4

Discrete-Time Linear Autoregressive Poisson Models

This chapter builds on the network Hawkes model introduced in the Chapter 3. We introduce linear autoregressive Poisson models — the discrete time analogue of the Hawkes process — and we derive efficient Gibbs sampling and stochastic variational inference algorithms, leveraging the Poisson superposition principle as before. This chapter marks the transition from continuous time models to the discrete time models that occupy this and subsequent chapters. As we will see, these discrete time formulations are in some ways easier to work with. We can easily extend them to non-Poisson spike count models, and we can interface with a diverse array of probabilistic matrix decomposition models. However, the discrete nature of spike counts still poses some serious inferential hurdles, which this thesis aims to overcome.

In addition to bridging from continuous to discrete, this chapter also addresses issues of computational complexity. The complexity of our Hawkes process inference algorithm scaled, in the worst case, quadratically with the number of spikes, since we had to sample a "parent" for each spike. By designing block parallel Gibbs updates, we were able to obtain linear complexity in the number of spikes. However, when the firing rates are high, this is still the bottleneck of our algorithm. Here, we reduce this complexity to be independent of the number of spikes by adopting a discrete time approach. Moreover, we derive efficient *stochastic* variational inference algorithms (Hoffman et al., 2013) that work with subsets of time bins in each iteration and thereby scale to massive datasets.

4.1 Probabilistic Model

The fundamental limitation of the previously developed continuous time models is that the domain of the auxiliary variable, ω_m, grows with the number of events which occurred before time s_m. For datasets with high rates of activity, this can quickly become the limiting factor of the inference algorithm. At the same time, it is often reasonable to assume that events do not interact on time scales shorter than Δt. This motivates a discrete time formulation in which we group events into bins of width Δt and ignore potential interactions between events in the same bin. Then the rate becomes,

$$\begin{aligned}\lambda_{t,n} &= \lambda_n^{(0)} + \sum_{n'=1}^{N} \sum_{d=1}^{D} s_{t-d,n'} \cdot h_{n'\to n}[d], \\ s_{t,n} &\sim \text{Poisson}(\lambda_{t,n} \cdot \Delta t),\end{aligned} \tag{4.1}$$

where $s_{t,n}$ is the number of spikes fired by neuron n in the t-th time bin and $h_{n'\to n}[d]$ is an impulse response function describing the influence that events on neuron n' have on the rate of process n at discrete time lag d. As we will show, under this formulation the auxiliary variables only assume a fixed set of values independent of the rate.

As in the last chapter, we introduce a network model as a prior distribution over the impulse response weights. Following the approach of the previous chapter, we decompose the impulse response function into the product of a binary variable that specifies whether or not a connection exists, a scalar weight that specifies the strength of the interaction if present, and a probability mass function that specifies the time course of interaction:

$$h_{n\to n'}[d] = a_{n\to n'} \cdot w_{n\to n'} \cdot \hbar[d;\, \boldsymbol{\theta}_{n\to n'}]$$

for $d \in \{1, \ldots, D\}$. The function $\hbar[d] : \{1, \ldots, D\} \to [0, 1]$ is now a probability mass function, which we model as a convex combination of normalized basis functions, $\boldsymbol{\phi}_b$,

$$\hbar[d;\, \boldsymbol{\theta}_{n\to n'}] \triangleq \sum_{b=1}^{B} \theta_{n\to n'}^{(b)} \cdot \phi_b[d],$$
$$\sum_{d=1}^{D} \phi_b[d] \cdot \Delta t = 1,$$
$$\sum_{b=1}^{B} \theta_{n\to n'}^{(b)} = 1.$$

We enforce the latter constraint with a Dirichlet prior $\boldsymbol{\theta}_{n\to n'} \sim \mathrm{Dir}(\boldsymbol{\gamma})$. The basis functions are typically taken to be normalized Gaussian bumps or rectified cosine functions spaced over the interval $1, \ldots, D$. For example, in the following experiments we used,

$$\widetilde{\phi}_b[d] = \exp\left\{-\frac{1}{2\sigma^2}(d-\mu_b)^2\right\},$$
$$\phi_b[d] = \frac{\widetilde{\phi}_b[d]}{\Delta t \sum_{d'=1}^{D} \widetilde{\phi}_b[d']},$$

with means, μ_b, evenly spaced on $[1, D]$, and $\sigma = \frac{D}{B-1}$.

Plugging this impulse response model into Eq. 4.1 yields,

$$\begin{aligned}\lambda_{t,n'} &= \lambda_{n'}^{(0)} + \sum_{n=1}^{N}\sum_{b=1}^{B} a_{n\to n'} \cdot w_{n\to n'} \cdot \theta_{n\to n'}^{(b)} \sum_{t'=1}^{t-1} s_{t',n} \cdot \phi_b[d] \\ &= \lambda_{n'}^{(0)} + \sum_{n'=1}^{N}\sum_{b=1}^{B} a_{n\to n'} \cdot w_{n\to n'} \cdot \theta_{n\to n'}^{(b)} \cdot \widehat{s}_{t,n,b},\end{aligned}$$

where

$$\widehat{s}_{t,n,b} \triangleq \left(\boldsymbol{s}_n * \boldsymbol{\phi}_b\right)[t]$$

is the discrete convolution of the n-th spike train with the b-th basis function evaluated at the t-th time bin. Since both the spike trains and the basis functions are given, these can be precomputed.

4.2 Inference with Gibbs Sampling

As before, we begin by introducing auxiliary parent variables for each entry $s_{t,n}$. By the superposition theorem for Poisson processes, each event can be attributed to either the background rate or one of the impulse responses.

Let $\omega_{t,n'}^{(n,b)} \in \{0, \ldots, s_{t,n'}\}$ denote how many of the events that occurred in the t-th time bin on the n'-th neuron are attributed to the b-th basis function of the n-th neuron. Similarly, let $\omega_{t,n'}^{(0)}$ denote the number of events attributed to the background process. We combine these auxiliary variables into vectors, $\boldsymbol{\omega}_{t,n'} \triangleq \left[\omega_{t,n'}^{(0)}, \omega_{t,n'}^{(1,1)}, \ldots, \omega_{t,n'}^{(N,B)}\right]$.

Due to the Poisson superposition principle, these parent variables are conditionally multinomial distributed. For time t and neuron n', we resample

$$\boldsymbol{\omega}_{t,n'} \sim \mathrm{Mult}\left(s_{t,n'}, \boldsymbol{u}_{t,n'}\right) \quad u_{t,n'}^{(0)} = \frac{\lambda_{n'}^{(0)}[t]}{\lambda_{n'}[t]}, \quad u_{t,n'}^{(n,b)} = \frac{\hat{s}_{t,n,b} \cdot a_{n\to n'} \cdot w_{n\to n'} \cdot \theta_{n\to n'}^{(b)}}{\lambda_{n'}[t]}.$$

Given this attribution, the likelihood factorizes into a product of Poisson distributions,

$$\begin{aligned} p(\boldsymbol{\omega} \mid \boldsymbol{\lambda}) = &\left[\prod_{t=1}^{T}\prod_{n'=1}^{N} \mathrm{Poisson}(\omega_{t,n'}^{(0)} \mid \lambda_{n'}^{(0)}\Delta t)\right] \\ &\times \left[\prod_{t=1}^{T}\prod_{n=1}^{N}\prod_{n'=1}^{N}\prod_{b=1}^{B} \mathrm{Poisson}(\omega_{t,n'}^{(n,b)} \mid \hat{s}_{t,n,b} \cdot a_{n\to n'} \cdot w_{n\to n'} \cdot \theta_{n\to n'}^{(b)} \cdot \Delta t)\right]. \end{aligned}$$

Gibbs sampling the background rates. We use conjugate priors for the constant background rates, weights, and impulse responses. For the constant background rates we have, $\lambda_{n'}^{(0)} \sim \mathrm{Gamma}(\alpha_\lambda, \beta_\lambda)$, which results in the conditional distribution

$$\begin{aligned} \lambda_{n'}^{(0)} \mid \{\omega_{t,n'}^{(0)}\} &\sim \mathrm{Gamma}(\alpha_\lambda^{(n)}, \beta_\lambda^{(n)}), \\ \alpha_\lambda^{(n)} &= \alpha_\lambda + \sum_{t=1}^{T} \omega_{t,n'}^{(0)}, \\ \beta_\lambda^{(n)} &= \beta_\lambda + T\Delta t\,. \end{aligned}$$

Gibbs sampling impulse responses. The likelihood of the impulse responses, $\boldsymbol{\theta}_{n\to n'}$ is proportional to a Dirichlet distribution. Combined with a Dirichlet($\boldsymbol{\gamma}$) prior this yields

$$\boldsymbol{\theta}_{n\to n'} \mid \{\omega_{t,n}^{(n',b)}\}, \boldsymbol{\gamma} \sim \mathrm{Dir}\left(\boldsymbol{\gamma}_{n\to n'}\right),$$
$$\gamma_{n\to n'}^{(b)} = \gamma_b + \sum_{t=1}^{T} \omega_{t,n'}^{(n,b)} .$$

Gibbs sampling the weighted adjacency matrix. As before, the weights are conjugate with a gamma prior, $\mathrm{Gamma}(\kappa, \nu_{n\to n'})$, where the scale is presumed to be given by the network prior. Given the adjacency matrix $\boldsymbol{A}$ and the auxiliary parent variables, the conditional distribution is,

$$w_{n\to n'} \mid a_{n\to n'} = 1 \sim \mathrm{Gamma}(\widetilde{\kappa}^{(n,n')}, \widetilde{\nu}^{(n,n')}),$$
$$\widetilde{\kappa}^{(n,n')} = \kappa + \sum_{t=1}^{T} \sum_{b=1}^{B} \omega_{t,n'}^{(n,b)},$$
$$\widetilde{\nu}^{(n,n')} = \nu_{n\to n'} + \sum_{t=1}^{T} s_{t,n}.$$

As in the previous chapter, in order to resample $\boldsymbol{A}$, we iterate over each entry and sample from the conditional distribution after integrating out the parents. We assume the parameters of the network prior can be sampled efficiently — a reasonable assumption for many exchangeable random network models.

The continuous time representation introduces a latent "parent" variable for each event in the dataset, and the parent can be any one of the events that occurred in the preceding window of influence. Call the number of potential parents M. The discrete time representation has a multinomial random variable for each time bin that contains at least one event, and the support of this multinomial is always a fixed size, $NB + 1$. When the rate of events is high, $NB + 1 \ll M$, allowing for dramatic improvements in efficiency in the discrete case.

Figure 4.1 shows the time per full Gibbs sweep as a function of the number of events per discrete time bin for the discrete and continuous formulations. The discrete formu-

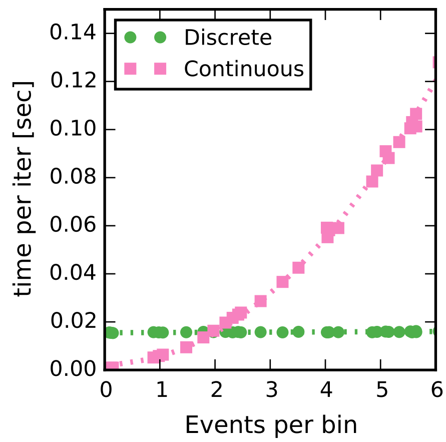

Figure 4.1: Comparison of run time per Gibbs sweep for the discrete and continuous network Hawkes formulations. Best fit lines added.

lation incurs a constant penalty whereas the continuous formulation quickly grows with the event rate. For low rates, the continuous formulation can be advantageous, but the discrete model is vastly superior in many realistic settings. For example, in Chapter 3 we worked with trades on the S&P100, which occur tens or hundreds of times per second for each stock. Since the complexity of our continuous time algorithm grew with the number of events, we had to downsample the data to consider only the times when stock prices changed significantly. However, we were also looking for interactions on time scales of one minute, very large compared to the rate of trades. Thus, it is reasonable to consider a discrete time model in which the number of trades is counted in, say, 1sec bins instead. The discrete time methods of this chapter would allow us to work directly with this type of trade-level activity and still scale to days or weeks of data.

4.3 Stochastic Variational Inference

The discrete time formulation offers advantageous complexity compared to the continuous analogue, but in order to maintain the invariance of the posterior distribution, we must still work with the entire set of parents each iteration. In many cases, a subset, or "mini-batch," of time bins can provide substantial information about the global parameters of the model, and rapid progress can be made by iterating quickly over subsets of the data. This motivates our derivation of a stochastic variational inference algorithm (Hoffman et al., 2013) for this discrete time model.

Variational methods optimize a lower bound on the marginal likelihood by minimizing the KL-divergence between a tractable approximating distribution and the true posterior. Since the local parents variables, $\boldsymbol{\omega}$, are conditionally independent given the global parameters ($\boldsymbol{A}$, $\boldsymbol{W}$, $\boldsymbol{\theta}$, etc.), our variational approach will easily extend to the stochastic setting in which we compute unbiased estimates of the gradient of the variational objective using mini-batches of data.

The primary impediment to deriving a variational approximation is the non-conjugacy of the spike-and-slab prior on the weights. To overcome this, we approximate the spike-and-slab prior with a mixture of gamma distributions, as has previously explored by Grabska-Barwinska et al. (2013):

$$
\begin{aligned}
p(\boldsymbol{A}, \boldsymbol{W} \mid \{\boldsymbol{z}_n\}, \boldsymbol{\vartheta}) &= \prod_{n,n'} p(a_{n\to n'} \mid \boldsymbol{z}_n, \boldsymbol{z}_{n'}, \boldsymbol{\vartheta})\, p(w_{n\to n'} \mid a_{n\to n'}, \boldsymbol{z}_n, \boldsymbol{z}_{n'}, \boldsymbol{\vartheta}) \\
p(a_{n\to n'} \mid \boldsymbol{z}_n, \boldsymbol{z}_{n'}, \boldsymbol{\vartheta}) &= \mathrm{Bern}(a_{n\to n'} \mid \rho_{n\to n'}), \\
p(w_{n\to n'} \mid a_{n\to n'}, \boldsymbol{z}_n, \boldsymbol{z}_{n'}, \boldsymbol{\vartheta}) &= \begin{cases} \mathrm{Gamma}(w_{n\to n'} \mid \kappa, \nu_{n\to n'}, a_{n\to n'} = 1), \\ \mathrm{Gamma}(w_{n\to n'} \mid \kappa_0, \nu_0, a_{n\to n'} = 0), \end{cases}
\end{aligned}
$$

where, as before, $\rho_{n\to n'}$ and $\nu_{n\to n'}$ are functions of the latent variables, $\boldsymbol{z}_n$ and $\boldsymbol{z}_{n'}$, and the parameters $\boldsymbol{\vartheta}$. We have approximated the "spike" in the spike-and-slab model with a gamma distribution parameterized by κ_0 and ν_0. As $\kappa_0 \to 0$ and $\nu_0 \to \infty$, the gamma distribution approaches a spike at zero.

This approximate probabilistic model is now amenable to mean field variational inference. We use a fully-factorized variational approximation, with the exception of a joint factor for each connection, $(a_{n \to n'}, w_{n \to n'})$.

$$
\begin{aligned}
q(a_{n \to n'}) &= \mathrm{Bern}(a_{n \to n'} \,|\, \widetilde{p}_{n \to n'}), \\
q(w_{n \to n'} \,|\, a_{n \to n'}) &= \begin{cases} \mathrm{Gamma}(w_{n \to n'} \,|\, \widetilde{\kappa}_1^{(n,n')}, \widetilde{\nu}_1^{(n,n')}, a_{n \to n'} = 1), \\ \mathrm{Gamma}(w_{n \to n'} \,|\, \widetilde{\kappa}_0^{(n,n')}, \widetilde{\nu}_0^{(n,n')}, a_{n \to n'} = 0) \,. \end{cases}
\end{aligned}
$$

Since the model is fully conjugate, the factors are easily derived.

Variational updates for parent variables, $q(\boldsymbol{\omega}_{t,n'})$ For the parent variables, the variational updates are

$$
\begin{aligned}
q(\boldsymbol{\omega}_{t,n'}) &= \mathrm{Mult}(\boldsymbol{\omega}_{t,n'} \,|\, s_{t,n'}, \widetilde{\boldsymbol{u}}_{t,n'}), \\
\widetilde{u}_{t,n'}^{(0)} &= \frac{1}{Z} \exp\left\{ \mathbb{E}_{\boldsymbol{\lambda}}[\ln \lambda_{n'}^{(0)}] \right\}, \\
\widetilde{u}_{t,n'}^{(n,b)} &= \frac{1}{Z} \hat{s}_{t,n,b} \exp\left\{ \mathbb{E}_{\boldsymbol{\theta}}[\ln \theta_{n \to n'}^{(b)}] + \mathbb{E}_W[\ln w_{n \to n'}] \right\},
\end{aligned}
$$

where Z is the normalization constant.

Variational updates for background rates, $q(\lambda_n^{(0)})$ The variational form parameters of the gamma distribution over background rates are

$$
\begin{aligned}
q(\lambda_n^{(0)}) &= \mathrm{Gamma}(\lambda_n^{(0)} \,|\, \widetilde{\alpha}_\lambda^{(n)}, \widetilde{\beta}_\lambda^{(n)}), \\
\widetilde{\alpha}_\lambda^{(n)} &= \alpha_\lambda + \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\omega}} \left[\omega_{t,n}^{(0)} \right], \\
\widetilde{\beta}_\lambda^{(n)} &= \beta_\lambda + T \Delta t \,.
\end{aligned}
$$

Variational approximation for impulse response parameters, $q(\boldsymbol{\theta}_{n \to n'})$ With the conjugate prior formulation the variational parameter updates for the Dirichlet distributed impulse response parameters are

$$
\begin{aligned}
q(\boldsymbol{\theta}_{n\to n'}) &= \mathrm{Dir}(\boldsymbol{n}_{n\to n'} \,|\, \widetilde{\boldsymbol{\gamma}}^{(n,n')}), \\
\widetilde{\gamma}_b^{(n,n')} &= \gamma_b + \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\omega}}\left[\omega_{t,n'}^{(n,b)}\right] .
\end{aligned}
$$

Variational approximation for the weighted adjacency matrix. The primary motivation for adopting a weakly sparse mixture of gamma distributions is to derive an efficient variational inference algorithm. The mixture-of-gammas prior is conjugate with the Poisson observations, and hence the variational distribution is also a mixture of gammas:

$$
\begin{aligned}
q(w_{n\to n'} \,|\, a_{n\to n'} = 1) &= \mathrm{Gamma}(w_{n\to n'} \,|\, \widetilde{\kappa}_1^{(n,n')}, \widetilde{\nu}_1^{(n,n')}) \\
\widetilde{\kappa}_1^{(k,k')} &= \kappa + \sum_{t=1}^{T}\sum_{b=1}^{B} \mathbb{E}_{\boldsymbol{\omega}}\left[\omega_{t,n'}^{(n,b)}\right] \\
\widetilde{\nu}_1^{(n,n')} &= \mathbb{E}_{\nu}[\nu_{n\to n'}] + \sum_{t=1}^{T} s_{t,n} ,
\end{aligned}
$$

and likewise for the "spike" factor,

$$
\begin{aligned}
q(w_{n\to n'} \,|\, a_{n\to n'} = 0) &= \mathrm{Gamma}(w_{n\to n'} \,|\, \widetilde{\kappa}_0^{(n,n')}, \widetilde{\nu}_0^{(n,n')}) \\
\widetilde{\kappa}_0^{(k,k')} &= \kappa_0 + \sum_{t=1}^{T}\sum_{b=1}^{B} \mathbb{E}_{\boldsymbol{\omega}}\left[\omega_{t,n'}^{(n,b)}\right] \\
\widetilde{\nu}_0^{(n,n')} &= \nu_0 + \sum_{t=1}^{T} s_{t,n} .
\end{aligned}
$$

This leaves us with $q(a_{n\to n'})$, which is Bernoulli distributed with parameter $\widetilde{p}_{n\to n'}$. The optimal parameter is given by,

$$\frac{\widetilde{p}_{n\to n'}}{1-\widetilde{p}_{n\to n'}} = \frac{\exp\{\mathbb{E}[\ln \rho_{n\to n'}]\}}{\exp\{\mathbb{E}[\ln(1-\rho_{n\to n'})]\}} \times \frac{(\exp\{\mathbb{E}[\ln \nu_{n\to n'}]\})^{\kappa}}{\Gamma(\kappa)} \times \frac{\Gamma(\widetilde{\kappa}_1^{(n,n')})}{(\widetilde{\nu}_1^{(n,n')})^{\widetilde{\kappa}_1^{(n,n')}}} \times \frac{\Gamma(\kappa_0)}{(\nu_0)^{\kappa_0}} \times \frac{(\widetilde{\nu}_0^{(n,n')})^{\widetilde{\kappa}_0^{(n,n')}}}{\Gamma(\widetilde{\kappa}_0^{(n,n')})}.$$

As with Gibbs sampling, we assume a variational approximation for the network model can be derived, and provide access to the necessary expectations, $\mathbb{E}[\ln \rho_{n\to n'}]$, $\mathbb{E}[\ln(1-\rho_{n\to n'})]$, $\mathbb{E}[\nu_{n\to n'}]$ and $\mathbb{E}[\ln \nu_{n\to n'}]$.

The spike counts in each time bin are conditionally independent given the network weights and the adjacency matrix — a common pattern exploited by stochastic variational inference (SVI) algorithms (Hoffman et al., 2013). These methods optimize the variational objective using stochastic gradient methods that work with mini-batches of data. Often, a mini-batch of data can provide valuable information about the global parameters, in our case the network and background rates. Quickly iterating over these global parameters allows us to reach good modes of the posterior distribution in a fraction of the time that standard variational Bayes and Gibbs sampling require, since those methods must process the entire dataset before making an update. SVI does require some tuning, however. In particular, we must set a mini-batch size and a step size schedule. In this work, we fix the mini-batch size to $T_{\mathsf{mb}} = 1024$ and set the step size at iteration i to $(i+1)^{-0.5}$. These parameters may be tuned with general purpose hyperparameter optimization techniques (Snoek et al., 2012).

4.4 Synthetic Results

We assess the performance of the proposed inference algorithms on a synthetic dataset generated by a strongly sparse Hawkes process with $N = 50$ neurons. We used a stochastic block model network prior with $K = 5$ clusters, each consisting of ten densely connected processes ($p_{k\to k} = 0.4$), with sparse connections to processes in other clusters ($p_{k\to k'} = 0.01$). All weights share the same scale of $\nu = 5.0$, though this information is not provided *a priori*. We simulate $T = 10^5$ time bins in steps of size $\Delta t = 1$. The neurons have an mean

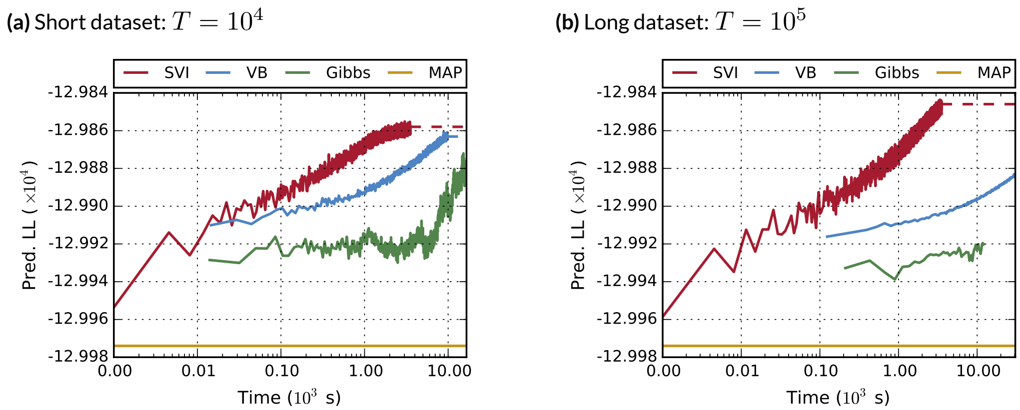

Figure 4.2: Predictive log likelihood versus wall clock time for three Bayesian inference algorithms on a dataset of $N = 50$ neurons and $T = 10^4$ and $T = 10^5$ time bins on the left and right, respectively.

background rate of 1.0 event per time bin and, due to the network interactions, the average total rate of the processes is 16.7 ± 12.0 events per bin. Referring to Figure 4.1, this is a regime that favors the discrete model. We initialized by performing MAP estimation on the first $T_{\text{init}} = 10^4$. Then we trained the model using Gibbs sampling, batch variational Bayesian inference, and stochastic variational inference,

We trained the models on only the first 10^4 time bins, the same that were used for initialization. We evaluated the algorithms in terms of their predictive log likelihood on a held-out dataset of length $T_{\text{test}} = 10^3$. Figure 4.2a shows the results as a function of wall-clock time. We find that SVI obtains competitive predictive log likelihood in a matter of minutes. Batch VB and Gibbs converge at a considerably slower rate, though they eventually match the SVI predictive likelihood after hours of computation. The MAP estimate, even with cross validated regularization, underperforms the other competing algorithms.

This trend is exaggerated when we consider the entire training set of size $T = 10^5$. Figure 4.2b illustrates the power of SVI in handling these large time datasets. Considerable information about the global parameters (e.g., the network) can be gained from just a minibatch of time points. Hence, we can make rapid improvements in predictive log likelihood very quickly. By contrast, each step of the Gibbs and batch VB algorithms is approximately 10 times slower, and even after computing sufficient statistics over the entire dataset, the

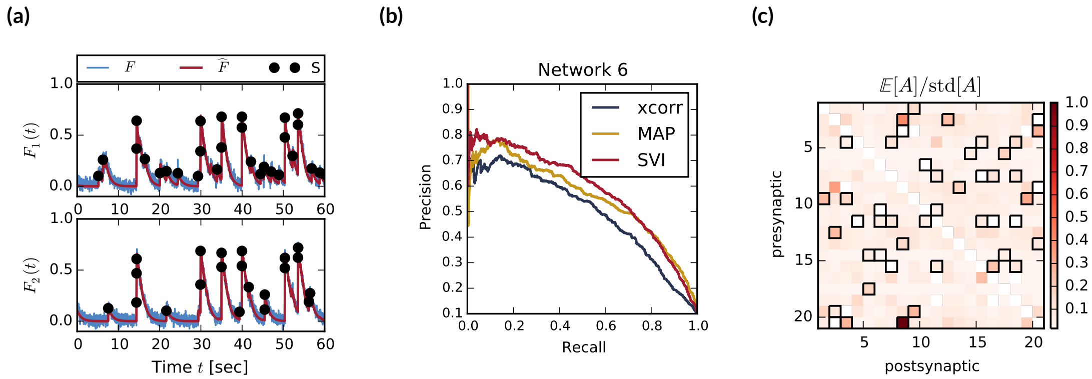

Figure 4.3: Application of the network Hawkes model to a connectomics challenge. **(a)** The data is in the form of a calcium fluorescence trace, which we preprocess to extract neural spike times. **(b)** We measure performance on a link prediction task using a precision-recall curve and find that the posterior estimates of SVI provide the best estimates on some networks. In addition to an estimate of the connection probability and weight, SVI provides an estimate of the posterior uncertainty. **(c)** Inferred $\mathbb{E}_q[\boldsymbol{A}]/\mathrm{std}_q[\boldsymbol{A}]$ for the first 20 neurons. True connections are outlined in black.

algorithm is only able to make limited progress per iteration.

4.5 Connectomics Results

We tested these inference algorithms on the data from the Chalearn neural connectomics challenge* (Stetter et al., 2012). The data consist of calcium fluorescence traces, $\boldsymbol{F}$, from six networks of $N = 100$ neurons each. We use ten minutes of data at 50Hz sampling frequency to yield $T = 3 \times 10^6$ entries in $\boldsymbol{S}$. In this case, the networks are purely excitatory, and each action potential, or spike, increases the probability of the downstream neurons firing as a result. This matches the underlying intuition of the Hawkes process model, making it a natural choice.

In order to apply the Hawkes model, we first convert the fluorescence traces into a spike count matrix using OOPSI, a Bayesian inference algorithm based on a model of calcium fluorescence (Vogelstein et al., 2010). The output is a filtered fluorescence trace, $\widehat{\boldsymbol{F}}$, and a probability of spike for each time bin. We threshold this at probability 0.7 to get a $T \times N$

*http://connectomics.chalearn.org

	Network 1		Network 2		Network 3		Network 4		Network 5	
Algorithm	ROC	PRC	ROC	PRC	ROC	PRC	ROC	PRC	ROC	PRC
xcorr	0.596	0.139	0.591	0.133	**0.701**	**0.198**	0.745	0.296	0.798	0.359
MAP	0.607	0.174	**0.619**	**0.143**	0.698	0.178	**0.790**	0.334	**0.859**	0.408
SVI	**0.649**	**0.184**	0.605	0.141	0.673	0.176	0.774	**0.342**	0.844	**0.410**

Table 4.1: Comparison of inference algorithms on link prediction for five networks from the Chalearn connectomics challenge. Performance is measured by area under the ROC curve and area under the precision recall curve (PRC). In four of the five networks a Hawkes process model provides the best results.

binary spike matrix, $\boldsymbol{S}$. This preprocessing is shown in Figure 4.3a.

Figure 4.3b shows the precision-recall curve we used to evaluate the algorithms' performance on network recovery. As a baseline, we compare against simple thresholding of the cross correlation matrix. On Network 6, SVI offers the best network inference. Table 4.1 shows the results on the other five networks using the same model parameters. On 4/5 of these networks, the Bayesian methods offer the best performance.

Figure 4.3c illustrates one of the main advantages of the fully Bayesian inference algorithm – calibrated estimates of posterior uncertainty. Here we show the SVI algorithm's estimate of the posterior mean of $\boldsymbol{A}$ normalized by the posterior standard deviation for a subset of 20 neurons from Network 6. We also outline the true connections to show that the most confident predictions are more likely to correspond to true connections. Such estimates of the posterior uncertainty are not available with standard heuristic methods or point estimates.

4.6 Conclusion

This brief chapter provided a link between the ideas introduced in Chapter 3 — namely the combination of network models and point process observations — to the discrete time autoregressive models of the next few chapters. We also showed how the conditional independence of the spike counts could be leveraged in a stochastic variational inference algorithm that scales to long recording durations. The key, again, was the Poisson superposition principle, which allowed a simple auxiliary variable formulation. Combining this formulation with an approximate spike-and-slab model led to a fully-conjugate model that admitted an efficient inference algorithm.

In the next chapter, we will continue to build on these ideas, but we will address a major limitation of this approach. The Poisson superposition principle only applies to *linear* models. Since the rate must be nonnegative, linear models cannot have inhibitory interactions with negative weights. We will show how this limitation can be overcome with another clever auxiliary variable trick.

References

Yashar Ahmadian, Jonathan W Pillow, and Liam Paninski. Efficient Markov chain Monte Carlo methods for decoding neural spike trains. *Neural Computation*, 23(1):46–96, 2011.

Misha B Ahrens, Michael B Orger, Drew N Robson, Jennifer M Li, and Philipp J Keller. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature Methods*, 10(5):413–420, 2013.

Laurence Aitchison and Peter E Latham. Synaptic sampling: A connection between PSP variability and uncertainty explains neurophysiological observations. *arXiv preprint arXiv:1505.04544*, 2015.

Laurence Aitchison and Máté Lengyel. The Hamiltonian brain. *arXiv preprint arXiv:1407.0973*, 2014.

David J Aldous. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598, 1981.

Charles H Anderson and David C Van Essen. Neurobiological computational systems. *Computational Intelligence Imitating Life*, pages 1–11, 1994.

Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1-2):5–43, 2003.

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

Michael J Barber, John W Clark, and Charles H Anderson. Neural representation of probabilistic information. *Neural Computation*, 15(8):1843–64, August 2003.

Leonard E Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563, 1966.

Matthew J. Beal, Zoubin Ghahramani, and Carl E. Rasmussen. The infinite hidden Markov model. *Advances in Neural Information Processing Systems 14*, pages 577–585, 2002.

Jeffrey M Beck and Alexandre Pouget. Exact inferences in a neural implementation of a hidden Markov model. *Neural Computation*, 19(5):1344–1361, 2007.

Jeffrey M Beck, Peter E Latham, and Alexandre Pouget. Marginalization in neural circuits with divisive normalization. *The Journal of Neuroscience*, 31(43):15310–15319, 2011.

Jeffrey M Beck, Katherine A Heller, and Alexandre Pouget. Complex inference in neural circuits with probabilistic population codes and topic models. *Advances in Neural Information Processing Systems*, pages 3059–3067, 2012.

Yoshua Bengio and Paolo Frasconi. An input output HMM architecture. *Advances in Neural Information Processing Systems*, pages 427–434, 1995.

Pietro Berkes, Gergo Orbán, Máté Lengyel, and József Fiser. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331(6013): 83–7, January 2011.

Gordon J Berman, Daniel M Choi, William Bialek, and Joshua W Shaevitz. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99):20140672, 2014.

Philippe Biane, Jim Pitman, and Marc Yor. Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bulletin of the American Mathematical Society*, 38(4):435–465, 2001.

Christopher M Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

David M Blei. Build, compute, critique, repeat: Data analysis with latent variable models. *Annual Review of Statistics and Its Application*, 1:203–232, 2014.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

Carolyn R Block and Richard Block. *Street gang crime in Chicago*. US Department of Justice, Office of Justice Programs, National Institute of Justice, 1993.

Carolyn R Block, Richard Block, and Illinois Criminal Justice Information Authority. Homicides in Chicago, 1965-1995. ICPSR06399-v5. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], July 2005.

Charles Blundell, Katherine A Heller, and Jeffrey M Beck. Modelling reciprocating relationships with Hawkes processes. *Advances in Neural Information Processing Systems*, pages 2600–2608, 2012.

George EP Box. Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A (General)*, pages 383–430, 1980.

David H Brainard and William T Freeman. Bayesian color constancy. *Journal of the Optical Society of America A*, 14(7):1393–1411, 1997.

Kevin L Briggman, Henry DI Abarbanel, and William B Kristan. Optical imaging of neuronal populations during decision-making. *Science*, 307(5711):896–901, 2005.

David R. Brillinger. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics*, 59(3):189–200, August 1988.

David R Brillinger, Hugh L Bryant Jr, and Jose P Segundo. Identification of synaptic interactions. *Biological Cybernetics*, 22(4):213–228, 1976.

Michael Bryant and Erik B Sudderth. Truly nonparametric online variational inference for hierarchical Dirichlet processes. *Advances in Neural Information Processing Systems 25*, pages 2699–2707, 2012.

Lars Buesing, Johannes Bill, Bernhard Nessler, and Wolfgang Maass. Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. *PLoS Computational Biology*, 7(11):e1002211, November 2011.

Lars Buesing, Jakob H. Macke, and Maneesh Sahani. Learning stable, regularised latent models of neural population dynamics. *Network: Computation in Neural Systems*, 23: 24–47, 2012a.

Lars Buesing, Jakob H Macke, and Maneesh Sahani. Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. *Advances in Neural Information Processing Systems*, pages 1682–1690, 2012b.

Lars Buesing, Timothy A Machado, John P Cunningham, and Liam Paninski. Clustered factor analysis of multineuronal spike data. *Advances in Neural Information Processing Systems*, pages 3500–3508, 2014.

Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

Santiago Ramón Cajal. *Textura del Sistema Nervioso del Hombre y los Vertebrados*, volume 1. Imprenta y Librería de Nicolás Moya, Madrid, Spain, 1899.

Natalia Caporale and Yang Dan. Spike timing-dependent plasticity: a Hebbian learning rule. *Annual Review of Neuroscience*, 31:25–46, 2008.

Nick Chater and Christopher D Manning. Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, 10(7):335–344, 2006.

Zhe Chen, Fabian Kloosterman, Emery N Brown, and Matthew A Wilson. Uncovering spatial topology represented by rat hippocampal population neuronal codes. *Journal of Computational Neuroscience*, 33(2):227–255, 2012.

Zhe Chen, Stephen N Gomperts, Jun Yamamoto, and Matthew A Wilson. Neural representation of spatial topology in the rodent hippocampus. *Neural Computation*, 26(1): 1–39, 2014.

Sharat Chikkerur, Thomas Serre, Cheston Tan, and Tomaso Poggio. What and where: A Bayesian inference theory of attention. *Vision Research*, 50(22):2233–2247, 2010.

Yoon Sik Cho, Aram Galstyan, Jeff Brantingham, and George Tita. Latent point process models for spatial-temporal networks. *arXiv:1302.2671*, 2013.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945, 2004.

Aaron C Courville, Nathaniel D Daw, and David S Touretzky. Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, 10(7):294–300, 2006.

Ronald L Cowan and Charles J Wilson. Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex. *Journal of Neurophysiology*, 71(1):17–32, 1994.

W Maxwell Cowan, Thomas C Südhof, and Charles F Stevens. *Synapses*. Johns Hopkins University Press, 2003.

Mary Kathryn Cowles and Bradley P Carlin. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91: 883–904, 1996.

John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509, 2014.

Paul Dagum and Michael Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60(1):141–153, 1993.

Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: Volume I: Elementary Theory and Methods*. Springer Science & Business Media, 2 edition, 2003.

Peter Dayan and Larry F Abbott. *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press, 2001.

Peter Dayan and Joshua A Solomon. Selective Bayes: Attentional load and crowding. *Vision Research*, 50(22):2248–2260, 2010.

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.

Sophie Deneve. Bayesian spiking neurons I: inference. *Neural Computation*, 20(1):91–117, January 2008.

Luc Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, USA, 1986.

Christopher DuBois, Carter Butts, and Padhraic Smyth. Stochastic block modeling of relational event dynamics. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 238–246, 2013.

Seif Eldawlatly, Yang Zhou, Rong Jin, and Karim G Oweiss. On the use of dynamic Bayesian networks in reconstructing functional neuronal networks from spike train ensembles. *Neural Computation*, 22(1):158–189, 2010.

Marc O Ernst and Martin S Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002.

Sean Escola, Alfredo Fontanini, Don Katz, and Liam Paninski. Hidden Markov models for the stimulus-response relationships of multistate neural systems. *Neural Computation*, 23(5):1071–1132, 2011.

Warren John Ewens. Population genetics theory—the past and the future. In S. Lessard, editor, *Mathematical and Statistical Developments of Evolutionary Theory*, pages 177–227. Springer, 1990.

Daniel E Feldman. The spike-timing dependence of plasticity. *Neuron*, 75(4):556–71, August 2012.

Daniel J Felleman and David C Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1):1–47, 1991.

Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230, 1973.

Christopher R Fetsch, Amanda H Turner, Gregory C DeAngelis, and Dora E Angelaki. Dynamic reweighting of visual and vestibular cues during self-motion perception. *The Journal of Neuroscience*, 29(49):15601–15612, 2009.

Christopher R Fetsch, Alexandre Pouget, Gregory C DeAngelis, and Dora E Angelaki. Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, 15(1):146–154, 2012.

József Fiser, Pietro Berkes, Gergő Orbán, and Máté Lengyel. Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, 14(3):119–130, 2010.

Alyson K Fletcher, Sundeep Rangan, Lav R Varshney, and Aniruddha Bhargava. Neural reconstruction with approximate message passing (neuramp). *Advances in Neural Information Processing Systems*, pages 2555–2563, 2011.

Emily B Fox. *Bayesian nonparametric learning of complex dynamical phenomena*. PhD thesis, Massachusetts Institute of Technology, 2009.

Emily B Fox, Erik B Sudderth, Michael I Jordan, and Alan S Willsky. An HDP-HMM for systems with state persistence. *Proceedings of the International Conference on Machine Learning*, pages 312–319, 2008.

Jeremy Freeman, Greg D Field, Peter H Li, Martin Greschner, Deborah E Gunning, Keith Mathieson, Alexander Sher, Alan M Litke, Liam Paninski, Eero P Simoncelli, et al. Mapping nonlinear receptive field structure in primate retina at single cone resolution. *eLife*, 4:e05241, 2015.

Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, 11(2):127–38, February 2010.

Karl J Friston. Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.

Deep Ganguli and Eero P Simoncelli. Implicit encoding of prior probabilities in optimal neural populations. *Advances in Neural Information Processing Systems*, pages 6–9, 2010.

Peiran Gao and Surya Ganguli. On simplicity and complexity in the brave new world of large-scale neuroscience. *Current Opinion in Neurobiology*, 32:148–155, 2015.

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis*. CRC press, 3rd edition, 2013.

Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741, 1984.

Felipe Gerhard, Tilman Kispersky, Gabrielle J Gutierrez, Eve Marder, Mark Kramer, and Uri Eden. Successful reconstruction of a physiological circuit with known connectivity from spiking activity alone. *PLoS Computational Biology*, 9(7):e1003138, 2013.

Samuel J Gershman, Matthew D Hoffman, and David M Blei. Nonparametric variational inference. *Proceedings of the International Conference on Machine Learning*, pages 663–670, 2012a.

Samuel J Gershman, Edward Vul, and Joshua B Tenenbaum. Multistability and perceptual inference. *Neural Computation*, 24(1):1–24, 2012b.

Sebastian Gerwinn, Jakob Macke, Matthias Seeger, and Matthias Bethge. Bayesian inference for spiking neuron models with a sparsity prior. *Advances in Neural Information Processing Systems*, pages 529–536, 2008.

Charles J Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, pages 473–483, 1992.

Walter R Gilks. *Markov Chain Monte Carlo*. Wiley Online Library, 2005.

Anna Goldenberg, Alice X Zheng, Stephen E Fienberg, and Edoardo M Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233, 2010.

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.

Noah Goodman, Vikash Mansinghka, Daniel M Roy, Keith Bonawitz, and Joshua B Tenenbaum. Church: a language for generative models. *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 220––229, 2008.

Noah D Goodman, Joshua B Tenenbaum, and Tobias Gerstenberg. Concepts in a probabilistic language of thought. Technical report, Center for Brains, Minds and Machines (CBMM), 2014.

Agnieszka Grabska-Barwinska, Jeff Beck, Alexandre Pouget, and Peter Latham. Demixing odors-fast inference in olfaction. *Advances in Neural Information Processing Systems*, pages 1968–1976, 2013.

SG Gregory, KF Barlow, KE McLay, R Kaul, D Swarbreck, A Dunham, CE Scott, KL Howe, K Woodfine, CCA Spencer, et al. The DNA sequence and biological annotation of human chromosome 1. *Nature*, 441(7091):315–321, 2006.

Thomas L Griffiths, Charles Kemp, and Joshua B Tenenbaum. Bayesian models of cognition. In Ron Sun, editor, *The Cambridge Handbook of Computational Psychology*. Cambridge University Press, 2008.

Roger B Grosse, Chris J Maddison, and Ruslan R Salakhutdinov. Annealing between distributions by averaging moments. *Advances in Neural Information Processing Systems*, pages 2769–2777, 2013.

Roger B Grosse, Zoubin Ghahramani, and Ryan P Adams. Sandwiching the marginal likelihood using bidirectional Monte Carlo. *arXiv preprint arXiv:1511.02543*, 2015.

Yong Gu, Dora E Angelaki, and Gregory C DeAngelis. Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, 11(10):1201–1210, 2008.

Fangjian Guo, Charles Blundell, Hanna Wallach, and Katherine A Heller. The Bayesian echo chamber: Modeling influence in conversations. *arXiv preprint arXiv:1411.2674*, 2014.

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

Moritz Helmstaedter, Kevin L Briggman, Srinivas C Turaga, Viren Jain, H Sebastian Seung, and Winfried Denk. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461):168–174, 2013.

Geoffrey E Hinton. How neural networks learn from experience. *Scientific American*, 1992.

Geoffrey E Hinton and Terrence J Sejnowski. Optimal perceptual inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1983.

Daniel R Hochbaum, Yongxin Zhao, Samouil L Farhi, Nathan Klapoetke, Christopher A Werley, Vikrant Kapoor, Peng Zou, Joel M Kralj, Dougal Maclaurin, Niklas Smedemark-Margulies, et al. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nature Methods*, 2014.

Peter D Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in Neural Information Processing Systems*, 20:1–8, 2008.

Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

Douglas N. Hoover. Relations on probability spaces and arrays of random variables. Technical report, Institute for Advanced Study, Princeton, 1979.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

Patrik O Hoyer and Aapo Hyvarinen. Interpreting neural response variability as Monte Carlo sampling of the posterior. *Advances in neural information processing systems*, pages 293–300, 2003.

Yanping Huang and Rajesh P. N. Rao. Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5):580–593, September 2011.

David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154, 1962.

Hemant Ishwaran and Mahmoud Zarepour. Exact and approximate sum representations for the Dirichlet process. *Canadian Journal of Statistics*, 30(2):269–283, 2002.

Tomoharu Iwata, Amar Shah, and Zoubin Ghahramani. Discovering latent influence in online social activities via shared cascade Poisson processes. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 266–274, 2013.

Mehrdad Jazayeri and Michael N Shadlen. Temporal context calibrates interval timing. *Nature Neuroscience*, 13(8):1020–1026, 2010.

Mehrdad Jazayeri and Michael N Shadlen. A neural mechanism for sensing and reproducing a time interval. *Current Biology*, 25(20):2599–2609, 2015.

Matthew J Johnson. *Bayesian time series models and scalable inference*. PhD thesis, Massachusetts Institute of Technology, June 2014.

Matthew J Johnson and Alan S Willsky. Bayesian nonparametric hidden semi-Markov models. *Journal of Machine Learning Research*, 14(1):673–701, 2013.

Matthew J Johnson and Alan S Willsky. Stochastic variational inference for Bayesian time series models. *Proceedings of the International Conference on Machine Learning*, 32:1854–1862, 2014.

Matthew J Johnson, Scott W Linderman, Sandeep R Datta, and Ryan P Adams. Discovering switching autoregressive dynamics in neural spike train recordings. *Computational and Systems Neuroscience (Cosyne) Abstracts*, 2015.

Lauren M Jones, Alfredo Fontanini, Brian F Sadacca, Paul Miller, and Donald B Katz. Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proceedings of the National Academy of Sciences*, 104(47):18772–18777, 2007.

Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, 1999.

Eric R Kandel, James H Schwartz, Thomas M Jessell, et al. *Principles of neural science*, volume 4. McGraw-Hill New York, 2000.

David Kappel, Stefan Habenschuss, Robert Legenstein, and Wolfgang Maass. Network plasticity as Bayesian inference. *PLoS Computational Biology*, 11(11):e1004485, 2015a.

David Kappel, Stefan Habenschuss, Robert Legenstein, and Wolfgang Maass. Synaptic sampling: A Bayesian approach to neural network plasticity and rewiring. *Advances in Neural Information Processing Systems*, pages 370–378, 2015b.

Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

Jason ND Kerr and Winfried Denk. Imaging in vivo: watching the brain in action. *Nature Reviews Neuroscience*, 9(3):195–205, 2008.

Roozbeh Kiani and Michael N Shadlen. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, 324(5928):759–64, May 2009.

John F. C. Kingman. *Poisson Processes (Oxford Studies in Probability)*. Oxford University Press, January 1993. ISBN 0198536933.

David C Knill and Whitman Richards. *Perception as Bayesian inference*. Cambridge University Press, 1996.

Konrad P Körding and Daniel M Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–7, January 2004.

Alp Kucukelbir, Rajesh Ranganath, Andrew Gelman, and David Blei. Automatic variational inference in Stan. *Advances in Neural Information Processing Systems*, pages 568–576, 2015.

Stephen W Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16(1):37–68, 1953.

Harold W Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.

Kenneth W Latimer, Jacob L Yates, Miriam LR Meister, Alexander C Huk, and Jonathan W Pillow. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*, 349(6244):184–187, 2015.

Tai Sing Lee and David Mumford. Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 20(7):1434–1448, 2003.

Robert Legenstein and Wolfgang Maass. Ensembles of spiking neurons with noise support optimal probabilistic inference in a dynamically changing environment. *PLoS Computational Biology*, 10(10):e1003859, 2014.

William C Lemon, Stefan R Pulver, Burkhard Höckendorf, Katie McDole, Kristin Branson, Jeremy Freeman, and Philipp J Keller. Whole-central nervous system functional imaging in larval Drosophila. *Nature Communications*, 6, 2015.

Michael S Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):R53–R78, 1998.

Percy Liang, Slav Petrov, Michael I Jordan, and Dan Klein. The infinite PCFG using hierarchical Dirichlet processes. *Proceedings of Empirical Methods in Natural Language Processing*, pages 688–697, 2007.

David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031, 2007.

Jeff W Lichtman, Jean Livet, and Joshua R Sanes. A technicolour approach to the connectome. *Nature Reviews Neuroscience*, 9(6):417–422, 2008.

Scott W Linderman and Ryan P. Adams. Discovering latent network structure in point process data. *Proceedings of the International Conference on Machine Learning*, pages 1413–1421, 2014.

Scott W Linderman and Ryan P Adams. Scalable Bayesian inference for excitatory point process networks. *arXiv preprint arXiv:1507.03228*, 2015.

Scott W Linderman and Ryan P Johnson, Matthew Jand Adams. Dependent multinomial models made easy: Stick-breaking with the Pólya-gamma augmentation. *Advances in Neural Information Processing Systems*, pages 3438–3446, 2015.

Scott W Linderman, Christopher H Stock, and Ryan P Adams. A framework for studying synaptic plasticity with neural spike train data. *Advances in Neural Information Processing Systems*, pages 2330–2338, 2014.

Scott W Linderman, Ryan P Adams, and Jonathan W Pillow. Inferring structured connectivity from spike trains under negative-binomial generalized linear models. *Computational and Systems Neuroscience (Cosyne) Abstracts*, 2015.

Scott W Linderman, Matthew J Johnson, Matthew W Wilson, and Zhe Chen. A nonparametric Bayesian approach to uncovering rat hippocampal population codes during spatial navigation. *Journal of Neuroscience Methods*, 263:36–47, 2016a.

Scott W Linderman, Aaron Tucker, and Matthew J Johnson. Bayesian latent state space models of neural activity. *Computational and Systems Neuroscience (Cosyne) Abstracts*, 2016b.

Fredrik Lindsten, Michael I Jordan, and Thomas B Schön. Ancestor sampling for particle Gibbs. *Advances in Neural Information Processing Systems*, pages 2600–2608, 2012.

Shai Litvak and Shimon Ullman. Cortical circuitry implementing graphical models. *Neural Computation*, 21(11):3010–3056, 2009.

James Robert Lloyd, Peter Orbanz, Zoubin Ghahramani, and Daniel M Roy. Random function priors for exchangeable arrays with applications to graphs and relational data. *Advances in Neural Information Processing Systems*, 2012.

Wei Ji Ma and Mehrdad Jazayeri. Neural coding of uncertainty and probability. *Annual Review of Neuroscience*, 37:205–220, 2014.

Wei Ji Ma, Jeffrey M Beck, Peter E Latham, and Alexandre Pouget. Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11):1432–8, November 2006.

David JC MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992.

Jakob H Macke, Lars Buesing, John P Cunningham, M Yu Byron, Krishna V Shenoy, and Maneesh Sahani. Empirical models of spiking in neural populations. *Advances in neural information processing systems*, pages 1350–1358, 2011.

Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R Bialas, Nolan Kamitaki, Emily M Martersteck, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.

Vikash Mansinghka, Daniel Selsam, and Yura Perov. Venture: a higher-order probabilistic programming platform with programmable inference. *arXiv preprint arXiv:1404.0099*, 2014.

David Marr. *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press, 1982.

Paul Miller and Donald B Katz. Stochastic transitions between neural states in taste processing and decision-making. *The Journal of Neuroscience*, 30(7):2559–2570, 2010.

T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023—-1032, 1988.

Shakir Mohamed, Zoubin Ghahramani, and Katherine A Heller. Bayesian and L1 approaches for sparse unsupervised learning. *Proceedings of the International Conference on Machine Learning*, pages 751–758, 2012.

Jesper Møller, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482, 1998.

Michael L Morgan, Gregory C DeAngelis, and Dora E Angelaki. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, 59(4):662–673, 2008.

Abigail Morrison, Markus Diesmann, and Wulfram Gerstner. Phenomenological models of synaptic plasticity based on spike timing. *Biological Cybernetics*, 98(6):459–478, 2008.

Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

Radford M Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001.

Radford M. Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, pages 113–162, 2010.

John A Nelder and R Jacob Baker. Generalized linear models. *Encyclopedia of Statistical Sciences*, 1972.

Bernhard Nessler, Michael Pfeiffer, Lars Buesing, and Wolfgang Maass. Bayesian computation emerges in generic cortical microcircuits through spike-timing-dependent plasticity. *PLoS Computational Biology*, 9(4):e1003037, 2013.

Mark EJ Newman. The structure and function of complex networks. *Society for Industrial and Applied Mathematics (SIAM) Review*, 45(2):167–256, 2003.

Krzysztof Nowicki and Tom A B Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.

Seung Wook Oh, Julie A Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M Henry, et al. A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214, 2014.

Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3):267–273, 1982.

John O'Keefe and Lynn Nadel. *The Hippocampus as a Cognitive Map*, volume 3. Clarendon Press, 1978.

Peter Orbanz and Daniel M Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):437–461, 2015.

Peter Orbanz and Yee Whye Teh. Bayesian nonparametric models. In *Encyclopedia of Machine Learning*, pages 81–89. Springer, 2011.

Adam M Packer, Darcy S Peterka, Jan J Hirtz, Rohit Prakash, Karl Deisseroth, and Rafael Yuste. Two-photon optogenetics of dendritic spines and neural circuits. *Nature Methods*, 9(12):1202–1205, 2012.

Liam Paninski. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15(4):243–262, January 2004.

Liam Paninski, Yashar Ahmadian, Daniel Gil Ferreira, Shinsuke Koyama, Kamiar Rahnama Rad, Michael Vidne, Joshua Vogelstein, and Wei Wu. A new look at state-space models for neural data. *Journal of Computational Neuroscience*, 29(1-2):107–126, 2010.

Andrew V Papachristos. Murder by structure: Dominance relations and the social structure of gang homicide. *American Journal of Sociology*, 115(1):74–128, 2009.

Il Memming Park and Jonathan W Pillow. Bayesian spike-triggered covariance analysis. *Advances in Neural Information Processing Systems*, pages 1692–1700, 2011.

Patrick O Perry and Patrick J Wolfe. Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2013.

Biljana Petreska, Byron Yu, John P Cunningham, Gopal Santhanam, Stephen I Ryu, Krishna V Shenoy, and Maneesh Sahani. Dynamical segmentation of single trials from population neural data. *Advances in Neural Information Processing Systems*, pages 756–764, 2011.

David Pfau, Eftychios A Pnevmatikakis, and Liam Paninski. Robust learning of low-dimensional dynamics from large neural ensembles. *Advances in Neural Information Processing Systems*, pages 2391–2399, 2013.

Jonathan W. Pillow and James Scott. Fully Bayesian inference for neural models with negative-binomial spiking. *Advances in Neural Information Processing Systems*, pages 1898–1906, 2012.

Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

Eftychios A Pnevmatikakis, Daniel Soudry, Yuanjun Gao, Timothy A Machado, Josh Merel, David Pfau, Thomas Reardon, Yu Mu, Clay Lacefield, Weijian Yang, et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron*, 2016.

Nicholas G Polson, James G Scott, and Jesse Windle. Bayesian inference for logistic models using Pólya-gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349, 2013.

Ruben Portugues, Claudia E Feierstein, Florian Engert, and Michael B Orger. Whole-brain activity maps reveal stereotyped, distributed networks for visuomotor behavior. *Neuron*, 81(6):1328–1343, 2014.

Alexandre Pouget, Jeffrey M Beck, Wei Ji Ma, and Peter E Latham. Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, 16(9):1170–1178, 2013.

Robert Prevedel, Young-Gyu Yoon, Maximilian Hoffmann, Nikita Pak, Gordon Wetzstein, Saul Kato, Tina Schrödel, Ramesh Raskar, Manuel Zimmer, Edward S Boyden, et al. Simultaneous whole-animal 3d imaging of neuronal activity using light-field microscopy. *Nature Methods*, 11(7):727–730, 2014.

Lawrence R Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

Adrian E Raftery and Steven Lewis. How many iterations in the Gibbs sampler? *Bayesian Statistics*, pages 763–773, 1992.

Rajesh Ranganath, Sean Gerrish, and David M Blei. Black box variational inference. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 33:275––283, 2014.

Rajesh P. N. Rao. Bayesian computation in recurrent neural circuits. *Neural Computation*, 16(1):1–38, January 2004.

Rajesh P. N. Rao. Neural models of Bayesian belief propagation. In *Bayesian brain: Probabilistic approaches to neural computation*, pages 236–264. MIT Press Cambridge, MA, 2007.

Rajesh P. N. Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1): 79–87, January 1999.

Danilo J Rezende, Daan Wierstra, and Wulfram Gerstner. Variational learning for recurrent spiking networks. *Advances in Neural Information Processing Systems*, pages 136–144, 2011.

Fred Rieke, David Warland, Rob de Ruyter van Steveninck, and William Bialek. *Spikes: exploring the neural code*. MIT press, 1999.

Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

Dan Roth. On the hardness of approximate reasoning. *Artificial Intelligence*, 82(1):273–302, 1996.

Maneesh Sahani. *Latent variable models for neural data analysis*. PhD thesis, California Institute of Technology, 1999.

Maneesh Sahani and Peter Dayan. Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity. *Neural Computation*, 2279:2255–2279, 2003.

Joshua R Sanes and Richard H Masland. The types of retinal ganglion cells: current status and implications for neuronal classification. *Annual Review of Neuroscience*, 38:221–246, 2015.

Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4: 639–650, 1994.

Ben Shababo, Brooks Paige, Ari Pakman, and Liam Paninski. Bayesian inference and online experimental design for mapping neural microcircuits. *Advances in Neural Information Processing Systems*, pages 1304–1312, 2013.

Vahid Shalchyan and Dario Farina. A non-parametric Bayesian approach for clustering and tracking non-stationarities of neural spikes. *Journal of Neuroscience Methods*, 223: 85–91, 2014.

Lei Shi and Thomas L Griffiths. Neural implementation of hierarchical Bayesian inference by importance sampling. *Advances in Neural Information Processing Systems*, 2009.

Yousheng Shu, Andrea Hasenstaub, and David A McCormick. Turning on and off recurrent balanced cortical activity. *Nature*, 423(6937):288–293, 2003.

Jack W Silverstein. The spectral radii and norms of large dimensional non-central random matrices. *Stochastic Models*, 10(3):525–532, 1994.

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

Eero P Simoncelli. Optimal estimation in sensory systems. *The Cognitive Neurosciences, IV*, 2009.

Anne C Smith and Emery N Brown. Estimating a state-space model from point process observations. *Neural Computation*, 15(5):965–91, May 2003.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, pages 2951–2959, 2012.

Sen Song, Kenneth D Miller, and Lawerence F Abbott. Competitive Hebbian learning through spike-timing-dependent synaptic plasticitye. *Nature Neuroscience*, 3(9):919–26, September 2000. ISSN 1097-6256.

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 Computational Biology*, 11(10):1–30, 10 2015. doi: 10.1371/journal.pcbi.1004464.

Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: a structural description of the human brain. *PLoS Computational Biology*, 1(4):e42, 2005.

Olav Stetter, Demian Battaglia, Jordi Soriano, and Theo Geisel. Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. *PLoS Computational Biology*, 8(8):e1002653, 2012.

Ian Stevenson and Konrad Koerding. Inferring spike-timing-dependent plasticity from spike train data. *Advances in Neural Information Processing Systems*, pages 2582–2590, 2011.

Ian H Stevenson, James M Rebesco, Nicholas G Hatsopoulos, Zach Haga, Lee E Miller, and Konrad P Körding. Bayesian inference of functional connectivity and network structure from spikes. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 17(3):203–213, 2009.

Alan A Stocker and Eero P Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4):578–85, April 2006.

Yee Whye Teh and Michael I Jordan. Hierarchical Bayesian nonparametric models with applications. *Bayesian Nonparametrics*, pages 158–207, 2010.

Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581, 2006.

Joshua B Tenenbaum, Thomas L Griffiths, and Charles Kemp. Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, 10(7):309–318, 2006.

Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022):1279–1285, 2011.

Luke Tierney and Joseph B Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86, 1986.

Wilson Truccolo, Uri T. Eden, Matthew R. Fellows, John P. Donoghue, and Emery N. Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, 93(2):1074–1089, 2005. doi: 10.1152/jn.00697.2004.

Philip Tully, Matthias Hennig, and Anders Lansner. Synaptic and nonsynaptic plasticity approximating probabilistic inference. *Frontiers in Synaptic Neuroscience*, 6(8), 2014.

Srini Turaga, Lars Buesing, Adam M Packer, Henry Dalgleish, Noah Pettit, Michael Hausser, and Jakob Macke. Inferring neural population dynamics from multiple partial recordings of the same neural circuit. *Advances in Neural Information Processing Systems*, pages 539–547, 2013.

Leslie G Valiant. *Circuits of the Mind*. Oxford University Press, Inc., 1994.

Leslie G Valiant. Memorization and association on a realistic neural model. *Neural Computation*, 17(3):527–555, 2005.

Leslie G Valiant. A quantitative theory of neural computation. *Biological Cybernetics*, 95 (3):205–211, 2006.

Jurgen Van Gael, Yunus Saatci, Yee Whye Teh, and Zoubin Ghahramani. Beam sampling for the infinite hidden Markov model. *Proceedings of the International Conference on Machine Learning*, pages 1088–1095, 2008.

Michael Vidne, Yashar Ahmadian, Jonathon Shlens, Jonathan W Pillow, Jayant Kulkarni, Alan M Litke, EJ Chichilnisky, Eero Simoncelli, and Liam Paninski. Modeling the impact of common noise inputs on the network activity of retinal ganglion cells. *Journal of Computational Neuroscience*, 33(1):97–121, 2012.

Joshua T Vogelstein, Brendon O Watson, Adam M Packer, Rafael Yuste, Bruno Jedynak, and Liam Paninski. Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophysical Journal*, 97(2):636–655, 2009.

Joshua T Vogelstein, Adam M Packer, Timothy A Machado, Tanya Sippy, Baktash Babadi, Rafael Yuste, and Liam Paninski. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, 104(6):3691–3704, 2010.

Hermann von Helmholtz and James Powell Cocke Southall. *Treatise on Physiological Optics: Translated from the 3rd German Ed.* Optical Society of America, 1925.

Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

Yair Weiss, Eero P Simoncelli, and Edward H Adelson. Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6):598–604, 2002.

Mike West, P Jeff Harrison, and Helio S Migon. Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, 80(389):73–83, 1985.

John G White, Eileen Southgate, J Nichol Thomson, and Sydney Brenner. The structure of the nervous system of the nematode Caenorhabditis elegans: the mind of a worm. *Philosophical Transactions of the Royal Society of London: Series B (Biological Sciences)*, 314:1–340, 1986.

Louise Whiteley and Maneesh Sahani. Attention in a Bayesian framework. *Frontiers in Human Neuroscience*, 6, 2012.

Alexander B Wiltschko, Matthew J Johnson, Giuliano Iurilli, Ralph E Peterson, Jesse M Katon, Stan L Pashkovski, Victoria E Abraira, Ryan P Adams, and Sandeep Robert Datta. Mapping sub-second structure in mouse behavior. *Neuron*, 88(6):1121–1135, 2015.

Jesse Windle, Nicholas G Polson, and James G Scott. Sampling Pólya-gamma random variates: alternate and approximate techniques. *arXiv preprint arXiv:1405.0506*, 2014.

Frank Wood and Michael J Black. A nonparametric Bayesian alternative to spike sorting. *Journal of Neuroscience Methods*, 173(1):1–12, 2008.

Frank Wood, Jan Willem van de Meent, and Vikash Mansinghka. A new approach to probabilistic programming inference. *arXiv preprint arXiv:1507.00996*, 2015.

Tianming Yang and Michael N Shadlen. Probabilistic reasoning by neurons. *Nature*, 447 (7148):1075–80, June 2007.

Byron M. Yu, John P. Cunningham, Gopal Santhanam, Stephen I. Ryu, Krishna V. Shenoy, and Maneesh Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of Neurophysiology*, 102:614–635, 2009.

Alan Yuille and Daniel Kersten. Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10(7):301–308, 2006.

Richard S Zemel, Peter Dayan, and Alexandre Pouget. Probabilistic interpretation of population codes. *Neural Computation*, 10(2):403–30, February 1998.

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.

Mingyuan Zhou, Lingbo Li, Lawrence Carin, and David B Dunson. Lognormal and gamma mixed negative binomial regression. *Proceedings of the International Conference on Machine Learning*, pages 1343–1350, 2012.