(s,t) &= \int_s^t P\big(\Pi^1((v,t]) = 0\big) d \exp\left(-\int_v^t \alpha_0(r)\, dr\right) \\ &= \mathbb{E}_{\Pi^1}\left(\int_s^t \mathbf{1}\left(\Pi^1((v,t]) = 0\right) d \exp\left(-\int_v^t \alpha_0(r) dr\right)\right) \\ &= \mathbb{E}_{\Pi^1}\left(\int_{\tau^1}^t d \exp\left(-\int_v^t \alpha_0(r) dr\right)\right) \\ &= 1 - \mathbb{E}_{\Pi^1}\left(\exp\left(-\int_{\tau^1}^t \alpha_0(r) dr\right)\right), \end{aligned} \tag{7}$$

where $\tau^1 = s \vee \inf\{r : \Pi^1((r,t]) = 0\}$ is an *honest* random time depending only on the PPP Π^1 (see Appendix B). Fubini's theorem has been applied between the first and the second equality operator, interchanging the expectation and the integral sign.

It is useful at this stage to give more context regarding the link between a PPP and the survival probability. From (6), we may notice that the factor $\exp(-\int_v^t \alpha_1(r) dr)$ is the survival probability of a process with a hazard function $\alpha_1(t)$. In survival analysis terminology, this implies that the process remains and "survives" in state 1 during the time interval $(v,t]$, thereby experiencing no jumps of the type $1 \to 0$ governed by $\alpha_1(t)$. Probabilistically, it is interpreted as the probability of observing no points in the interval $(v,t]$ for a nonhomogeneous PPP Π^1, a statement which we may express as $P(\Pi^1((v,t]) = 0)$.

Next, and similarly, we construct an independent PPP with intensity $\alpha_0(t)$ and define $\tau^0 = s \vee \inf\{r : \Pi^0((r,t]) = 0\}$, an honest random time depending only on the PPP Π^0 so that

$$\begin{aligned} P_{10}(s,t) &= \mathbb{E}_{\Pi^0}\left(\int_s^t \mathbf{1}\left(\Pi^0((v,t]) = 0\right) d \exp\left(-\int_v^t \alpha_1(r) dr\right)\right) \\ &= \mathbb{E}_{\Pi^0}\left(\int_{\tau^0}^t d \exp\left(-\int_v^t \alpha_1(r) dr\right)\right) \\ &= 1 - \mathbb{E}_{\Pi^0}\left(\exp\left(-\int_{\tau^0}^t \alpha_1(r) dr\right)\right). \end{aligned} \tag{8}$$

The honest random times τ^1 and τ^0 are independent with respective probability distributions

$$f^{(k)}(u;\theta) = \alpha_k(u)^\eta \exp\left(-\int_u^t \alpha_k(s) ds\right), \mathbf{1}\,(s \le u < t), \tag{9}$$

where θ is the collection of the transition rates and $\eta = \mathbf{1}(u > s)$, $k = 0, 1$.

3 | Unbiased Estimators of the Transition Probabilities

3.1 | Using the PPP Π^1 With Intensity $\alpha_1(t)$

Given a vector of honest random times τ^1 depending only on the PPP Π^1 with intensity $\alpha_1(t)$ and the transition rates, we introduce conditional transition probabilities as:

$$P_{00}^{\tau^1}(s,t) = \exp\left(-\int_{\tau^1}^t \alpha_0(r) dr\right), \; P_{01}^{\tau^1}(s,t) = 1 - P_{00}^{\tau^1}(s,t),$$

$$P_{11}^{\tau^1}(s,t) = \exp\left(-\int_s^t (\alpha_0(r) + \alpha_1(r)) dr\right) + P_{01}^{\tau^1}(s,t),$$

(follows from (4))

$$P_{10}^{\tau^1}(s,t) = 1 - P_{11}^{\tau^1}(s,t). \tag{10}$$

Theorem 1. *The conditional transition probability matrix $P^{\tau^1}(s,t)$, with elements given in (10) and τ^1 being honest random times resulting from the PPP with intensity $\alpha_1(t)$, is an unbiased estimator of $\mathbf{P}(s,t)$ defined in (4). Further, the variance of each element $P_{k\ell}^{\tau^1}(s,t)$ is given by*

$$\mathrm{Var}\left(P_{k\ell}^{\tau^1}(s,t)\right) = \mathbb{P}(\tau_1^0 \vee \tau_2^0 \le \tau^1) - \mathbb{P}(\tau^0 \le \tau^1)^2,$$

where τ^0 is the honest time associated on the interval $[s,t)$ with a PPP with intensity $\alpha_0(t)$, independent of the PPP with intensity $\alpha_1(t)$, and τ_1^0 and τ_2^0 are the realizations of τ^0.

Proof. The unbiasedness of the estimators given in (10) $P_{k\ell}^{\tau^1}(s,t)$ $k, \ell \in \{0,1\}$ follow from (7) so that

$$\mathbb{E}_{\Pi^1}\left[P_{k\ell}^{\tau^1}(s,t)\right] = P_{k\ell}(s,t), \quad k, \ell \in \{0,1\}.$$

It is clear that $\forall\, k, \ell \in \{0,1\}$, the conditional transition probabilities $P_{k\ell}^{\tau^1}(s,t)$ as functions of the honest time τ^1 have the same variance. By conditioning and using the independence of τ^1 and τ^0 we obtain the second moment of the estimator $P_{00}^{\tau^1}(s,t)$ as follows.

$$\begin{aligned} \mathbb{P}\left(\tau_1^0 \vee \tau_2^0 \le \tau^1\right) &= \mathbb{E}\left[\mathbb{P}\left(\tau_1^0 \vee \tau_2^0 \le \tau^1 \middle| \tau^1\right)\right] \\ &= \mathbb{E}\left[\mathbb{P}\left(\tau_1^0 \le \tau^1 \middle| \tau^1\right)\mathbb{P}\left(\tau_2^0 \le \tau^1 \middle| \tau^1\right)\right] \end{aligned}$$

(by independence of τ_1^0 and τ_2^0)

$$= \mathbb{E}\left[\exp\left(-\int_{\tau^1}^{t}\alpha_0(r)dr\right)^2\right] = \mathbb{E}\left[P_{00}^{\tau^1}(s,t)^2\right]$$

while $\mathbb{E}\left[P_{00}^{\tau^1}(s,t)\right] = \mathbb{P}(\tau^0 \le \tau^1)$. Hence the theorem. □

3.2 | Using the PPP Π^0 With Intensity $\alpha_0(t)$

By interchanging the role of 0 and 1 in (10) with honest random times τ^0 depending only on the PPP Π_0 with intensity $\alpha_0(t)$, we obtain alternative unbiased estimators of the transition probabilities:

$$P_{11}^{\tau^0}(s,t) = \exp\left(-\int_{\tau^0}^{t}\alpha_1(r)dr\right),\ P_{10}^{\tau^0}(s,t) = 1 - P_{11}^{\tau^0}(s,t),$$
$$P_{00}^{\tau^0}(s,t) = \exp\left(-\int_{s}^{t}(\alpha_0(r)+\alpha_1(r))dr\right) + P_{10}^{\tau^0}(s,t),$$
$$P_{01}^{\tau^0}(s,t) = 1 - P_{00}^{\tau^0}(s,t).$$

A theorem similar to Theorem 1 can be given in terms of τ^0, and its proof follows the same line of thinking.

Theorem 2. *The conditional transition probability matrix $P^{\tau^0}(s,t)$ is an unbiased estimator of $\mathbf{P}(s,t)$ defined in (4). Further, the variance of each element $P_{k\ell}^{\tau^0}(s,t)$ is given by*

$$\mathrm{Var}(P_{k\ell}^{\tau^0}(s,t)) = \mathbb{P}(\tau_1^1 \vee \tau_2^1 \le \tau^0) - \mathbb{P}(\tau^1 \le \tau^0)^2$$

where τ^1 and τ^0 are the honest times associated in the interval $[s,t)$ with two independent PPPs with rates $\alpha_1(t)$ and $\alpha_0(t)$, respectively, and τ_1^1 and τ_2^1 are two independent realizations of τ^1.

Remark. Given both vectors τ^1 and τ^0, we can provide another unbiased estimator of the conditional transition probability matrix

$$P_{00}^{\tau^1,\tau^0}(s,t) = \exp\left(-\int_{\tau^1}^{t}\alpha_0(r)dr\right), P_{01}^{\tau^1,\tau^0}(s,t) = 1 - P_{00}^{\tau^1,\tau^0}(s,t),$$
$$P_{11}^{\tau^1,\tau^0}(s,t) = \exp\left(-\int_{\tau^0}^{t}\alpha_1(r)dr\right), P_{10}^{\tau^1,\tau^0}(s,t) = 1 - P_{11}^{\tau^1,\tau^0}(s,t).$$

Remark. To reduce the variance, we could also take the average of the unbiased estimators of the transition probabilities with several independent copies of τ^1 or τ^0, possibly choosing the representation with smaller variance.

3.3 | Using (τ^1, τ^0) Pairs

We have two independent PPPs, Π^1 with intensity $\alpha_1(t)$ and Π^0 with intensity $\alpha_0(t)$. τ^1 and τ^0 are as defined in Section 2.3 and as illustrated in Figure 2. We can then write the transition probabilities (see equation (6)) using the pair (τ^1, τ^0) as:

$$P_{01}(s,t) = \int_s^t \exp\left(-\int_v^t \alpha_1(r)dr\right) d\exp\left(-\int_v^t \alpha_0(r)dr\right)$$
$$= 1 - \mathbb{P}_{\Pi^0\otimes\Pi^1}(\tau^0 \le \tau^1) = \mathbb{P}_{\Pi^0\otimes\Pi^1}(\tau^0 > \tau^1)$$

$$P_{00}(s,t) = 1 - P_{01}(s,t) = \mathbb{P}_{\Pi^0\otimes\Pi^1}(\tau^1 \ge \tau^0)$$
$$P_{10}(s,t) = \mathbb{P}_{\Pi^0\otimes\Pi^1}(\tau^1 > \tau^0)$$
$$P_{11}(s,t) = 1 - P_{10}(s,t) = \mathbb{P}_{\Pi^0\otimes\Pi^1}(\tau^0 \ge \tau^1).$$

It is obvious that $P_{01}(s,t) < P_{11}(s,t)$ and $P_{10}(s,t) < P_{00}(s,t)$. Note that

$$\{\tau^0 = \tau^1\} = \{\tau^0 = \tau^1 = s\} = \{\Pi^0((s,t]) = \Pi^1((s,t]) = 0\}$$

up to a set of probability zero under the product distribution.

We can now give unbiased estimators of the transition probabilities as indicator functions of the inequalities. It is clear that these estimators depend on the honest times; however, we have omitted this explicit dependence in the notation for the sake of simplicity.

$$\widetilde{P}_{01}(s,t) = \mathbf{1}(\tau^0 > \tau^1),\ \widetilde{P}_{00}(s,t) = \mathbf{1}(\tau^1 \ge \tau^0)$$
$$\widetilde{P}_{10}(s,t) = \mathbf{1}(\tau^1 > \tau^0),\ \widetilde{P}_{11}(s,t) = \mathbf{1}(\tau^0 \ge \tau^1). \quad (11)$$

We refer to Appendix C for further insight into the properties of the two PPPs. We use (11) in the data augmentation algorithm for estimation of the unknown rates.

4 | Data Augmentation Algorithm for an Intermittently Observed Process

We now use the theory developed in Section 3.3 for a fixed time grid $0 = t_0 < t_1 < \cdots < t_n$. Let Π^k be a PPP with intensity $\alpha_k(t)$ on the interval $(t_0, t_n]$ and for each consecutive times let

$$\tau_i^k = t_{i-1} \vee \inf\{r : \Pi^k((r,t_i]) = 0\}, \quad i = 1,\ldots,n, k = 0,1.$$

Note that these honest random times τ_i^k's are independent across i, and that it is not required to explicitly simulate the entire PPP Π^k; it is sufficient to sample the relevant independent random times τ_i^k for each i. From the above definition, for example, of τ^1, remark that $P\left(\Pi^1(r,t_i] = 0\right) = \exp(-\int_r^{t_i}\alpha_1(s)ds)$, and this precisely has the uniform distribution in $(0,1]$ when we set $r = \tau^1$. The cumulative hazard function $A_1(r,t_i] = \int_r^{t_i}\alpha_1(s)ds$ represents the expected number of transitions from 1 to 0 in $(r,t_i]$. Also note that $P\left(\Pi^1(r,t_i] = 0\right) = 0$ implies that $P\left(\Pi^1(s,t_i] = 0\right) = 0,\ \forall\ s \ge r$.

As a direct consequence, we then have the following sampling scheme for τ_i^1's and τ_i^0's: for y_i^k independent Exponential(1) random variables, define

$$\tau_i^k = t_{i-1} \vee \inf\left\{r : \int_r^{t_i}\alpha_k(s)ds \le y_i^k\right\},\ \eta_i^k = \mathbf{1}(\tau_i^k > t_{i-1}) \quad (12)$$

and likelihood contributions

$$f_i^{(k)}(\tau_i^k;\theta) = \alpha_k(\tau_i^k)^{\eta_i^k}\exp\left(-\int_{\tau_i^k}^{t_i}\alpha_k(s)ds\right)\mathbf{1}\ (t_{i-1} \le \tau_i^k < t_i). \quad (13)$$

From (12), observe that when $\alpha_k(s) > 0$ almost everywhere, then r is simply the value such that $\int_r^{t_i}\alpha_k(s)ds = y_i^k$. Also note that when simulating values τ_i^k, it must be checked whether they lie inside

the interval $[t_{i-1}, t_i)$; if not, set $\tau_i^k = t_{i-1}$, under which case we say that truncation has occurred ($\eta_i^k = 0$) and the density $f_i^{(k)}(\tau_i^k; \theta)$ reduces to a point mass probability.

We furthermore designate by θ^k the unknown parameters of the transition rate $\alpha_k(t)$, $k = 0, 1$ and $\theta = (\theta^0, \theta^1)$, and we also assume that the realizations of the process are observed for J subjects, $\{X_j(t)\}$ at time points $\{t_{0,j}, t_{1,j}, \ldots, t_{n_j,j}\}$, $j = 1, \ldots, J$. Then the observed data likelihood for θ is

$$\begin{aligned} L(\theta; X) = P(X|\theta) &= \prod_{j=1}^{J} p(X(t_{0,j})) \prod_{i=1}^{n_j} P_{X(t_{i-1,j})X(t_i,j)}(t_{i-1,j}, t_{i,j}) \\ &= \prod_{j=1}^{J} p(X(t_{0,j})) L_j(\theta; X_j) \propto \prod_{j=1}^{J} L_j(\theta; X_j), \end{aligned} \tag{14}$$

where $p(X(t_{0,j}))$ is the initial distribution. When the initial states are given and our interest is not in the estimation of the initial distribution, we can drop the term containing the initial distribution and work with the terms containing θ. We assume that the parameters θ^0 and θ^1 are independent a priori and let their prior density functions be $\pi_k(\theta^k), k = 0, 1$. The posterior distribution of θ given the data is proportional to the product of the likelihood (14) and the priors $\pi_0(\theta^k)\pi_1(\theta^k)$. It is to be noted that the posterior distribution is difficult to obtain because of the lack of an analytical form for the transition probabilities. Below, we describe a Bayesian data augmentation method that utilizes the honest times and the transition probabilities (11).

4.1 | Bayesian Data Augmentation Algorithm

The next step involves augmenting the observed data, hereafter denoted by X, with the auxiliary honest time variables $(\tau_{i,j}^1, \tau_{i,j}^0)$ for the purpose of parameter estimation. We now also define the vector quantities $\boldsymbol{\tau}^0$ and $\boldsymbol{\tau}^1$ to designate the collection over the index $j = \{1, ..., J\}$ of honest times which results in n_j honest times for subject j. Recall from (11) that the transition probabilities may be estimated using the honest times as follows.

$$\widetilde{P}_{01}(t_{i-1,j}, t_{i,j}) = \mathbf{1}\,(\tau_{i,j}^0 > \tau_{i,j}^1), \widetilde{P}_{00}(t_{i-1,j}, t_{i,j}) = \mathbf{1}\,(\tau_{i,j}^1 \geq \tau_{i,j}^0), \tag{15}$$

$$\widetilde{P}_{10}(t_{i-1,j}, t_{i,j}) = \mathbf{1}\,(\tau_{i,j}^1 > \tau_{i,j}^0), \widetilde{P}_{11}(t_{i-1,j}, t_{i,j}) = \mathbf{1}\,(\tau_{i,j}^0 \geq \tau_{i,j}^1). \tag{16}$$

Then, assuming that the honest times and θ are given, the conditional probability of the observed data X gives the observed data likelihood for $(\theta, \boldsymbol{\tau}^1, \boldsymbol{\tau}^0)$ and is

$$\begin{aligned} \widetilde{L}(\theta, \boldsymbol{\tau}^1, \boldsymbol{\tau}^0; X) &= P(X|\theta, \boldsymbol{\tau}^1, \boldsymbol{\tau}^0) \\ &= \prod_{j=1}^{J} p(X(t_{0,j})) \prod_{i=1}^{n_j} \widetilde{P}_{X(t_{i-1,j})X(t_{i,j})}(t_{i-1,j}, t_{i,j}), \\ &\propto \prod_{j=1}^{J} \prod_{i=1}^{n_j} \widetilde{P}_{X(t_{i-1,j})X(t_{i,j})}(t_{i-1,j}, t_{i,j}), \end{aligned} \tag{17}$$

with the property

$$\mathbb{E}_{\Pi^1 \otimes \Pi^0}\left[\widetilde{L}(\theta, \boldsymbol{\tau}^1, \boldsymbol{\tau}^0; X)\right] = L(\theta; X).$$

This precisely shows that the likelihood $\widetilde{L}(\theta, \boldsymbol{\tau}^1, \boldsymbol{\tau}^0; X)$ is an unbiased estimator of the likelihood $L(\theta; X)$.

We make an observation that equation (17) does not depend on θ.

Note then that the joint density of the augmented data $(X, \boldsymbol{\tau}^1, \boldsymbol{\tau}^0)$ given θ can be expressed as the product

$$\begin{aligned} \mathcal{L}(\theta; X, \boldsymbol{\tau}^1, \boldsymbol{\tau}^0) &= P(X, \boldsymbol{\tau}^1, \boldsymbol{\tau}^0|\theta) = P(X|\theta, \boldsymbol{\tau}^1, \boldsymbol{\tau}^0) P(\boldsymbol{\tau}^1, \boldsymbol{\tau}^0|\theta) \\ &= \widetilde{L}(\theta, \boldsymbol{\tau}^1, \boldsymbol{\tau}^0; X) \prod_{j=1}^{J} \prod_{i=1}^{n_j} \prod_{k=0}^{1} f_{i,j}^{(k)}(\tau_{i,j}^k; \theta) \\ &\propto \prod_{j=1}^{J} \prod_{i=1}^{n_j} \widetilde{P}_{X(t_{i-1,j})X(t_{i,j})}(t_{i-1,j}, t_{i,j}) \\ &\quad \prod_{j=1}^{J} \prod_{i=1}^{n_j} \prod_{k=0}^{1} f_{i,j}^{(k)}(\tau_{i,j}^k; \theta). \end{aligned} \tag{18}$$

The Gibbs sampler applied to the posterior distribution of (θ^0, θ^1) and missing honest times involves specification of the following full conditional probabilities:

$$\begin{aligned} P(\boldsymbol{\tau}^1, \boldsymbol{\tau}^0|.) &\propto P(\boldsymbol{\tau}^1, \boldsymbol{\tau}^0|\theta) P(X|\theta, \boldsymbol{\tau}^1, \boldsymbol{\tau}^0) \\ &= P(\boldsymbol{\tau}^1, \boldsymbol{\tau}^0|\theta) \prod_{j=1}^{J} \prod_{i=1}^{n_j} \widetilde{P}_{X(t_{i-1,j})X(t_{i,j})}(t_{i-1,j}, t_{i,j}) \end{aligned} \tag{19}$$

$$\begin{aligned} P_k(\theta^k|.) &\propto \pi(\theta^k) P(\boldsymbol{\tau}^k|\theta^k), k = 0, 1 \\ &= \pi_k(\theta^k) \prod_{j=1}^{J} \prod_{i=1}^{n_i} f_{i,j}^{(k)}(\tau_{i,j}^k; \theta^k), \\ &= \pi_k(\theta^k) \prod_{j=1}^{J} \prod_{i=1}^{n_i} \alpha_k(\tau_{i,j}^k)^{\eta_{i,j}^k} \exp\left(- \int_{\tau_{i,j}^k}^{t_{i,j}} \alpha_k(s) ds \right). \end{aligned} \tag{20}$$

Note that $\alpha_k(t)$ is a function of θ^k. It is clear from (19) that the honest times can be sampled separately for each inter-observation interval. In order to provide intuition for the algorithm, remark that we first sample $(\tau_{i,j}^1, \tau_{i,j}^0)$ from (19) for each i and j, and the parameters θ should be sampled from (20), where all variables are being conditioned on their most recent sampled values. Note that the conditional distributions of $\theta^0|X, \boldsymbol{\tau}^0$ and of $\theta^1|X, \boldsymbol{\tau}^1$ given in (20) are independent and hence are also marginal distributions given $X, \boldsymbol{\tau}^0, \boldsymbol{\tau}^1$. If sampling from any of these full conditional probabilities is Difficult, then a Metropolis Hastings step can be used.

Step 0. Specify initial values of θ^0 and θ^1, the parameters of $\alpha_0(\cdot)$ and $\alpha_1(\cdot)$ respectively.

Next, repeat:

Step 1. For each individual $j = 1, ...J$ and each interval $[t_{i-1,j}, t_{i,j}]$, $i = 1, \ldots, n_j$, repeatedly sample $(\tau_{i,j}^0, \tau_{i,j}^1)$ from equation (12) until the pair is compatible with the observed states at the start and end of the inter-observation interval according to the compatibility rules encoded by (15) and (16). This way, we eventually get a value of 1 for the second product term in (19).

Step 2. After we have sampled and accepted the pairs of honest times, we need to update θ^0 and θ^1 by either sampling directly from (20) when possible or in separate Metropolis steps.

Ultimately, the above sampler relies on the fact that the $\{\tau^k_{i,j}\}$ are sampled from their correct conditional distribution given the data X and θ, as well as the fact that, conditioned on $\{\tau^k_{i,j}\}$, the parameters are conditionally independent of the observed states.

We also point out that sampling of $(\tau^0_{i,j}, \tau^1_{i,j})$ can be carried out in parallel for each i and j since they are sampled independent of each other. Also, the choice of initial values of θ is crucial, especially if the observed data are sparse, as will be discussed in the next section. Additional reading on a wider selection of Bayesian computational methods, which include those discussed above, may be found in (Robert and Casella 2004) for interested readers.

5 | Example

To illustrate the data augmentation algorithm described in Section 4.1, we consider Weibull-type transition rates which have the functional forms:

$$\alpha_k(t) = \lambda_k \gamma_k t^{\gamma_k - 1},\ \gamma_k, \lambda_k > 0, t > 0, \tag{21}$$

where γ_k and λ_k, $k = 0, 1$ are the shape and rate parameters respectively. The transition probabilities of Equation (4) in this case are, for $k \neq l, k, l = 0, 1$:

$$P_{kl}(s,t) = \exp\big(-\lambda_k\, t^{\gamma_k} - \lambda_l\, t^{\gamma_l}\big)\gamma_k \lambda_k \times \int_s^t \exp\big(\lambda_k\, v^{\gamma_k} + \lambda_l\, v^{\gamma_l}\big)\, v^{\gamma_k - 1} dv. \tag{22}$$

The advantage of our approach, therefore, lies in the fact that we do not need to use (22), and consequently we avoid the numerical integration.

Setting $\gamma_0 = \gamma_1 = 1$ gives the constant rates and corresponds to a time-homogeneous process. The above expression simplifies to

$$P_{kl}(s,t) = \frac{\lambda_k}{\lambda_0 + \lambda_1}(1 - \exp\{-(\lambda_0 + \lambda_1)(t - s)\}). \tag{23}$$

We first describe the data augmentation algorithm in the Weibull case, assuming independent Gamma priors for the four parameters. The density of the Gamma distribution with the shape and rate parameters (a, b) is given as

$$\pi(y; a, b) = \frac{b^a}{\Gamma(a)} y^{a-1} \exp\{-by\},\ \ a, b > 0, y > 0.$$

5.1 | Updating the Weibull Parameters

Here $\theta = (\gamma_0, \lambda_0, \gamma_1, \lambda_1)$. The integrated Weibull rates are

$$\int_r^t \alpha_k(s) ds = \lambda_k(t^{\gamma_k} - r^{\gamma_k}),\ \ k = 0, 1.$$

For given $(\tau^1_{i,j}, \tau^0_{i,j})$, $i = 1, \ldots, n_j$, $j = 1, \ldots, J$ and X, the full conditional distribution of $\theta^k = (\gamma_k, \lambda_k)$ is proportional to (20). Under the independent gamma priors for $\gamma_k \sim \text{Gamma}(\xi_k, \beta_k)$ and $\lambda_k \sim \text{Gamma}(a_k, b_k)$, we note that (20), for each k, simplifies as:

$$P(\gamma_k, \lambda_k | \boldsymbol{\tau}^1, \boldsymbol{\tau}^0, \mathbf{X}) \propto \pi(\gamma_k; \xi_k, \beta_k) \gamma_k^{\sum_{j,i} \eta^k_{i,j}} \prod_{j,i} (\tau^k_{i,j})^{\gamma_k \eta^k_{i,j}} \lambda_k^{a_k + \sum_{j,i} \eta^k_{i,j} - 1} \exp\{-\lambda_k (b_k + \sum_{j,i} (t^{\gamma_k}_{i,j} - (\tau^k_{i,j})^{\gamma_k})\}. \tag{24}$$

Hence, the full conditional distribution of λ_k given also γ_k is the gamma distribution with parameters $(a_k + \sum_{j,i} \eta^k_{i,j})$ and $(b_k + \sum_{j,i} (t^{\gamma_k}_{i,j} - (\tau^k_{i,j})^{\gamma_k}))$ as the shape and the scale parameters, respectively and $\eta^k_{i,j} = \mathbf{1}(\tau^k_{i,j} > t_{i-1,j})$. In this case, the gamma prior is a conjugate prior for λ_k and we can sample directly from its full conditional distribution. The log-marginal density of γ_k, obtained by first integrating out λ_k from (24), is proportional to

$$\log(P(\gamma_k | \boldsymbol{\tau}^1, \boldsymbol{\tau}^0, \mathbf{X})) \propto (\xi_k - 1)\log(\gamma_k) - \beta_k \gamma_k + \log(\gamma_k) \sum_{j,i} \eta^k_{i,j} + \gamma_k \sum_{j,i} \eta^k_{i,j} \log(\tau^k_{i,j}) + (a_k + \sum_{j,i} \eta^k_{i,j}) \log(b_k + \sum_{j,i} (t^{\gamma_k}_{i,j} - (\tau^k_{i,j})^{\gamma_k}). \tag{25}$$

We now apply the Bayesian data augmentation algorithm described at the end of Section 4.1 for obtaining their posterior distributions. Start with some initial values of θ (we will return to this later).

Step 1. The steps described below apply to all intervals $[t_{i-1,j}, t_{i,j}]$, $i = 1, \ldots, n_i$, $j = 1, \ldots, J$. Simulate $(\tau^k_{i,j}, \eta^k_{i,j})$, $k = 0, 1$ using the rejection sampling described earlier where $\eta^k_{i,j} = \mathbf{1}(\tau^k_{i,j} > t_{i-1,j})$ and

$$\tau^k_{i,j} = t_{i-1,j} \vee (t^{\gamma_k}_{i,j} - y_{i,j}/\lambda_k)^{1/\gamma_k} = \{t^{\gamma_k}_{i-1,j} \vee (t^{\gamma_k}_{i,j} - y_{i,j}/\lambda_k)\}^{1/\gamma_k},$$

where $y_{i,j}$ are independent sample from the Exponential (1) distribution.

Step 2. To update γ_k, we use a log-normal proposal γ^*_k with mean $\log(\gamma_k)$ and variance σ^2, and update γ_k with Metropolis-Hastings acceptance log-ratio (where the last term comes from the Jacobian of the log-transformation)

$$\left(\xi_k + \sum_{j,i} \eta^k_{i,j} - 1\right)(\log(\gamma^*_k) - \log(\gamma_k)) + \left(-\beta_k + \sum_{j,i} \eta^k_{i,j} \log(\tau^k_{i,j})\right)(\gamma^*_k - \gamma_k) + \left(a_k + \sum_{j,i} \eta_{i,j}\right)\left\{\log\left(b_k + \sum_{j,i} (t^{\gamma_k}_{i,j} - (\tau^k_{i,j})^{\gamma_k})\right) - \log\left(b_k + \sum_{j,i} (t^{\gamma^*_k}_{i,j} - (\tau^k_{i,j})^{\gamma^*_k})\right)\right\} + (\log \gamma^*_k - \log \gamma_k).$$

Note that the last term cancels out with a term from the first one, and we get the log-acceptance ratio as:

$$\left(\xi_k + \sum_{j,i} \eta_{i,j}^k\right)\left(\log(\gamma_k^*) - \log(\gamma_k)\right)$$
$$+ \left(-\beta_k + \sum_{j,i} \eta_{i,j}^k \log(\tau_{i,j}^k)\right)(\gamma_k^* - \gamma_k)$$
$$+ \left(a_k + \sum_{j,i} \eta_{i,j}\right)\left\{\log\left(b_k + \sum_{j,i}(t_{i,j}^{\gamma_k} - (\tau_{i,j}^k)^{\gamma_k})\right)\right.$$
$$\left. - \log\left(b_k + \sum_{j,i}(t_{i,j}^{\gamma_k^*} - (\tau_{i,j}^k)^{\gamma_k^*})\right)\right\}.$$

Then, given γ_k, λ_k is sampled from its full conditional distribution

$$\lambda_k \sim \text{Gamma}(a_k + \sum_{j,i} \eta_{i,j}^k, b_k + \sum_{j,i}(t_{i,j}^{\gamma_k} - (\tau_{i,j}^k)^{\gamma_k})).$$

In the constant intensities model, $\gamma = 1$, we update only λ directly from its full conditional distribution. For the sake of completeness, we provide the detailed algorithm for the constant rates in Appendix D.

5.1.1 | Choice of the Initial Values

The Bayesian data augmentation algorithm described above can be sensitive to the initial values, especially when the observed times are sparse. We suggest using the crude estimates obtained by treating the sampling points with changes in the observed state as if they were continuous time transitions. For the Weibull case, we can assign one as the initial values for the shape parameters and estimate the rate parameter λ_0 as the ratio of the number of direct transitions from 0 to 1 to the sum of the lengths of observation intervals starting from state 0. The initial value of λ_1 can be obtained correspondingly. We expect that these initial values of the rates provide lower bounds for the maximum likelihood estimates.

In case of the constant rates model, the estimators of rates, if the observed times were exact transition times, would be the ones that are described above. Since they are not the exact transition times, more (or at least as many as there are observed) transitions within each observed interval are expected, so the numerator would be as large as in the estimator, and the denominator would be smaller or equal to the one in the estimator. Alternatively, one can use the `msm` package to obtain estimates of the constant rates for an intermittently observed process. Similar suggestions are also discussed in (Titman 2011).

5.1.2 | Choice of the Priors and Hyperparameters

We have chosen a gamma distribution as a prior distribution for the rate parameter λ_k because of the conjugacy, so that its posterior distribution is also a gamma distribution. The parameters of the gamma prior (a, b) can be interpreted as the prior belief on the total number of events and the total exposure, respectively. Such information can be used to get reasonable values of these parameters. In the absence of prior information, we set both $a = b = 0.1$. The improper prior distribution of the form $1/\gamma_k$ may not be a good choice here because of the sparsely observed data.

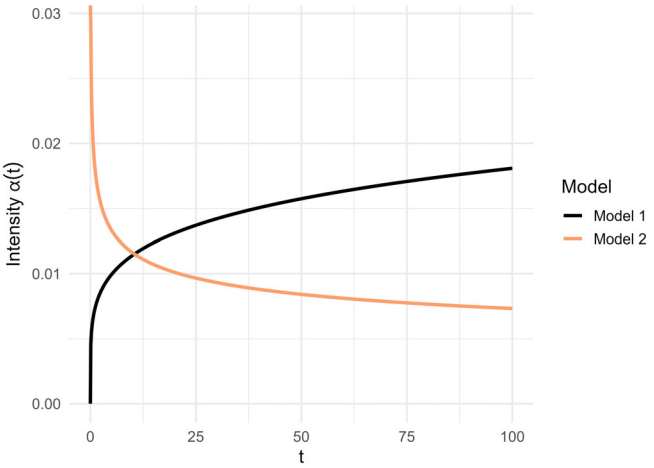


FIGURE 3 | Weibull intensity functions for $\alpha_0(t)$ (denoting a $0 \to 1$ transition) and shown by the black curve, as well as for $\alpha_1(t)$ for a $1 \to 0$ transition and shown by the orange curve.

5.1.3 | Model Selection

Approaches to model assessment based on comparison of parametric and nonparametric estimates, model expansion, etc., may be problematic in our case. We refer to section 5.2.3 of (Cook and Lawless 2018) for further discussion on this. In Appendix E, we briefly describe the model expansion approach, especially for the nested models that are considered here.

6 | Results

We first display the results for the Weibull rates (21) under two different scenarios: one is based on simulation of a continuous-time Markov process, and the other one uses real-life time grids from clinical visits data. In both scenarios, the initial distribution is set to $(0.5, 0.5)$ and the true shape and rate parameters: $\gamma_0 = 1.2$, $\gamma_1 = 0.8$, $\lambda_0 = 0.006$, and $\lambda_1 = 0.023$. For visualization purposes, respective plots of the Weibull intensity functions for $\alpha_0(t)$ and $\alpha_1(t)$ are shown in Figure 3.

Notice from the Weibull density (21) and Figure 3 that when the shape parameter $\gamma_k < 1$, then the intensity function decays as t increases. The black curve has $\gamma_0 = 1.2$ and the orange curve has $\gamma_1 = 0.8$. When $\gamma_k > 1$, the risk of experiencing a transition grows with time. The rate parameter λ_k controls how fast transitions happen. When we increase its value, the intensity curve proportionally shifts upwards, meaning a higher overall rate of transitions at all times.

Now, the workflow for the Bayesian inference consisted of running three separate chains, each having 10,000 iterations, and subsequently establishing a burn-in amount of iterations set to

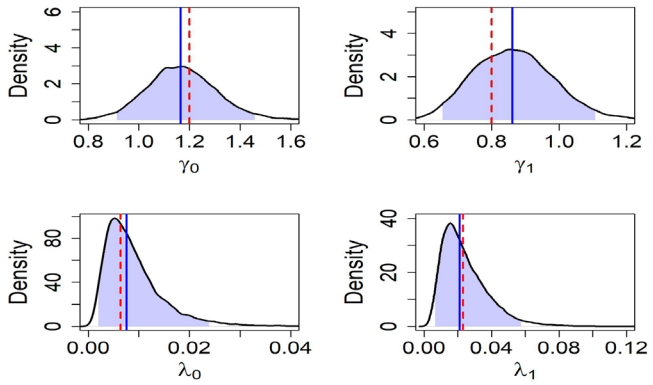


FIGURE 4 | Simulated time grid and Weibull rates. Posterior density plots for the four parameters λ_0, λ_1, γ_0 and γ_1 under the simulated time grid. The dashed red line shows the true value, the solid blue line indicates the posterior median, and the 95% credible intervals are shaded.

2000 for each chain. Then, posterior samples across the three chains were pooled, resulting in effectively 24,000 overall MCMC samples used to derive posterior statistics. We use the diagnostic tool of the multichain effective sample size (ESS) to ensure that convergence of the chains has been well reached and comment on the results. Note that a first chain was run with the initial values set according to the discussion in Section 5.1. Then, the two additional chains were initialized using slightly perturbed versions of these values, arbitrarily obtained by multiplying this initial set by factors slightly greater or less than 1, so that the chains started above and below the original values, respectively. We finally also present trace plots found in Appendix G (Figure G1) to visualize the behavior of the three chains and also establish convergence and proper mixing. We present the results under the homogeneous case in Appendix F (Tables F1 and F2, Figures F1 and F2).

6.1 | Scenario 1: Simulated Data

In this scenario, we first simulate $J = 100$ sample paths of a continuous-time Markov jump process with Weibull rates and a given time horizon $T = 100$. Then, we sample $n = 50$ time points uniformly over the interval $[0, T]$ and assign the occupied states from the simulated continuous path. Figure 4 shows the posterior densities of the four parameters, and the summary is shown in Table 1. It is clear from the plots that the posterior medians of the two shape parameters γ_0 and γ_1 fall above and below one, respectively, as expected. The posterior median of λ_1 is almost at the true value, while that of λ_0 is slightly above the true value. It is to be noted that an intermittently observed process may carry a different amount of information for each parameter, and hence, some parameters may be estimated more precisely than the others.

6.2 | Scenario 2: Real-Life Time Grid

Here we employ a real-life time grid from $J = 93$ subjects according to their clinical visits during a $T = 48$ month period, and simulate states occupied according to the Weibull rates. Thus, the real-time grid allows us to obtain observation times that actually correspond to real subjects attending their scheduled clinical assessments. Effectively, we are simulating the $X(t)$ process

TABLE 1 | Weibull rates. Posterior medians with (95% credible intervals) of parameters under simulated and real-life time grids.

Parameters	True	Simulated grid	Real-life grid
γ_0	1.2	1.17 (0.91, 1.46)	1.13 (1.01, 1.25)
λ_0	0.006	0.008 (0.002, 0.024)	0.010 (0.005, 0.024)
γ_1	0.8	0.86 (0.65, 1.11)	0.84 (0.70, 0.97)
λ_1	0.023	0.021 (0.007, 0.058)	0.019 (0.007, 0.049)

TABLE 2 | Summary statistics of the number of visits per subject in the real-life time grid.

Statistic	Range	Median (Q1, Q3)	Mean (SD)
Estimate	(9, 48)	29 (17, 39)	29 (12)

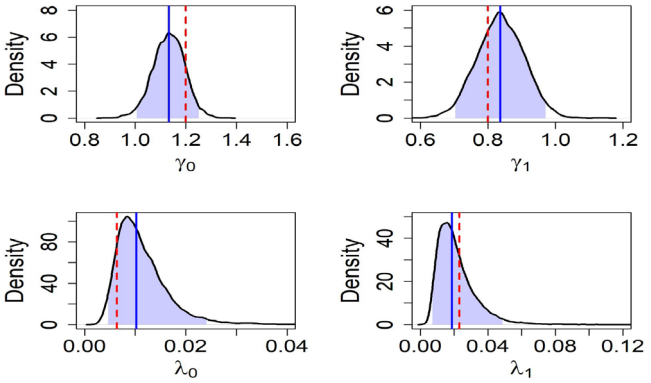


FIGURE 5 | Real-life time grid and Weibull rates. Posterior density plots for the four parameters λ_0, λ_1, γ_0, and γ_1 under the real-life time grid. As before, the dashed red line shows the true value, the solid blue line indicates the estimated value, and the 95% credible intervals are shaded.

using Weibull rates to generate the states at these real observation times. The subjects in this scenario have a varying number of observations, unlike in our simulated example, where n was fixed. This is what makes real-life clinical data complex, as it introduces irregularity in observation schedules and challenges in aligning individuals on a common assessment timeline. To better understand the real-life grid, we present descriptive statistics in Table 2. Figure 5 shows the posterior densities of the four parameters, and the summary is shown in Table 1. As in Section 6.1, the posterior medians of the two shape parameters γ_0 and γ_1 also fall above and below one, respectively. Just like in the simulated grid, λ_1 is estimated more precisely than λ_0 in the real-life time grid.

Lastly, from Figures 4 and 5, we can see that the uncertainty in the γ parameters has reduced (much narrower interval width) in the real-life grid as opposed to the simulated grid, but this is not the case for the λ parameters. One may initially argue that this means there is an increase in the amount of information about the decay (or increase) of the transition intensity. Yet taking a closer look at Table 2 shows that the average number of grid points (observation times) n for an individual j has nearly halved (suggesting that the information should also have been halved), and we know there are 0.93 times the amount of subjects compared to

TABLE 3 | Weibull rates. Multichain Effective Sample Sizes (ESS) for the Bayesian model parameters under simulated and real-life time grids.

Parameters	ESS simulated grid	ESS Real-life grid
γ_0	744	752
λ_0	768	712
γ_1	584	905
λ_1	595	747

the simulated grid. One possible explanation for why the uncertainty has actually been reduced for the shape parameters γ is that the real-life observation regime concentrates measurement values near the start of the process, reflecting clinical practice whereby observations are typically scheduled around the period when state changes are most likely to occur, thus making it easier to discern γ.

The results from Table 3 are very satisfactory for establishing that the chains explored the parameter space efficiently and produced stable results. Indeed, the relatively high values ranging from about 580 to 900 show that the computed posterior estimates are in fact based on a sufficient number of effectively independent samples. Moreover, the trace plots displayed in Figure G1, Appendix G confirm that all three chains have explored the same posterior region and that the mixing is adequate. There are no visibly persistent trends, and the plots are generally centered around a constant level. Lastly, the Gelman-Rubin statistic ($\hat{R}$) was also computed as a diagnostic tool to assess the convergence. All values for both the real and simulated time grids were very close to 1 (point estimates between 1.00 and 1.01 with upper confidence limits ≤ 1.02), suggesting that the chains have converged to the same stationary distribution and have mixed well. For both grids, the multivariate $\hat{R} \approx 1.0$ as well.

7 | Discussion

In this paper, we have proposed a simple and novel data augmentation algorithm with honest times in a Bayesian setting. We implement it using the full conditional distributions of the parameters and the honest times to estimate the transition intensities for two-state nonhomogeneous Markov models when the trajectories are observed intermittently on irregular and different time grids. Our approach is shown to be very adaptable because it can perform reliable parameter estimation under both homogeneous rates, as demonstrated by Tables F1 and F2 (the results are very similar to the `msm` package) and nonhomogeneous rates (Table 1), irrespective on the coarseness of the underlying time grids. When the integrated PPP intensities have an analytic form, our algorithm does not introduce time-discretization errors as the `nhm` package (Titman 2011) does when solving numerically the linear ODE for the transition matrices. The `msm` package (Jackson 2011) based on the spectral decomposition or series expansion, should be preferred for time-homogeneous Markov models to avoid unnecessary computations needed for the nonhomogeneous case.

The precision with which the model parameters are estimated depends on the sparseness of the observed times, that is, $t_{i,j} - t_{i-1,j}$. This point is explored in detail for time-homogeneous Markov models in section 5.2.2 of (Cook and Lawless 2018), and the readers interested in further details may refer to it. The derivatives of the transition probabilities and the Fisher information for a general multistate model with intermittent observations are given in the appendix in (Titman 2011). In general, it is difficult to use the Fisher information to comment directly on the precision and the time spacings. The proportions of truncated honest times are related to the information content of the intermittently sampled data.

It should be noted that common MCMC convergence diagnostic tools like the ESS or $\hat{R}$ may fail to correctly assess convergence under more specific circumstances such as having heavy-tailed or multimodal posteriors. We hypothesize that our set-up of sparse and irregular data may actually pose challenges with regard to weak identifiability and a Fisher information matrix that is near-singular, which may consequently complicate the convergence assessment of MCMC samples. This is why Vehtari et al. (2021) propose, among others, the use of rank-based $\hat{R}$ or ESS values based on quantiles as alternative diagnostic tools to try and remedy these issues (Vehtari et al. 2021). It would also be interesting to explore simulation-based calibration as another diagnostic tool to validate the Bayesian model by checking that the generated posterior samples are consistent with the true underlying data-generating process (Talts et al. 2018).

Given the simulated time grids, we only have access to a partially sampled trajectory of the process for each subject. As a next step, two research objectives are of interest from a clinical point of view, and our proposed data augmentation algorithm can be extended to meet these objectives. The first aim is to estimate the expected total time spent in state 0 over the time interval $[0, T]$. The second goal is to be able to predict the sample path for the selected time grid.

It is important to acknowledge that, for the purposes of illustration, we have only focused on Weibull-type rates as an example. Nonetheless, we point out that our algorithm also works with other time-dependent rates, such as Gompertz rates. It would also be worthwhile in further works to apply our algorithm in the case of time-varying covariates when the time data are sparse. Additionally, our proposed workflow can be extended to multiple correlated nonhomogeneous Markov two-state processes. Implications and extensions of the proposed method for general multistate models are also of interest. Beyond parametric models, in a forthcoming work, we are also combining the data augmentation via honest times together with a Bayesian nonparametric model based on generalized Gamma subordinators (see (Gasbarra et al. 2015) for an illustration on the use and applicability of Bayesian nonparametric modelling applied to survival-type data). In fact, we believe that explaining jump-based behavior using infinite-activity Lévy-based processes will provide a much-needed versatile and rich framework because we will not need to assume any specific parametric family of distributions for the underlying process governing the stochastic dynamics.

Acknowledgments

We thank the reviewers for their detailed and constructive comments that have led to an improved version of the manuscript. Etienne Sebag's research was supported by the Research Council of Finland (338507) and

University of Helsinki, Aalto University and KU Leuven Seed Fund. The authors thank Terhi Ollila, Department of Ophthalmology, Helsinki University Hospital, Helsinki, Finland, for providing real-time grid based on her doctoral study data. Open access publishing facilitated by Helsingin yliopisto, as part of the Wiley - FinELib agreement.

Funding

This work was supported by the Research Council of Finland (Grant No. 338507).

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Aksamit, A., T. Choulli, and M. Jeanblanc. 2021. "Thin Times and Random Times' Decomposition." *Electronic Journal of Probability* 26: 1–22. https://doi.org/10.1214/20-EJP569.

Aralis, H., and R. Brookmeyer. 2019. "A Stochastic Estimation Procedure for Intermittently-Observed Semi-Markov Multistate Models With Back Transitions." *Statistical Methods in Medical Research* 28, no. 3: 770–787. https://doi.org/10.1177/0962280217736342.

Barone, R., and A. Tancredi. 2022. "Bayesian Inference for Discretely Observed Continuous Time Multi-State Models." *Statistics in Medicine* 41, no. 19: 3789–3803. https://doi.org/10.1002/sim.9449.

Cheung, L. C., P. S. Albert, S. Das, and R. J. Cook. 2022. "Multistate Models for the Natural History of Cancer Progression." *British Journal of Cancer* 127, no. 7: 1279–1288. https://doi.org/10.1038/s41416-022-01904-5.

Cook, R. J., and J. F. Lawless. 2018. *Multistate Models for the Analysis of Life History Data*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315119731.

Gasbarra, D., E. Arjas, A. Vehtari, R. Slama, and N. Keiding. 2015. "The Current Duration Design for Estimating the Time to Pregnancy Distribution: A Nonparametric Bayesian Perspective." *Lifetime Data Analysis* 21, no. 4: 594–625. https://doi.org/10.1007/s10985-015-9333-0.

Gill, R. D., and S. Johansen. 1990. "A Survey of Product-Integration With a View Toward Application in Survival Analysis." *Annals of Statistics* 18, no. 4: 1501–1555. https://doi.org/10.1214/aos/1176347865.

Gronau, Q. F., A. Sarafoglou, D. Matzke, et al. 2017. "A Tutorial on Bridge Sampling." *Journal of Mathematical Psychology* 81: 80–97. https://doi.org/10.1016/j.jmp.2017.09.005.

Hobolth, A., and E. A. Stone. 2009. "Simulation From Endpoint-Conditioned, Continuous-Time Markov Chains on a Finite State Space, With Applications to Molecular Evolution." *Annals of Applied Statistics* 3, no. 3: 1204–1231. https://doi.org/10.1214/09-AOAS247.

Hubbard, R. A., L. Y. T. Inoue, and J. R. Fann. 2007. "Modeling Nonhomo-Geneous Markov Processes via Time Transformation." *Biometrics* 64, no. 3: 843–850. https://doi.org/10.1111/j.1541-0420.2007.00932.x.

Jackson, C. H. 2011. "Multi-State Models for Panel Data: The MSM Package for R." *Journal of Statistical Software* 38, no. 8: 1–28. https://doi.org/10.18637/jss.v038.i08.

Jeulin, T. 1980. *Semi-Martingales et Grossissement d'une Filtration*. Springer. https://doi.org/10.1007/bfb0093539.

Kendall, E. B., J. P. Williams, G. H. Hermansen, F. Bois, and V. H. Thanh. 2024. "Beyond Time-Homogeneity for Continuous-Time Multistate Markov Models." *Journal of Computational and Graphical Statistics* 34, no. 2: 668–682. https://doi.org/10.1080/10618600.2024.2388609.

Klausch, T., E. U. Akwiwu, M. A. van de Wiel, V. M. Coupe, and J. Berkhof. 2023. "A Bayesian Accelerated Failure Time Model for Interval Censored Three-State Screening Outcomes." *Annals of Applied Statistics* 17, no. 2: 1285–1306. https://doi.org/10.1214/22-AOAS1669.

Luo, Y., and C. Sherlock. 2025. "Bayesian Inference for the Markov-Modulated Poisson Process With an Outcome Process." *Journal of the Royal Statistical Society. Series C, Applied Statistics* 74, no. 5: 1239–1254. https://doi.org/10.1093/jrsssc/qlaf021.

Luo, Y., D. A. Stephens, A. Verma, and D. L. Buckeridge. 2021. "Bayesian Latent Multi-State Modeling for Non-Equidistant Longitudinal Electronic Health Records." *Biometrics* 77, no. 1: 78–90. https://doi.org/10.1111/biom.13261.

Machado, R. J. M., A. van den Hout, and G. Marra. 2021. "Penalised Maximum Likelihood Estimation in Multi-State Models for Interval-Censored Data." *Computational Statistics & Data Analysis* 153: 107057. https://doi.org/10.1016/j.csda.2020.107057.

Mehtäälä, J., K. Auranen, and S. Kulathinal. 2015. "Optimal Designs for Epidemiologic Longitudinal Studies With Binary Outcomes." *Statistical Methods in Medical Research* 24, no. 6: 803–818. https://doi.org/10.1177/0962280211430663.

Nevala, A., S. Heinävaara, T. Sarkeala, and S. Kulathinal. 2024. "Bayesian Hidden Markov Model for Natural History of Colorectal Cancer: Handling Misclassified Observations, Varying Observation Schemes and Unobserved Data." *Annals of Applied Statistics* 18, no. 4: 3050–3070. https://doi.org/10.1214/24-AOAS1922.

Robert, C. P. 2025. "Bayesian Inference: Theory, Methods, Computations." *Chance* 38, no. 1: 64. https://doi.org/10.1080/09332480.2025.2473298.

Robert, C. P., and G. Casella. 2004. *Monte Carlo Statistical Methods*. 2nd ed. Springer Texts in Statistics. Springer. https://doi.org/10.1007/978-1-4757-4145-2.

Taha, R. T., S. S. Ahmed, Q. Y. Hatim, and M. J. Abu-AlShaeer. 2024. "Advanced Bayesian Methods for Longitudinal Data Analysis in Public Health." *Journal of Ecohumanism* 3, no. 5: 270–289. https://doi.org/10.62754/joe.v3i5.3906.

Talts, S., M. Betancourt, D. Simpson, A. Vehtari, and A. Gelman. 2018. "Validating Bayesian Inference Algorithms With Simulation-Based Calibration." arXiv:1804.06788.

Tancredi, A. 2019. "Approximate Bayesian Inference for Discretely Observed Continuous-Time Multi-State Models." *Biometrics* 75, no. 3: 966–977. https://doi.org/10.1111/biom.13019.

Thom, H., D. Incerti, C. Jackson, F. Lamrock, and C. Williams. 2025. "Continuous Time Multistate Models." In *R for Health Technology Assessment*, edited by G. Baio, H. Thom, and P. Pechlivanoglou, 299–329. Chapman and Hall/CRC. https://doi.org/10.1201/9781003031819-11.

Titman, A. C. 2011. "Flexible Nonhomogeneous Markov Models for Panel Observed Data." *Biometrics* 67, no. 3: 780–787. https://doi.org/10.1111/j.1541-0420.2010.01550.x.

Titman, A. C. 2019. "nhm: Non-Homogeneous Markov and Hidden Markov Multistate Models." R Package. https://CRAN.R-project.org/%20package=nhm.

Tom, B. D. M., and V. T. Farewell. 2011. "Intermittent Observation of Time-Dependent Explanatory Variables: A Multistate Modelling Approach." *Statistics in Medicine* 30, no. 30: 3520–3531. https://doi.org/10.1002/sim.4429.

Vehtari, A., A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner. 2021. "Rank-Normalization, Folding, and Localization: An Improved R-Hat for Assessing Convergence of MCMC (With Discussion)." *Bayesian Analysis* 16: 667–718. https://doi.org/10.1214/20-BA1221.

Williams, J. P., C. B. Storlie, T. M. Therneau, C. R. Jack Jr., and J. Hannig. 2020. "A Bayesian Approach to Multistate Hidden Markov Models: Application to Dementia Progression." *Journal of the American Statistical Association* 115, no. 529: 16–31. https://doi.org/10.1080/01621459.2019.1594831.

Zingoni, Z. M., T. F. Chirwa, J. Todd, and E. Musenge. 2021. "A Review of Multistate Modelling Approaches in Monitoring Disease Progression: Bayesian Estimation Using the Kolmogorov-Chapman Forward Equations." *Statistical Methods in Medical Research* 30, no. 5: 1373–1392. https://doi.org/10.1177/0962280221997507.

Appendix A: Solution to the Kolmogorov Forward Equations (3)

The ODEs in (3) obtained using the Kolmogorov forward equations are:

$$\begin{aligned}\frac{\partial}{\partial t}P_{01}(s,t) &= P_{00}(s,t)\alpha_0(t) - P_{01}(s,t)\alpha_1(t)\\ &= (1-P_{01}(s,t))\alpha_0(t) - P_{01}(s,t)\alpha_1(t)\\ &= \alpha_0(t) - (\alpha_0(t)+\alpha_1(t))P_{01}(s,t),\\ \frac{\partial}{\partial t}P_{10}(s,t) &= -P_{10}(s,t)\alpha_0(t) + P_{11}(s,t)\alpha_1(t)\\ &= -P_{10}(s,t)\alpha_0(t) + (1-P_{10}(s,t))\alpha_1(t)\\ &= \alpha_1(t) - (\alpha_0(t)+\alpha_1(t))P_{10}(s,t)\end{aligned}$$

We can now use the following result to solve the above system of nonhomogeneous differential equations.

Solution to a system of nonhomogeneous differential equation of the type

$$y'(x) = A(x)y(x) + b(x)$$

where $y(x)$, $y'(x)$ and $b(x)$ are $n\times 1$ vectors and $A(x)$ is a $n\times n$ matrix, and is given as

$$y(x) = U(t)c_0 + U(t)\int_s^t U^{-1}(w)b(w)dw,$$

where $U(t) = \exp\{\int_s^t A(u)du\}$ c_0 is the constant vector of integration and is obtained, for example, using $y(x_0) = y_0$.

Appendix B: Honest Times

Definition. A random time τ is an *honest time* w.r.t. to the filtration $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$ if and only if for every $t > 0$ there exists an $\mathcal{F}_t$-measurable random variable τ_t such that $\tau = \tau_t$ on the event $\{\tau < t\}$. Then it is always possible to choose τ_t such that $\tau_t \leq \tau$ $\mathbb{P}$-a.s. (Aksamit et al. 2021; Jeulin 1980).

Intuitively an honest time is the last time something happens, so that when it happens we see in our information filtration that it is happening without necessarily knowing that it is happening for the last time, and because of that it is not a stopping time.

For example, if $0 = \sigma_0 \leq \sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_k \leq \ldots$ are stopping times in the filtration $\mathbb{F}$ and $T > 0$, the $\mathcal{F}_T$-measurable random time

$$\tau_T = \sup_{k\in\mathbb{N}}\{\sigma_k \mathbf{1}(\sigma_k \leq T)\}$$

is an honest time but not necessarily a stopping time, because

$$\{\tau_T \leq t\} = \bigcap_{k\in\mathbb{N}}(\underbrace{\{\sigma_k > T\}}_{\in\mathcal{F}_T} \cup \underbrace{\{\sigma_k \leq t\}}_{\in\mathcal{F}_t})$$

which is not necessarily in $\mathcal{F}_t$ when $t < T$. Moreover on the event $\{\tau_T \leq t\}$, $\tau_T = \tau_t$, which is $\mathcal{F}_t$-measurable.

Appendix C: Further Insight Into the Two PPPs

Given the two independent PPP with intensity $\alpha_0(t)$ and $\alpha_1(t)$ and the initial state $X(s)$, then one could, in principle, construct the trajectory of $(X(r) : s < r \leq t)$. The interpretation of the aforementioned construction is as follows. When $X(s) = 0$, then the condition $X(t) = 1$ holds if, and only if, the random times defined above satisfy $\tau^0 > \tau^1$, regardless of the number of possible transitions of X between the states 0 and 1 in the interval $(s,t]$. This directly means that, in the interval $(s,t]$, there is at least one $0\to 1$ transition, and after the last $0\to 1$ transition, there are no more $1\to 0$ transitions, hence $X(t) = 1$. We therefore interpret τ^0 as the last time point such that there is at least one transition from $0\to 1$ in the interval $(s,\tau^0]$ and no such transition between $(\tau^0,t]$. Similarly, τ^1 is the last time point such that there is at least one transition from $1\to 0$ in the interval $(s,\tau^1]$ and no such transition between $(\tau^1,t]$. Indeed, for the process to be in state 1 at time t, the last transition can not be from 1 to 0 and hence, $\tau^1 < \tau^0$. This is equivalent to saying that τ^1 is the first time point such that the corresponding PPP has no jump points inside the interval $(\tau^1,t]$. We have

$$\begin{aligned}P_{11}(s,t) - P_{01}(s,t) &= P_{00}(s,t) - P_{10}(s,t)\\ &= \exp\Big(-\int_s^t \big(\alpha_0(u)+\alpha_1(u)\big)du\Big)\end{aligned}$$

The assumption

$$\int_s^\infty \big(\alpha_0(v)+\alpha_1(v)\big)dv = \infty$$

implies mixing in total variation

$$\lim_{t\to\infty}\sum_{j=1,2}|P_{1j}(s,t) - P_{2j}(s,t)| = 0$$

without stationarity conditions. Indeed for two coupled copies of the Markov process, X_t and X'_t, driven by the same pair of independent Poisson point processes Π^1 and Π^0 with respective intensities $\alpha_1(t)$ and $\alpha_0(t)$, started at time s with $X_s = 1$ and $X'_s = 0$, the stopping time

$$\tau = \inf\big\{t : \Pi^1((s,t]) + \Pi^0((s,t]) > 0\big\}$$

is a coupling time, such that $X_t = X'_t,\ \forall t \geq \tau$, and the total variation distance is given by $\mathrm{TV}(X_t, X'_t) = 2P(\tau > t),\ \forall\ t > s$.

Appendix D: Constant Transition Rates and Data Augmentation Algorithm

Here we describe the data augmentation algorithm using both τ^1 and τ^0.

1. Start with some initial values of (λ_0,λ_1), and then simulate the initial values of $(\tau^1_{i,j},\tau^0_{i,j})$ as in Step 1, Section 5.1 with $\gamma_0 = \gamma_1 = 1$. Initial value of, for example, λ_0, can be obtained as the ratio of the number of intervals starting in 0 and ending in 1 to the sum of the interval lengths starting in 0. We sample the pairs by rejection sampling.
   Accept the pair $(\tau^1_{i,j},\tau^0_{i,j})$ when they satisfy the condition constrained by the states at $t_{i-1,j}$ and $t_{i,j}$.

2. For given $(\tau^1_{i,j},\tau^0_{i,j}), i = 1,\ldots,n_j, j = 1,\ldots,J$ and X, the posterior distribution of (λ_0,λ_1) is proportional to the likelihood in (18) and the independent gamma prior. Note that the first term in (18) is one due to the way $(\tau^1_{i,j},\tau^0_{i,j}), i = 1,\ldots,n_j, j = 1,\ldots,J$ are sampled.
   Also, the likelihood factorizes into two parts; one involving only λ_0 and the other one involving only λ_1. We can write the posterior distribution of λ_k by collecting all terms involving λ_k as follows.

$$\begin{aligned}P(\lambda_k|\boldsymbol{\tau}^1,\boldsymbol{\tau}^0,\mathbf{X}) &\propto \lambda_k^{a_k-1}\exp\{-b_k\lambda_k\}\prod_{j=1}^{J}\prod_{i=1}^{n_j} f^{(k)}_{i,j}(\tau^k_{i,j};\lambda_k)\\ &= \lambda_k^{a_k+\sum_j \sum_i \eta^k_{i,j}-1}\exp\{-\lambda_k(b_k + \sum_j\sum_i (t_{i,j}-\tau^k_{i,j})\},\end{aligned}$$

   which is a gamma density with $(a_k + \sum_{j,i}\eta^k_{i,j})$ and $(b_k + \sum_{j,i}(t_{i,j} - \tau^k_{i,j}))$ as the shape and the scale parameters, respectively and $\eta^k_{i,j} = \mathbf{1}(\tau^k_{i,j} > t_{i-1,j})$.

3. Simulate values of (λ_0,λ_1) from the above posterior distributions.

4. Simulate pairs $(\tau^1_{i,j},\tau^0_{i,j})$ as in step 1.

5. Repeat until the convergence.

Appendix E: Model Selection in the Class of Nested Models

We can use the concept of Bayes factor and the model indicator $m = 0, 1$. In case of exponential and Weibull case, the shape parameter γ is equal to one or is different from one. We compare two Bayes models, $\mathcal{M}_m = \{\mathcal{P}_m, \pi_m^1\}$, $m = 0, 1$ (section 8.2.1, (Robert 2025)). The common prior is a mixture of the Dirac-measure at $\gamma = 1$ (say $\widetilde{\pi}_0^1$) and the prior for the continuous γ (say $\widetilde{\pi}_1^1$). Let the prior for the parameters that are not subjected to the Bayes test, rate parameters of the Weibull model, have priors $\pi_{kl}^1(\lambda_{kl}), k, l = 0, 1$. The Bayes test for testing the point-null hypothesis $H_0 : \mathcal{M}_0$ versus $H_1 : \mathcal{M}_1$ can be based on the Bayes factor

$$B_{01} = \frac{m_0(\mathbf{X})}{m_1(\mathbf{X})}, \tag{E1}$$

where $m_k(\mathbf{X})$ is the marginal distribution of the data $\mathbf{X}$ given the model $\mathcal{M}_k$ and

$$\begin{aligned} m_0(\mathbf{X}) &= \int_\theta \mathbb{P}(\mathbf{X}|\theta)\pi_0^1(\theta)d\theta \\ &= \int_{\lambda_0,\lambda_1} \mathbb{P}(\mathbf{X}|\theta_0)\pi_{01}^1(\lambda_0)\pi_{10}^1(\lambda_1)d\lambda_0 d\lambda_1, \end{aligned}$$

where $\theta_0 = (\gamma_0 = 1, \lambda_0, \gamma_1 = 1, \lambda_1)$. For continuous prior π^0 of $\theta = (\gamma_0, \lambda_0, \gamma_1, \lambda_1)$ we have the denominator

$$m_1(\mathbf{X}) = \int_\theta \mathbb{P}(\mathbf{X}|\theta)\pi_1^1(\theta)d\theta,$$

where $\pi_1^1(\theta) = \widetilde{\pi}_1^1(\gamma_0)\pi_{01}^1(\lambda_0)\widetilde{\pi}_1^1(\gamma_1)\pi_{10}^1(\lambda_1)$. We have assumed gamma prior distributions for all four parameters. We need to compute the marginal likelihoods. For this purpose we can use one or all of the four algorithms described in a tutorial article (Gronau et al. 2017).

Appendix F: Example: Time-Homogeneous Markov Process

Here, we present the results under the homogeneous case ($\gamma_0 = \gamma_1 = 1$) for both a simulated time grid (using the same configuration of $J = 100$, $n = 50$ and $T = 100$) as well as the same real-life visiting times of subjects.

For the simulated process, we consider two sets of rate parameters; one has λ_0 and λ_1 fixed close to each other ($\lambda_0 = 0.047$ and $\lambda_1 = 0.051$) and the other where they are set further apart from each other ($\lambda_0 = 0.1$ and $\lambda_1 = 1.0$). Table F1 shows the results obtained using the proposed data augmentation algorithm (Appendix D) and the `R` package `msm`. The results are similar in both scenarios.

Note that we employ the same MCMC routine as for the Weibull example in the main text, running three chains each of 10,000 iterations, and each started from different values. We set the burn-in amount of 2000 for each chain. The MCMC convergence diagnostic tools such as ESS or $\hat{R}$ were checked, but the results are not formally reported. The values were excellent for the simulated grids under both pairs of parameter values (close and far), and satisfactory for the real-life time grids. The posterior plots for the time-homogeneous workflow are displayed in Figures F1 and F2.

TABLE F1 | Simulated data for a time-homogeneous Markov model.

Parameter	Bayesian data augmentation	`msm`
$\lambda_0 = 0.047$	0.048 (0.041, 0.054)	0.048 (0.042, 0.055)
$\lambda_1 = 0.051$	0.052 (0.045, 0.060)	0.052 (0.045, 0.060)
$\lambda_0 = 0.100$	0.103 (0.087, 0.122)	0.103 (0.087, 0.122)
$\lambda_1 = 1.000$	1.127 (0.974, 1.338)	1.134 (0.971, 1.324)

Note: Posterior medians/maximum likelihood estimate with respective 95% CIs using different estimation methods. Two sets of parameters are considered; one where the two rates are close and the other one where the two rates are far apart.

TABLE F2 | Real-life time grid. Posterior medians/maximum likelihood estimate with respective 95% CIs using different estimation methods.

Parameter	BDA	`msm`	`nhm`	Exact
$\lambda_0 = 0.047$	0.047 (0.038, 0.067)	0.046 (0.036, 0.058)	0.040 (N/A, N/A)	0.047 (0.0378, 0.0610)
$\lambda_1 = 0.051$	0.049 (0.039, 0.069)	0.047 (0.038, 0.060)	0.044 (N/A, N/A)	0.049 (0.0391, 0.0638)

Abbreviations: BDA, Bayesian Data Augmentation; Exact, Exact likelihood based estimation.

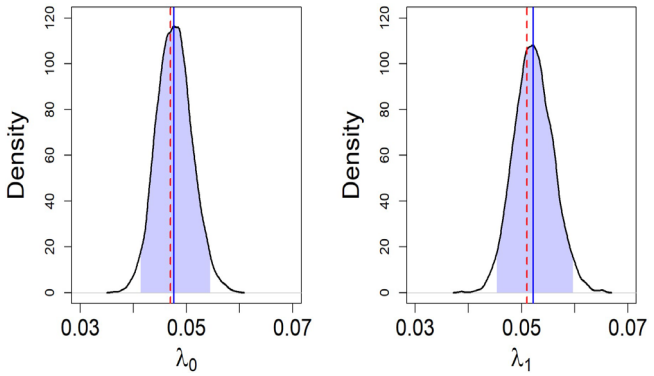


FIGURE F1 | Simulated time grid and the time-homogeneous transition intensities. Posterior density plots for λ_0 and λ_1 under the simulated time grid. The dashed red line shows the true value, the solid blue line indicates the estimated value, and the 95% credible intervals are shaded.

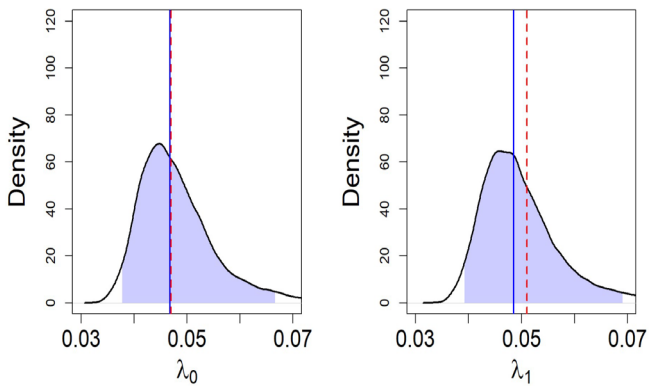
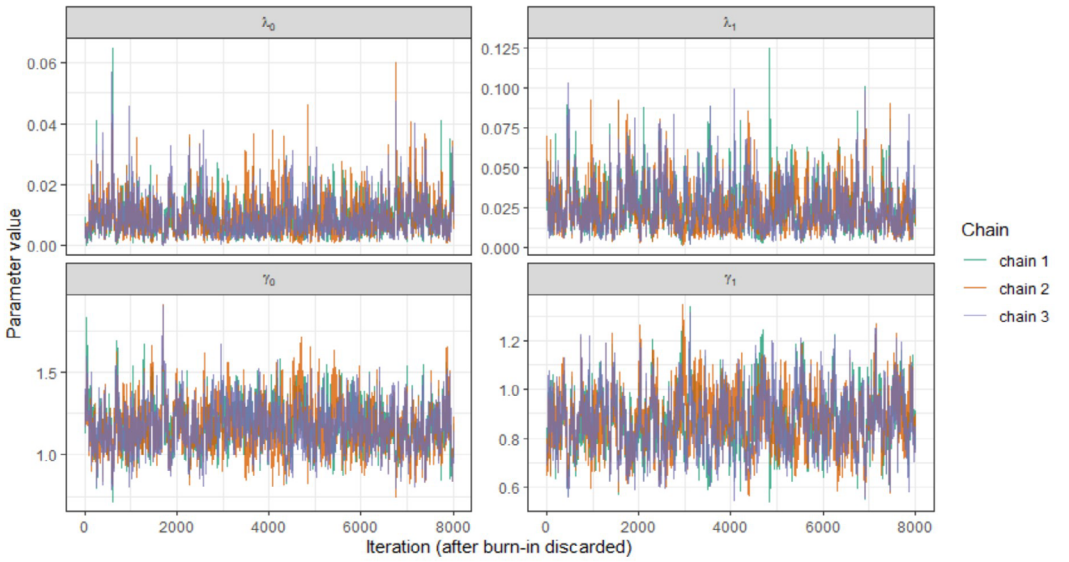


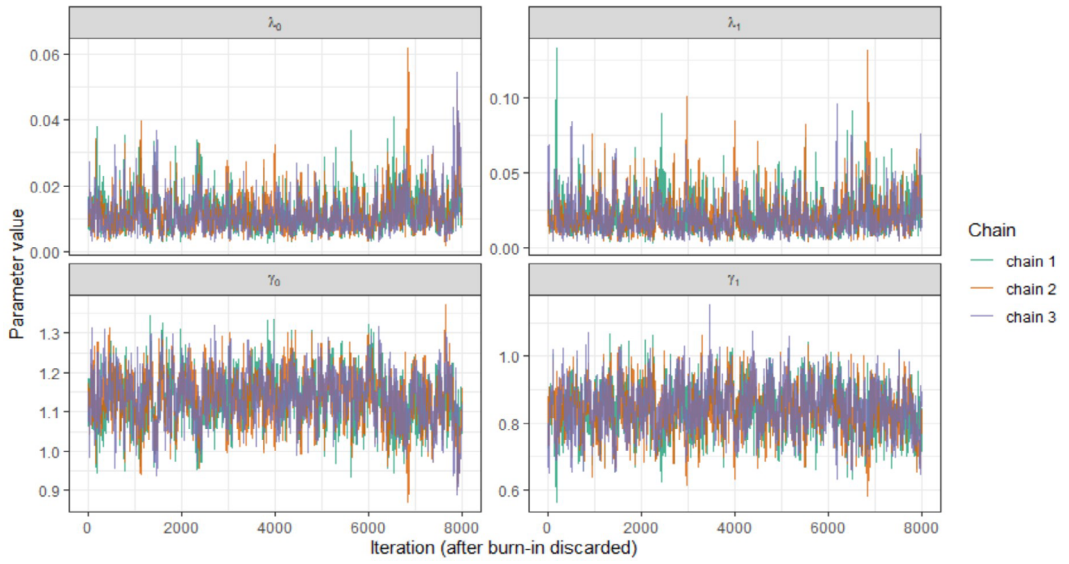
FIGURE F2 | Real-life time grid and the time-homogeneous transition intensities. Posterior Density plots λ_0 and λ_1 under the real-life time grid. The dashed red line shows the true value, the solid blue line indicates the estimated value, and the 95% credible intervals are shaded.

We additionally note that estimation of the parameters in the time homegeneous case is possible by directly using the likelihood function in Equation (14). For the real-life time grid, we use ($\lambda_0 = 0.047$ and $\lambda_1 = 0.051$) to simulate the states occupied at each observed time corresponding to a scheduled clinical time from real-life practice. We then use the exact likelihood, the proposed data augmentation algorithm, and the `R` packages `msm` and `nhm`. The results comparing the four estimation approaches are shown in Table F2. The nhm package fails to provide the confidence intervals due to numerical issues relating to the computation of the Hessian.

Appendix G: Example: Trace Plots for Weibull Example



(a) Simulated grid



(b) Real-life grid

FIGURE G1 | MCMC diagnostics: Trace plots for the three chains. (a) Simulated grid, (b) Real-life grid.