Bayesian Inference for Latent Hawkes Processes

Scott W. Linderman∗
Columbia University
swl2133@columbia.edu

Yixin Wang∗
Columbia University
yixin.wang@columbia.edu

David M. Blei
Columbia University
david.blei@columbia.edu

Abstract

Hawkes processes are multivariate point processes that model excitatory interactions among vertices in a network. Each vertex emits a sequence of discrete events: points in time with associated content, or *marks*. Unlike Poisson processes, Hawkes processes allow events on one vertex to influence the subsequent rate of events on downstream vertices. With this property, they are ideally suited to model a variety of phenomena, like social network interactions, the spread of earthquake aftershocks, and spiking activity in neural circuits. This paper addresses a natural question: what if some vertices, marks, or time intervals are hidden from view? Such scenarios often arise in practice, as when neuroscientists work with recordings of subsets of neurons, or when social exchanges are not fully observed. These *latent* Hawkes processes pose a serious inferential challenge: we must perform inference in a model whose latent variables are marked point processes. We derive Bayesian inference and learning algorithms for latent Hawkes processes and demonstrate their efficacy on a variety of synthetic data and neural data.

1 Introduction

Many natural phenomena give rise to discrete sets of events in time and space. Social networks relay messages from one user to another, each message carrying a timestamp and some associated content such as a string of text, a picture, or a video. Neurons in the brain communicate via action potentials, or *spikes*, that induce post-synaptic potentials on downstream neurons, thereby influencing future spiking activity. Earthquakes trigger aftershocks, each characterized by a location and an amplitude. As scientists, our goal is to understand the properties of these systems, and the key to this understanding is a model of the data generating process. Common to all three of these examples is the notion of *mutual excitation*: events beget future events. Hawkes processes [1] capture this property.

Hawkes processes are defined by a network and a point process observation model. The vertices of the network correspond to users in the social network or neurons in the brain, and each vertex gives rise to a set of events in time, which may also include corresponding content, or marks. For each vertex, the marked events are modeled as a draw from a point process with an underlying rate. The network structure plays a role in determining that rate, where the edges specify how events on one vertex influence the future rate of events on downstream vertices. Specifically, the edges are characterized by excitatory spatiotemporal impulse responses; for each event, an impulse response is added to the rate of all connected vertices. These excitatory interactions distinguish the Hawkes process from the standard Poisson process and enable it to capture excitatory spatiotemporal phenomena, like those we described above.

In this work, we consider the case where some marks, vertices, or time intervals are hidden. Such scenarios often arise in practice: due to experimental limitations, it may not be possible to observe an entire population of neurons simultaneously or to accurately identify all seismic events. In social network data, some messages may be encrypted or partially redacted, leading to missing threads of

∗ Authors contributed equally.

Advances in Approximate Bayesian Inference Workshop, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

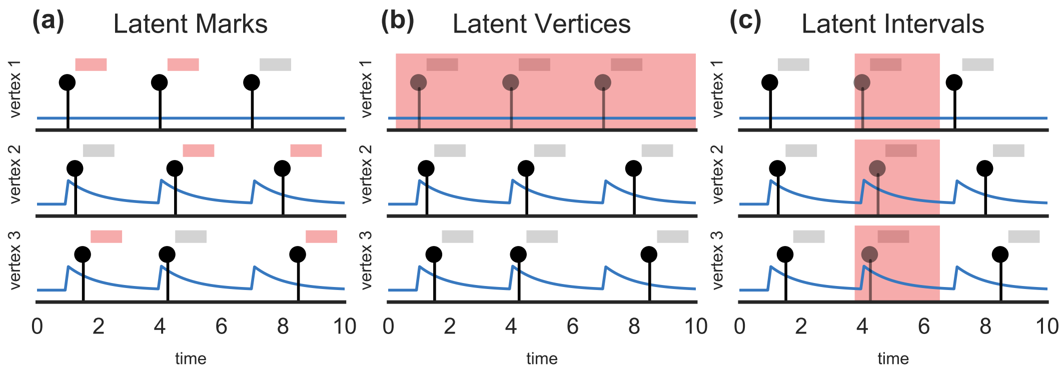

Figure 1: *Various types of latent Hawkes processes.* Black spikes denote events and gray boxes denote corresponding marks, if they are observed. The temporal component of the underlying rate, $\lambda_v(t, y)$, is shown in blue. Here, vertex 1 has a constant background rate, and each event on vertex 1 induces impulse responses on vertices 2 and 3. Red shading denotes the various types missingness: (a) We observe the times of all events, but some marks are hidden. (b) All events on the first vertex are latent; this is known as a Neyman-Scott process. (c) Some intervals of time are occluded.

communication. Our goal in these situations is to not only infer the missing data, but to also learn the network of interactions that give rise to it.

To that end, we develop the *latent Hawkes process*. We derive a stochastic expectation maximization algorithm for Bayesian learning and inference, using expectations of the latent data to update our estimate of the global network parameters. Hawkes processes pose a unique inferential challenge: the latent data assume the form of a point process with an unknown number of events and potentially high-dimensional marks. Our principal contribution is an inference method for latent marked Hawkes processes. In particular, we show how inference can be performed with data-driven and Rao-Blackwellized sequential Monte Carlo algorithms, leveraging variational inference in the inner loop to approximate the particle weights. See Appendix A for related work.

2 Latent Hawkes Processes

While Poisson processes are foundational models for spatiotemporal data [2], many real-world systems violate the assumption of independent intervals. Hawkes processes [1] remedy this shortcoming by allowing events to influence the future rate. They do so via a *conditional intensity function* over time and marks, $\lambda_v(t, y \,|\, \mathcal{H}_t)$, that depends upon the event history, $\mathcal{H}_t = \{v_n, t_n, y_n : t_n < t\}$, where $v_n \in \{1, \ldots, V\}$ denotes the vertex on which the n-th event occurred, $t_n \in [0, T]$ denotes its time, and $y_n \in \mathcal{Y}$ denotes its mark. The process has a parameterized background rate, $b_v(t, y; \theta)$, and on top of this baseline, each event adds a parametric, nonnegative *impulse response* $f_{v_n \to v}(t, y; t_n, y_n, \theta)$ to the subsequent rate,

$$\lambda_v(t, y \,|\, \mathcal{H}_t, \theta) = b_v(t, y;\, \theta) + \sum_{n=1}^{N} f_{v_n \to v}(t, y;\, t_n, y_n, \theta)\, \mathbb{I}[t > t_n].$$

Given this conditional intensity function, the log likelihood of a set of events decomposes into two terms: the negative integrated rate and the sum of instantaneous log rates,

$$\log p(\{v_n, t_n, y_n\} \,|\, \theta) = -\sum_{v=1}^{V} \int_0^T \int_{\mathcal{Y}} \lambda_v(t, y \,|\, \mathcal{H}_t, \theta)\, \mathrm{d}t\, \mathrm{d}y + \sum_{n=1}^{N} \log \lambda_{v_n}(t_n, y_n \,|\, \mathcal{H}_{t_n}, \theta).$$

Previous work has provided a variety of methods for learning the parameters θ in the context of fully-observed data [e.g. 3, 4] but in the partially-observed case, we must perform joint inference of both the model parameters and the latent data. We distinguish various types of missing data, as they warrant different inferential approaches.

Missing marks. Consider a spy eavesdropping on trans-Atlantic messages. Some of the messages are transmitted in plain-text with the sender, receiver, and content of the message intact; others are encrypted so that only the source and destination are observed. Her mission—should she choose to accept it—is to infer the likely contents of those encrypted messages based on patterns of communication in the plain-text messages. She may model this as a latent Hawkes process in which the times and vertices are observed but the marks are partially missing, as illustrated in Fig. 1a. Let t_n denote the time of the n-th message, v_n specify its sender, and $y_n = (r_n, c_n)$ specify the receiver and the contents of the message, respectively. In some messages, c_n is missing, as indicated by the red boxes in Fig. 1a.

Latent vertices. A more interesting type of missingness is when entire vertices are obscured. In this case, we must infer the latent timestamps as well as their marks. One of our motivating examples comes from the field of document analysis. We have a dataset of diplomatic cables sent between embassies over the course of a decade, and we are interested in modeling these documents to learn about communication patterns. While much of the communication is routine, there are many messages in response to real-world events, like the Entebbi raid in Uganda, which coincided with the bicentennial of the United States on July 4, 1976. These two events generated a flurry of atypical diplomatic cables.

This can be viewed as a Hawkes process with one latent vertex (for world events) and V observed vertices (the embassies) connected by a special network structure: the latent, world event vertex only influeces the embassy vertices, not the other way around. This is shown in Fig. 1b. If we further restrict this such that the embassies are *only* driven by world events—i.e. such that there are no embassy-to-embassy interactions—this would be an instance of a Neyman-Scott process [5].

Latent intervals. Finally, consider a neuroscientist seeking to map out synaptic connections in cortex. His recording tools enable high-fidelity recordings of neural activity, but since he can only focus his microscope on a limited region of tissue, he can only record from a subset of cells at a time. By repositioning the microscope over the course of the experiment, he records a variety of different subsets. In this example, the vertices are the neurons and the events are the timestamps of spikes measured on the observed neurons.

The weights of the Hawkes process may be taken as a measure of connection strength between a directed pair of neurons. Since the microscope is moved throughout the experiment, there are intervals in which no data is collected from certain neurons. However, in order to estimate the connectivity, this missing data must be taken into account; otherwise, spurious correlations could arise if two neurons are driven by the same upstream, latent cell. Fig. 1c illustrates this type of latent Hawkes processes. We study this problem in Section C.2.

3 Bayesian Learning and Inference

Learning and inference in latent Hawkes processes is fundamentally a latent variable problem. As such, we start with an expectation-maximization algorithm [6] that alternates between *inference*—computing expected log likelihoods—and *learning*—taking gradients with respect to the model parameters. The critical computation is the expectation of the complete data log joint probability, $\mathcal{L}(\theta) = \mathbb{E}_{p(\boldsymbol{z} \mid \boldsymbol{x}, \theta_{\mathsf{curr}})}[\log p(\boldsymbol{x}, \boldsymbol{z} \mid \theta)]$, where $\boldsymbol{z}$ are the latent marks and/or timestamps, $\boldsymbol{x}$ are the observed marks and/or timestamps, and θ are the model parameters that govern the impulse responses, for example. Given this expectation, we follow $\nabla_\theta \mathcal{L}$ to perform a stochastic gradient step, update the global parameter estimate, and iterate.

The unique challenge of latent Hawkes processes is that *the latent variables take the form of a marked point process*. Specifically, since the number of latent events is undetermined, we propose a variety of methods for performing inference over sets of unknown cardinality. Monte Carlo methods are well-suited to this challenge, stochastically simulating sets of size. In particular, the causal, temporal structure of the latent Hawkes process naturally suggests sequential Monte Carlo (SMC) methods [7–9].

Data-Driven Sequential Monte Carlo. We develop a sequential Monte Carlo (SMC) [7–9] approach. It leverages the autoregressive nature of Hawkes processes (the instantaneous rate is only a function of preceding events) to sequentially propose and resample particles. First, let $\{s_i\}_{i=0}^{I}$, $s_0 = 0$, $s_i < s_{i+1}$, $s_I = T$, define a *scaffold* that partitions the time range $[0, T]$ into I disjoint intervals. The value of the p-th particle in the i-th interval, $z_i^{(p)}$, is a set of latent events: $z_i^{(p)} = \{(v_n^{(p)}, t_n^{(p)}, y_n^{(p)}) : s_{i-1} < t_n^{(p)} \leq s_i\}$. Contrast this with the set of observed events x_i for the same interval. We only propose latent events for vertices whose data is missing in that interval.

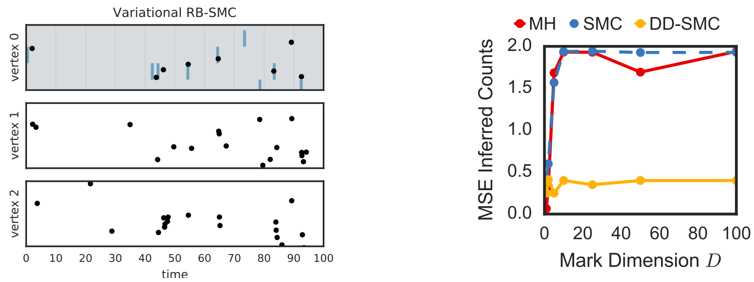

Figure 2: *Recovering marked data with SMC. Left:* An example of a latent Hawkes process with a single hidden vertex driving twenty observed vertices, all with 100-dimensional marks. Yellow shading denotes the inferred posterior over times and marks using Rao-Blackwellized SMC. *Right:* As the dimension of the marks increases, standard MH and SMC fail, while data-driven SMC provides accurate inferences even for high dimensional data.

We generate a candidate set of latent events for interval i by sampling a proposal distribution, $z_i^{(p)} \sim r(z_i \mid x_{1:i}, z_{1:i-1}^{(p)}, \theta)$, and weighting the newly updated particles with the function,

$$\omega(z_{1:i}^{(p)}) = \frac{p(x_{1:i}, z_{1:i}^{(p)} \mid \theta)}{p(x_{1:i-1}, z_{1:i-1}^{(p)} \mid \theta)\; r(z_i^{(p)} \mid x_{1:i}, z_{1:i-1}^{(p)}, \theta)}.$$

The key design choice is the proposal distribution. In particular, the high dimensionality of the marks calls for delicate choices of the proposal distribution to control the variance of the SMC estimates. To this end, we utilize data-driven proposals, leveraging our intuition that latent marks are often similar to observed marks.

Rao-Blackwellized Sequential Monte Carlo. Alternatively, note that in the un-marked case, inference of the latent times is relatively simple, and standard SMC works well. This motivates a *Rao-Blackwellized* approach, in which we marginalize the latent marks and infer only the timestamps and vertices. The weights are then given by $\omega(\widetilde{z}_{1:i}^{(p)})$ where $\widetilde{z}_i^{(p)}$ denotes the particles without marks. (Note that here the proposal is a density on $\widetilde{z}_i$.)

There is a caveat here though: the weights now need the marginal likelihood $p(x_{1:i}, \widetilde{z}_{1:i}^{(p)}, \theta)$, which requires an intractable integral over marks. We propose to use a variational lower bound on the marginal likelihood instead. Specifically, we optimize a parametric variational distribution $q(y_{1:i}^{(p)}; \eta) \approx p(y_{1:i}^{(p)} \mid x_{1:i}, \widetilde{z}_{1:i}^{(p)}, \theta)$; the ELBO $\mathcal{L}(\eta)$ gives our lower bound. This approximation biases our SMC estimates, but the Rao-Blackwellization should reduce its variance. In other words, we trade bias in variational approximation for lower variance due to Rao-Blackwellization. We can view this as variational inference by defining a surrogate ELBO: $\prod_{i=1}^{T} \mathbb{E}(\frac{1}{P} \sum_{p=1}^{P} \omega(\widetilde{z}_{1:i}^{(p)})$.

4 Experiments

We simulated a latent Hawkes process with one latent process and three observed processes—i.e. a Neyman-Scott process—with Gaussian marks $y_n \in \mathbb{R}^D$. Figure 2a illustrates a draw from this model: black dots denote time, and the y position denotes the first mark coordinate. We inferred the posterior distribution of latent events with data-driven SMC and Rao- Blackwellized. For comparison, we also used standard SMC and a simple Metropolis-Hastings MCMC algorithm. The yellow shading in Figure 2a shows the posterior distribution of the first mark coordinate using the Rao-Blackwellized approach; Figure 2b shows that as the dimensionality of the marks increases, naïve proposals fail to provide good estimates, and the data-driven approach becomes necessary. We show further experimental results in the appendix.

We studied latent Hawkes processes—marked, multivariate point processes with missing event times, vertices, or marks. The countably infinite sets of latent time stamps and marks posed severe challenges not faced in standard Hawkes processes. We proposed Bayesian algorithms for learning and inference in this challenging setting. To demonstrate the efficacy of our approach, we evaluated our methods using both simulated data and real neural data (see Appendix C) under a variety of missingness regimes. In future work, we will explore variational sequential Monte Carlo [10]—a variational method for learning proposals—as an alternative to and an extension of Rao-Blackwellized SMC.

References

[1] Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83, 1971.

[2] Peter J Diggle. *Statistical analysis of spatial and spatio-temporal point patterns*. CRC Press, 3rd edition, 2013.

[3] Aleksandr Simma and Michael I Jordan. Modeling events with cascades of Poisson processes. *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2010.

[4] Scott W Linderman and Ryan P Adams. Discovering latent network structure in point process data. In *Proceedings of The 31st International Conference on Machine Learning*, pages 1413–1421, 2014.

[5] Jerzy Neyman and Elizabeth L Scott. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(1):1–43, 1958.

[6] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[7] Jun S Liu and Rong Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American statistical association*, 93(443):1032–1044, 1998.

[8] Pierre Del Moral. Non-linear filtering: interacting particle resolution. *Markov processes and related fields*, 2(4):555–581, 1996.

[9] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.

[10] Christian A Naesseth, Scott W Linderman, Rajesh Ranganath, and David M Blei. Variational sequential Monte Carlo. *arXiv preprint arXiv:1705.11140*, 2017.

[11] Ke Zhou, Hongyuan Zha, and Le Song. Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 16, 2013.

[12] Manuel Gomez-Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1019–1028, 2010.

[13] Charles Blundell, Jeff Beck, and Katherine A Heller. Modelling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems*, pages 2600–2608, 2012.

[14] Fangjian Guo, Charles Blundell, Hanna Wallach, and Katherine Heller. The Bayesian echo chamber: Modeling social influence via linguistic accommodation. In *Artificial Intelligence and Statistics*, pages 315–323, 2015.

[15] Mehrdad Farajtabar, Yichen Wang, Manuel Gomez Rodriguez, Shuang Li, Hongyuan Zha, and Le Song. Coevolve: A joint point process model for information diffusion and network co-evolution. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 1954–1962. Curran Associates, Inc., 2015.

[16] Liangda Li, Hongbo Deng, Anlei Dong, Yi Chang, and Hongyuan Zha. Identifying and labeling search tasks via query-based hawkes processes. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 731–740, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2956-9.

[17] Xinran He, Theodoros Rekatsinas, James Foulds, Lise Getoor, and Yan Liu. Hawkestopic: A joint model for network inference and topic modeling from text-based cascades. In *International Conference on Machine Learning*, pages 871–880, 2015.

[18] Yichen Wang, Nan Du, Rakshit Trivedi, and Le Song. Coevolutionary latent feature processes for continuous-time user-item interactions. In *Advances in Neural Information Processing Systems*, pages 4547–4555, 2016.

[19] Hanjun Dai, Yichen Wang, Rakshit Trivedi, and Le Song. Deep coevolutionary network: Embedding user and item features for recommendation. 2017.

[20] Ushio Tanaka and Yosihiko Ogata. Identification and estimation of superposed Neyman–Scott spatial cluster processes. *Annals of the Institute of Statistical Mathematics*, 66(4):687–702, 2014.

[21] David A Moore and Stuart J Russell. Signal-based Bayesian seismic monitoring. *arXiv preprint arXiv:1703.00561*, 2017.

[22] Sindhu Ghanta, Jennifer G Dy, Donglin Niu, Michael I Jordan, et al. Latent marked Poisson process with applications to object segmentation. *Bayesian Analysis*, 2016.

[23] Chris Lloyd, Tom Gunter, Michael Osborne, and Stephen Roberts. Variational inference for Gaussian process modulated Poisson processes. In *International Conference on Machine Learning*, pages 1814–1822, 2015.

[24] Vinayak Rao, Ryan P Adams, and David D Dunson. Bayesian inference for matérn repulsive processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3): 877–897, 2017.

[25] Peter J Green. Reversible jump Markov chain Monte Carlo computation and bayesian model determination. *Biometrika*, pages 711–732, 1995.

[26] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[27] Daniel Soudry, Suraj Keshri, Patrick Stinson, Min-hwan Oh, Garud Iyengar, and Liam Paninski. Efficient" shotgun" inference of neural connectivity from highly sub-sampled activity data. *PLoS Comput Biol*, 11(10):e1004464, 2015.

A Related Work

While Hawkes processes were introduced over four decades ago [1], they have witnessed a resurgence of interest from the machine learning community. A number of recent works have addressed *learning*, the problem of inferring the parameters of the background rates and impulse response functions [3, 11, 12, 4, 13, 14]. These works have generalized Hawkes processes in a number of ways, incorporating them into larger probabilistic models with hierarchical and dynamic network structure [4, 13, 15] and text diffusion models [16, 17, 14]. Our work is orthogonal to these, focusing instead on the case of latent variable inference in the face of missing data. The methods we develop could fruitfully be incorporated into these broader modeling frameworks. Previous work has also used Hawkes processes for latent user-item features for recommendation systems [18, 19]; however, in this work, only the marks are missing. Our work moves beyond this setting; we consider the general class of latent Hawkes processes that can depend on both observed and unobserved events with unknown times and marks.

A key component of our proposed algorithms is inference over latent stochastic processes, for which many Monte Carlo methods have been developed. Tanaka and Ogata [20] used nearest neighbor based likelihood functions to infer parameters of multiple superimposed Neyman-Scott processes. Recently, Moore and Russell [21] used reversible jump MCMC to infer a latent homogeneous Poisson process that drives Gaussian processes over wavelet parameters for seismic monitoring, and Ghanta et al. [22] developed a hybrid Gibbs sampler to infer a latent marked Poisson process for image segmentation. Lloyd et al. [23] proposed a variational inference algorithm for Gaussian process modulated Poisson processes. Rao et al. [24] studied Matern type-III repulsive processes, resulting in a novel MCMC algorithm, which also incorporated inference over latent point processes. The Monte Carlo methods underlying our method are shared by these previous works, but here we tailor them to inference in latent Hawkes processes with missing data, and we show how to leverage the special structure of these models to develop efficient learning and inference algorithms.

B MCMC Methods for Latent Events

Importance Sampling. This is the simplest of the methods we consider. We estimate expectations with respect to the posterior distribution of latent events by independently sampling *particles* $\{\boldsymbol{z}^{(p)}\}_{p=1}^{P}$ from a proposal distribution $r(\boldsymbol{z} \mid \boldsymbol{x}, \theta_{\mathsf{curr}})$, weighting them according to the ratio of the joint probability over the proposal probability, and using them to compute a Monte Carlo estimate of the expected log likelihood. A simple choice of proposal distribution is a homogeneous Poisson process with the model's background rate. This self-normalized importance sampling estimate is biased for finite numbers of particles, but still provides a reasonable baseline for our other methods.

Reversible Jump MCMC. A more sophisticated approach is to run a Markov chain to sample sets of latent events. Following previous work Moore and Russell [21], we use a reversible jump Metropolis-Hastings algorithm [25] consisting of three proposals: (i) add a new latent event to a randomly chosen missing interval; (ii) remove one latent event uniformly at random; and (iii) jitter one latent event's time and mark slightly. These operations are reversible, and their forward and backward proposal probabilities are easily calculated. The key to this approach is tuning hyperparameters, such as the number of "burn-in" iterations before samples are collected, since the early samples will be heavily biased by the initial distribution.

Sequential Monte Carlo. Finally, we develop a sequential Monte Carlo (SMC) [7, 8] approach. It builds on importance sampling but leverages the autoregressive nature of Hawkes processes (the instantaneous rate is only a function of preceding events) to sequentially simulate and resample particles. Given latent event particles, $\boldsymbol{z}_{1:i-1}^{(p)}$, and observed events $\boldsymbol{x}_{1:i-1}$, for intervals up to and including $i-1$, we draw a new set of latent events for interval i by sampling a proposal distribution, $\boldsymbol{z}_i^{(p)} \sim r(\boldsymbol{z}_i \mid \boldsymbol{x}_{1:i}, \boldsymbol{z}_{1:i-1}^{(p)}, \theta)$, and weighting the newly updated particles by the conditional probability of the i-th interval,

$$\omega_i^{(p)} = \frac{p(\boldsymbol{x}_i, \boldsymbol{z}_i^{(p)} \mid \boldsymbol{x}_{1:i-1}, \boldsymbol{z}_{1:i-1}^{(p)}, \theta)}{r(\boldsymbol{z}_i^{(p)} \mid \boldsymbol{x}_{1:i}, \boldsymbol{z}_{1:i-1}^{(p)}, \theta)} = \frac{p(\boldsymbol{x}_{1:i}, \boldsymbol{z}_{1:i}^{(p)} \mid \theta)}{p(\boldsymbol{x}_{1:i-1}, \boldsymbol{z}_{1:i-1}^{(p)} \mid \theta)\, r(\boldsymbol{z}_i^{(p)} \mid \boldsymbol{x}_{1:i}, \boldsymbol{z}_{1:i-1}^{(p)}, \theta)}.$$

Since the Hawkes process is autoregressive, the conditional probability of partial data up to the i-th interval is easy to compute. Moreover, note that the proposal may depend on observed data.

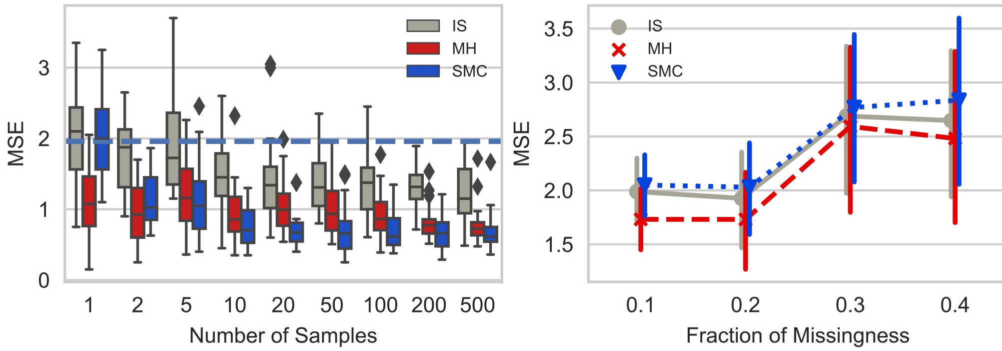

Figure 3: *Learning and inference in a synthetic latent Hawkes process (100 replications). Left:* Mean squared error of the number of events in each missing interval as the number of Monte Carlo samples is varied for each estimator, given the true model parameters. The dashed line is proposals from the background rate. The MCMC algorithm in our framework consistently outperforms. *Right:* Mean squared error (with 95% credible interval) of the number of events in each missing interval as a function of the percentage of missing data. Predictions by latent Hawkes processes have small mean squared error.

The critical component that distinguishes SMC from importance sampling is the *resampling* step. After each interval is updated, we normalize the particle weights and resample from their empirical distribution. This replicates particles with large weights and removes particles that are inconsistent. As we will see next, accurate proposals can yield significant performance improvements.

C Experiments

We demonstrate the potential of latent Hawkes processes in a variety of settings and missingness patterns. We first consider a case in which the underlying dynamics truly follow Hawkes processes. By masking off some marks, vertices, or time intervals, we show that fitting latent Hawkes processes models can recover the truth. Moreover, we apply these models to neural data; latent Hawkes processes are able to fill in the missing truth accurately.

C.1 Unmarked Synthetic Experiments

We first simulated a Neyman-Scott process of $V = 10$ nodes and 100 time steps with random parameters. We mask off the first node completely. We then simulated an unmarked Hawkes process of $V = 10$ nodes and 100 time steps with random parameters. We use a scaffold with uniformly-spaced intervals of 5 time steps and randomly chose $p = 10\%, 20\%, 30\%, 40\%$ of the intervals to be masked.

We consider the tasks of learning and inference separately. First, we assess the accuracy of the proposed latent event inference methods—importance sampling, Metropolis-Hastings, and SMC—in terms of their ability to estimate the number of events in each latent interval, as measured by mean squared error. We evaluate this performance as a function of the number of Monte Carlo samples generated (number of MCMC steps or particles). Figure 3a shows that the MCMC algorithms in our framework consistently outperform the baseline proposals from the background rate, and that Metropolis-Hastings and SMC perform best of all. The mean squared error stabilizes close to zero after 20 samples.

Second, we measure the accuracy of our stochastic EM algorithm in predicting events in missing intervals. In all learning experiments, we use Adam [26] with a step-size of 0.1 for stochastic optimization. We again use the three inference methods to approximate the expected log likelihood, but here we measure performance as a function of the percentage of missing data. Figure 3b shows that our inference algorithm predicts accurately up to as large as 40% of missingness. Predictions by latent Hawkes processes have small mean squared error. They are usually $\pm \sim 1$ event for each interval. Metropolis-Hastings algorithms consistently give the best prediction performance.

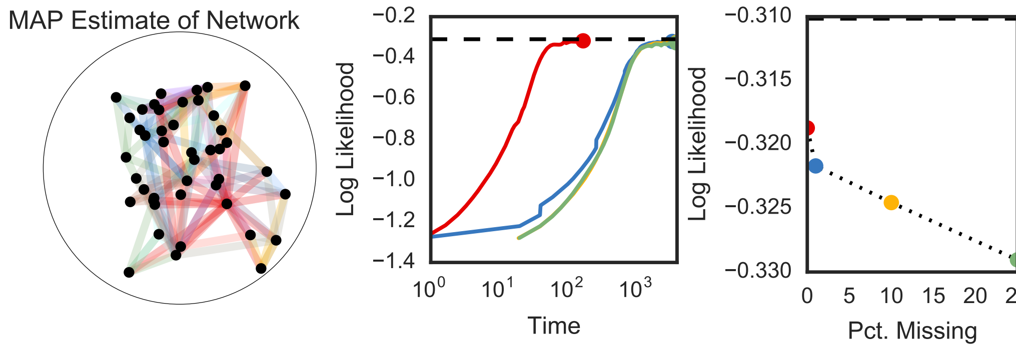

Figure 4: *Analysis of hippocampal place cells. Right:* Inferred network of interactions. Each black dot is positioned at the center of a cell's place field. The edges correspond to the top 10% of inferred weights, and they are color coded by presynaptic neuron. *Center:* Log likelihood as a function of computation time for increasing amounts of missing data (red=0%; blue=1%; yellow=10%; green=25%). *Right:* Final log likelihood as a function of percentage of missing data.

C.2 Analysis of Neural Data

Finally, we study a data set of $T = 1460$s (24.3min) of neural spiking activity from $V = 44$ hippocampal place cells, special neurons that fire in localized regions of space.[1] The dataset contains a total of 57,114 spikes with 1,298±983 spikes per neuron. Figure 4a shows the place-field centers for all cells, along with the inferred network from a Hawkes process fit to the complete dataset. As described in Section 2, we are motivated by recent work on estimating functional connectivity given partially obscured datasets that arise in practice due to experimental limitations [27]. To simulate this type of data, we used a scaffold with uniformly-spaced 10s intervals and we randomly mask increasing percentages of data.

The method of Soudry et al. [27] is for discrete data under a nonlinear autoregressive model, and is therefore not immediately comparable; however, we can again use the standard Hawkes process fitting using maximum likelihood on the complete data to provide an upper bound on the performance of our stochastic algorithms.

Figure 4b illustrates the log likelihood of the data as a function of wall-clock time for increasing amounts of missing data. The batch maximum likelihood estimate is shown in black; in red is the estimate of our stochastic EM algorithm with no missing data, in which case it reduces to stochastic gradient ascent. The blue, yellow, and green lines show performance with 1%, 10%, and 25% missing data, respectively. All converge to within 0.02nats/spike of the batch log likelihood, as shown in 4c, though performance steadily degrades with missing data. These promising results suggest that key properties of this cell population could be derived, even with substantial amounts of missing data.

[1] We thank the Wilson lab, MIT, for sharing this dataset.