Journal of the Royal Statistical Society Series C: Applied Statistics, 2025, pp. 1–3

doi: DOI HERE
Advance Access Publication Date: Day Month Year
Paper

PAPER

Supplementary materials for "Bayesian inference for the Markov-modulated Poisson process with an outcome process"

Yu Luo[1,*] and Chris Sherlock[2]

[1]Department of Mathematics, King's College London, London, U.K. and [2]Department of Mathematics and Statistics, Lancaster University, Bailrigg, U.K.
[*]Corresponding author. yu.luo@kcl.ac.uk
FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

This is the supplementary materials for Bayesian inference for the Markov-modulated Poisson process with an outcome process.

1. Using the Metropolis–Hastings algorithm for inference

We have presented an exact Gibbs sampler for inference on a Markov modulated Poisson process with an observation process at event times. If non-conjugate priors were used, one would need to use the Metropolis–Hastings algorithm. The Gibbs sampler has two appealing properties compared with the Metropolis–Hastings algorithm.

Firstly, it is automatic: the user can simply apply it to their data and the algorithm will run straight "out of the box". By contrast, the Metropolis–Hastings algorithm requires, firstly a choice of blocking - should all the $\boldsymbol{\beta}$ parameters be updated simultaneously then all of the $\mathbf{Q}$ parameters and then the vector $\boldsymbol{\lambda}$, should all be updated at once or is some other choice preferable? Then, for the commonly used random walk proposal, for example, proposal scalings and variance matrices need to be chosen. These either require initial tuning runs or some form of adaptive MCMC (e.g. Roberts and Rosenthal, 2009) which in turn needs a choice of hyperparameters and careful watching.

Secondly, sampling from the conditional posterior explores the *whole* of the conditional state space for the parameter and with guaranteed acceptance. For example, with Gaussian observations, the covariate vector $\boldsymbol{\beta}$ is sampled directly from its conditional posterior whatever the dimension, $\mathsf{dim}(\boldsymbol{\beta})$. By contrast with a random walk Metropolis proposal the magnitude of the proposed jump in $\boldsymbol{\beta}$ must shrink in proportion to $1/\sqrt{\mathsf{dim}(\boldsymbol{\beta})}$ (e.g. Roberts and Rosenthal, 2001). The Metropolis-adjusted Langevin algorithm and Hamiltonian Monte Carlo have better scaling, but both require the computation of gradients of the log posterior.

2. Non-homogeneous transitions

Inherent in the use of a time-homogeneous hidden Markov model is that the waiting times between the changes of state of the hidden process are exponentially distributed and that the probability of a particular change given that a state change has occurred does not depend on the amount of time already spent in the current state. However, it is possible that this is not the case in reality. We explain why inference using both the timing of the events, $\tau_1, \tau_2, \ldots$ and the observations $o_1, o_2, \ldots$ is difficult in the general scenario, but then illustrate a particular case where a natural extension of our algorithm is possible: when waiting times have an Erlang distribution.

It is helpful to re-examine why relatively tractable inference, as described in Sections 3 and 4, has been possible. Transitions of the underlying Markov model occur according to constant-rate Poisson processes. For example, in a chain with two live states and a death state. If $X_0 = 1$ then there are two Poisson processes, one with rate $q_{1,2}$ and one with rate $q_{1,3}$, and whichever has the first event determines the next change of state. Crucially, we have also modeled the events in the observation process as arising from a Poisson process; in our running example, since $X_0 = 1$, this has a rate of λ_1. The fact that both transitions *and* observation events occurred according to Poisson processes allowed us to consider an extended state space and write down the generator in Section 3:

$$\begin{pmatrix} \mathbf{Q} - \boldsymbol{\Lambda} & \boldsymbol{\lambda} \\ \mathbf{0} & 0 \end{pmatrix}.$$

Exponentiating the generator produced transition probabilities that enabled us to employ the forward-backward algorithm to simulate the states of the hidden chain at the event times of the observation process. Maximum-likelihood and Bayesian inference for homogeneous hidden Markov models is covered in depth in Cappé et al. (2005) and there have been many advances for the homogeneous case since. Methods for Bayesian inference on the Markov modulated Poisson process are described in Fearnhead and Sherlock (2006); Rao and Teh (2013); maximum-likelihood inference on a combined homogeneous HMM with an informative observation process is performed via the EM algorithm in Lange et al. (2015) and Mews et al. (2023). To the best of our knowledge this article is the first to address Bayesian inference the combination.

One natural mechanism for allowing for different waiting distributions and time-dependent transition probabilities is by making the Poisson processes for the transitions non-homogeneous. There has been work in the sphere of inference for non-parametric Poisson intensities for a general non-Markovian model where the states themselves are observed (Titman, 2011; Kendall et al., 2024; Putter and Spitoni, 2018), maximum-likelihood for non-homogeneous hidden Markov models using numerical integration (Titman, 2023), and Bayesian estimation for a Poisson process with a non-parametric intensity (Adams et al., 2009; Alie et al., 2023). Whilst the general technique of particle MCMC (Andrieu et al., 2010) can always be employed as long as it is possible to simulate from the hidden process, it involves further choices such as the number of particles and the type of particle filter to be used. We would like to extend our relatively straightforward methodology to the non-homogeneous case, but there is no fixed rate matrix to exponentiate and so the transition probabilities between event times of the observation process are intractable (e.g., Kendall et al., 2024); moreover there would, in general, be no conjugate prior, so parameter inference would require the Metropolis–Hastings algorithm. We, therefore, see no natural extension of our method to the general case. However, it is possible to extend our method to the special case of a model where the timings for the potential state changes have Erlang distributions.

An Erlang distribution is a Gamma distribution with an integer shape parameter, k, and a rate parameter q. Hence, an $\mathsf{Erlang}(k, q)$ random variable is a sum of k independent $\mathsf{Exp}(q)$ random variables. Transitions can, therefore be modeling by adding $k-1$ interim states for each actual state transition. We illustrate this by considering the three-state model from Section 5, where State 3 corresponds to death and where there are no transitions back to lower-numbered states. We allow for a single interim transition for each state change and label the states with the partial transitions that have occurred in brackets. We also allow the first partial transition towards death to occur from either State 1 or State 2, so that the marginal time to death does not have an Erlang distribution unless $q_{1,3} = q_{2,3}$; we note that in the Markov chain in Section 5, the marginal time to death does not have an exponential distribution unless $q_{1,3} = q_{2,3}$. The generator is

$$
\mathbf{Q} = \begin{array}{r|ccccccc|}
 & 1 & 1(2) & 1(3) & 1(2,3) & 2 & 2(3) & 3 \\
\hline
1 & - & q_{1,2} & q_{1,3} & 0 & 0 & 0 & 0 \\
1(2) & 0 & - & 0 & q_{1,3} & q_{1,2} & 0 & 0 \\
1(3) & 0 & 0 & - & q_{1,2} & 0 & 0 & q_{1,3} \\
1(2,3) & 0 & 0 & 0 & - & 0 & q_{1,2} & q_{1,3} \\
2 & 0 & 0 & 0 & 0 & - & q_{1,3} & 0 \\
2(3) & 0 & 0 & 0 & 0 & 0 & - & q_{2,3} \\
3 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array},
$$

where for simplicity of presentation, the minus sign, $-$, in each row stands for the negative quantity that will make the entries in the row sum to 0. The corresponding vector for the rate of the observation process is $\boldsymbol{\lambda} = [\lambda_1, \lambda_1, \lambda_1, \lambda_1, \lambda_2, \lambda_2, 0]^\top$.

3. Continuous-path versus Discrete-time Likelihoods

In a CTHMM scenario it is possible to integrate out the hidden Markov chain between observation times to give the likelihood

$$
\mathcal{L}(\mathbf{Q}; x_{[0,\tau]}) = \prod_{t=1}^{T} \exp\left\{\mathbf{Q}\Delta_t\right\}_{x_{\tau_t}, x_{\tau_{t+1}}},
$$

where $\Delta_t = \tau_{t+1} - \tau_t$ and $\tau_{T+1} = \tau_e$. That is, we impute the states at each observation time but we do not fill in the full continuous-time Markov chain between states. In our setting, this could be adapted to include the information that there were no observation events in each inter-event interval and that there were events at the observation times:

$$
\mathcal{L}(\mathbf{Q}, \boldsymbol{\lambda}; x_{[0,\tau]}) = \left[\prod_{t=1}^{T-1} \exp\left\{(\mathbf{Q} - \boldsymbol{\Lambda})\Delta_t\right\}_{x_{\tau_t}, x_{\tau_{t+1}}} \lambda_{x_{\tau_{t+1}}}\right] \exp\left\{(\mathbf{Q} - \boldsymbol{\Lambda})(\tau_e - \tau_T)\right\}_{x_{\tau_T}, x_{\tau_e}}
$$

as the transition probability. Only imputing the states at observation times would reduce the amount of computation a little and still permits a Gibbs step for updating the $\mathbf{B}$ matrix. As mentioned in the discussion, the evolution process occurs in continuous time and so it is natural to impute the full continuous-time process; however, there is also a more pragmatic reason to do this. If one only imputed the states at event times, one would be forced to perform inference using Metropolis—Hastings for both $\mathbf{Q}$ and $\boldsymbol{\lambda}$, and deal with the issues that we mention earlier in the supplementary material. The full underlying Markov chain is exactly what is needed for a Gibbs sampler on $\mathbf{Q}$ and $\boldsymbol{\lambda}$, and that is why our algorithm samples the full underlying Markov chain.

4. Application: Additional results

The results from the four-state Hidden Markov model appear in Table 1, and those for the three-state model but with observation times no longer informative on the state are given in Table 2.

Table 1. Posterior median of four-state MMPP parameters associated with 95% credible intervals.

Parameter	State 1	State 2	State 3
λ (Marked point process rate)	4.82 (4.70, 4.96)	5.51 (5.40, 5.62)	7.71 (7.54, 7.89)
ν (initial state distribution)	0.33 (0.30, 0.37)	0.35 (0.30, 0.40)	0.32 (0.28, 0.36)
Expected number of drugs dispensed (GP)	0.70 (0.66, 0.75)	4.25 (4.21, 4.28)	8.66 (8.62, 8.70)
Expected number of drugs dispensed (ED)	0.43 (0.37, 0.50)	4.01 (3.94, 4.08)	8.65 (8.58, 8.73)
Expected number of drugs dispensed (HOSP)	0.51 (0.45, 0.58)	3.81 (3.73, 3.88)	8.48 (8.39, 8.56)
Expected number of drugs dispensed (SPEC)	0.51 (0.45, 0.58)	4.15 (4.07, 4.22)	8.65 (8.56, 8.74)
$q_{1.}$ (Transition rate from State 1)	–	0.28 (0.24, 0.32)	0.02 (0.01, 0.04)
$q_{2.}$ (Transition rate from State 2)	–	–	0.18 (0.17, 0.20)
$q_{.4}$ (Transition rate to death state)	0.07 (0.06, 0.09)	0.07 (0.06, 0.08)	0.23 (0.21, 0.26)

Table 2. Posterior median of three-state CTHMM (with observation timing not informative on the state) parameters associated with 95% credible intervals.

Parameter	State 1	State 2
ν (initial state distribution)	0.76 (0.73, 0.79)	0.24 (0.21, 0.27)
Expected number of drugs dispensed (GP)	3.59 (3.55, 3.62)	8.37 (8.33, 8.41)
Expected number of drugs dispensed (ED)	3.73 (3.65, 3.80)	8.81 (8.73, 8.90)
Expected number of drugs dispensed (HOSP)	3.32 (3.24, 3.40)	8.52 (8.43, 8.61)
Expected number of drugs dispensed (SPEC)	3.51 (3.43, 3.59)	8.45 (8.35, 8.55)
$q_{1.}$ (Transition rate from State 1)	–	0.16 (0.14, 0.18)
$q_{.3}$ (Transition rate to death state)	0.09 (0.08, 0.11)	0.04 (0.03, 0.05)

To calculate the marginal likelihood, integrating out the hidden Markov process, let $\mathbf{L}(t;\boldsymbol{\beta})$ be a diagonal matrix with $L_{k,k} = f(o_t|X_{\tau_t} = k)$. When the first event is guaranteed to occur at time 0, the full likelihood for a the set of event times and observations for a single person is:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{Q}, \boldsymbol{\nu}) \propto \boldsymbol{\nu}^\top \mathbf{L}(0;\boldsymbol{\beta}) \exp[(\mathbf{Q} - \boldsymbol{\lambda})\tau_1] \left\{ \prod_{t=1}^{T} \boldsymbol{\Lambda} \mathbf{L}(t;\boldsymbol{\beta}) \exp[(\mathbf{Q} - \boldsymbol{\Lambda})(\tau_{t+1} - \tau_t)] \right\} \mathbf{1},$$

where $\tau_{T+1} = t_e$, the end of the observation window. For each of the three- and four-state CTMC-MMPP models we evaluate this expression at the model's posterior mean parameter estimate. For the BIC calculations, the three-state model has 14 parameters and the four-state model has 23; the number of patients is $n = 1000$.

References

Adams, R. P., I. Murray, and D. J. C. MacKay (2009). Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, New York, NY, USA, pp. 9–16. Association for Computing Machinery.

Alie, R., D. A. Stephens, and A. M. Schmidt (2023). On data augmentation in point process models based on thinning procedures. *arXiv preprint arXiv:2203.06743*.

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(3), 269–342.

Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer New York, NY.

Fearnhead, P. and C. Sherlock (2006). An exact Gibbs sampler for the Markov-modulated Poisson process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(5), 767–784.

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*, 1–15.

Lange, J. M., R. A. Hubbard, L. Y. Inoue, and V. N. Minin (2015). A joint model for multistate disease processes and random informative observation times, with applications to electronic medical records data. *Biometrics 71*(1), 90–101.

Mews, S., B. Surmann, L. Hasemann, and S. Elkenkamp (2023). Markov-modulated marked Poisson processes for modeling disease dynamics based on medical claims data. *Statistics in Medicine 42*(21), 3804–3815.

Putter, H. and C. Spitoni (2018). Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen-Johansen estimator. *Statistical Methods in Medical Research 27*(7), 2081–2092.

Rao, V. and Y. W. Teh (2013). Fast MCMC sampling for Markov jump processes and extensions. *Journal of Machine Learning Research 14*(11).

Roberts, G. O. and J. S. Rosenthal (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science 16*(4), 351 – 367.

Roberts, G. O. and J. S. Rosenthal (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics 18*(2), 349–367.

Titman, A. (2023). *nhm: Non-Homogeneous Markov and Hidden Markov Multistate Models.*

Titman, A. C. (2011). Flexible nonhomogeneous markov models for panel observed data. *Biometrics 67*(3), 780–787.