Additive Continuous-time Joint Partitioning of Neural Variability

Adam S. Charles (adamsc@princeton.edu), Jonathan W. Pillow (pillow@princeton.edu)

Princeton University

Princeton, NJ, 08544, USA

abstract

Accurate estimation of neural spike-rates is challenging due to fact that both stimulus-dependent spike-rates and trial-by-trial noise are continuously time-varying and that neural spiking is well known to exhibit super- or sub-Poisson behavior. In particular, the time-varying nature of the noise makes spike-count statistics sensitive to choices in temporal bin-size selection (Cohen & Kohn, 2011). While methods have been proposed for both binless rate estimation and non-Poisson activity (R. Goris, Movshon, & Simoncelli, 2014; Charles, Park, Weller, Horwitz, & Pillow, 2018), no current over-dispersion model can perform arbitrary continuous-time rate estimation. We present here such a model, where we model the stimulus-based rate as a Gaussian Process (GP), and the rate driving the observed spiking is an additive combination of the stimulus GP and a noise process (also modeled as a GP), passed through a rectifying nonlinearity. Our model significantly generalizes previous over-dispersion models by both removing the bin-size dependence, as well as allowing estimation of the latent continuous-time spike-rates. Our model also explains the difference in statistics across bin sizes by accounting for temporal correlations. Given the noise parameters, we can estimate the stimulus GP via a maximum a-posteriori optimization, using a Laplace approximation to marginalize over the noise instantiations. We demonstrate out model both on simulated data as well as macaque V1 activity.

Keywords: Spiking statistics, over-dispersion, Poisson

Introduction

Relating observed neural spike trains to stimuli or behavior is a key to uncovering the neural code. While some methods generate latent variable models that attempt to describe the entire processing pipeline, more often the spike rate is estimated from the spike trains for more flexible exploratory analysis. Depending on the analysis, this estimation has historically been accomplished by averaging the number of spikes per bin over multiple identical trials, i.e. the Peristimulus time histogram (PSTH) (Gerstein & Perkel, 1969), or by inferring the rate per bin under a Poisson model. These estimations can be challenging, however, due to a number of factors. For one, both these methods rely on a seemingly arbitrary choice of bin-size; a quantity that should have no impact of the analysis and yet many statistics calculated with the binned rates are sensitive to the choice in bin size (Cohen & Kohn, 2011). Additionally, the Poisson assumption over the spike rates does not match the observed higher-order statistics of neural firing. Specifically, many neurons display over- or under-dispersion, meaning that the variance of the spiking process is above or below the mean (Shadlen & Newsome, 1998; Geisler & Albrecht, 1997; Eden & Kramer, 2010). Not accounting for this behavior in the rate estimation can further bias the resulting estimates.

Recent advances have tackled each of these challenges separately. For the time-bins, both advanced methods to select optimal bin-sizes (Shimazaki & Shinomoto, 2007) or fully continuous time Poisson models that bypass the need to bin at all (Mena & Paninski, 2014; Truccolo, Eden, Fellows, Donoghue, & Brown, 2005) have been proposed. To account for over- and under- dispersion, new models have been proposed to either augment or replace the Poisson model (R. Goris et al., 2014; Charles et al., 2018; Stevenson, 2016; Gao, Busing, Shenoy, & Cunningham, 2015; Pillow & Scott, 2012; Sellers, Borle, & Shmueli, 2012; Moshitch & Nelken, 2014). No method has yet, however, has performed bin-less inference of latent spike-rates under an over- or under-dispersed model

We present here such a model for over-dispersed spike rates. Our method is based on a continuous time extension of previous work that models over-dispersion as latent noise that effects the spike rate through a nonlinearity (Charles et al., 2018). By extending the Poisson per-bin count model to a full Poisson process, and extending the latent Gaussian noise to a continuous-time Gaussian Process (GP), we are able to model both of these challenging aspects.

Model details

Our model begins with a Poisson process that has a continuously changing rate function where the spiking process for a neuron responding to a given stimulus at trial m, $y_m(t)$, is a

$$y_m(t) = \mathsf{PoissonProcess}\left(f_\theta(x(t) + n_{k,m}(t))\right), \quad (1)$$

where $x(t)$ is the stimulus-dependent term, $n_m(t)$ is a stochastic (continuous-time) modulation, and $f_\theta(\cdot)$ is a non-linear rectifying function (here we use $f_\theta(\cdot) = \exp(\cdot)$). Letting $T_m[i]$ indicate the ith event for trial m for a
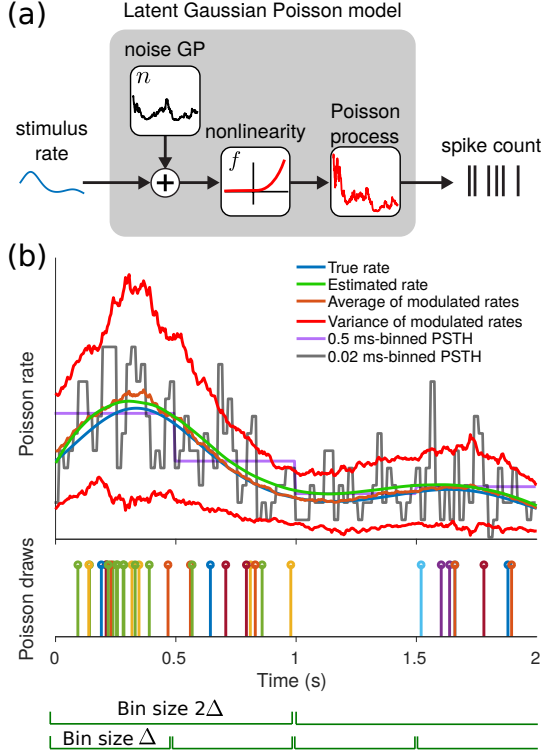
(a) Latent Gaussian Poisson model

Figure 1: (a) Our model is a continuous-time latent GP model, where the stimulus based activity is combined with additive noise and passed through a point-wise non-linearity before driving the spike train via an inhomogenious Poisson process. (b) An example draw from our model. The blue curve indicates the stimulus activity, the orange curve represents the average rate over 100 noise draws, and the red curves indicate the one-standard deviation spread. The larger variances at the higher rates shows the super-Poisson nature of our model. The purple and gray curves show how binned PSTH estimates either hide the latent rate's behavior (large bins) or give poor estimates of the rate under data constraints (small bins). Example spike trains drawn from this model are displayed in different colors. The green curve depicts the accuracy of our estimate to the latent stimulus rate.

given stimulus, the likelihood of the observed spike times given the underlying stimulus rate and additive modulation is

$$p(\{T_m[i]\}_{i=1,\dots,Y_m} \| x(t), n_m(t))$$
$$= e^{\left(\sum_{i=1}^{Y_m} x_k(T_m[i]) + n_m(T_m[i])\right)} e^{-\int_0^T e^{x(t)+n_m(t)} \, dt}.$$

To infer the stimulus-dependent rate $x(t)$, we introduce a Gaussian Process (GP) prior for both $x(t)$, and the ad-

ditive noise $n_n(t)$,

$$
\begin{aligned}
x(t) &\sim \mathcal{GP}\left(\mu_x, K_x(t_i, t_j)\right) \\
n_m(t) &\sim \mathcal{GP}\left(\mu_n, K_n(t_i, t_j)\right),
\end{aligned}
$$

where the covariance kernel $K_x(t_i, t_j)$ is taken to have a radial-basis function (RBF) form,

$$K(t_i, t_j) = \rho \exp\left(-\frac{|t_i - t_j|^p}{l^p}\right),$$

parametrized by the variance level ρ, the length-scale l and the exponential power p, which differ for $n_m(t)$ and $x(t)$. Since $n_m(t)$ are nuisance variables which we do not wish to infer, we instead focus on marginalizing out these variables, instead using the spiking data for all trials to only estimate $x(t)$. Additionally, it is impossible to infer $x(t)$ for all t on a continuous domain, so instead we infer $x(t)$ at a subset $\overline{T}_m[i]$ that sufficiently span the duration of the trial. Letting $\overline{x}$ and $\overline{n}_m$ be $x(t)$ and $n_m(t)$ evaluated at each $\overline{T}_m[i]$, we seek to solve

$$\arg\min_{\overline{x}} \int \prod_{m=1}^{M} \left[\int\int p(T_m|x, n_m, \overline{x}, \overline{n}_m) p(n_m|\overline{n}_m) \cdot \right.$$
$$\left. p(\overline{n}_m)\, d\overline{n}_m \, dn_m \right] p(x|\overline{x})\, dx\, p(\overline{x}),$$

which can have the integrals with respect to n_m and x calculated analytically to yield

$$\arg\min_{\overline{x}} e^{\frac{1}{2}\beta} \prod_{m=1}^{M} \left[\int e^{-(we^{\overline{x}})^T e^{\overline{n}_m}} e^{\mathbf{1}^T K_{n,\overline{n},m} K_{\overline{n},\overline{n}}^{-1} \overline{n}_m} \cdot \right.$$
$$\left. p(\overline{n}_m)\, d\overline{n}_m \right] e^{\mathbf{1}^T K_{x,\overline{x}} K_{\overline{x},\overline{x}}^{-1} \overline{x}} p(\overline{x}),$$

where $K_{n,\overline{n},m}$ and $K_{x,\overline{x}}$ are the noise and stimulus covariance matrices between the estimate times and the observed spike times on the mth trial, $K_{\overline{n},\overline{n}}$ and $K_{\overline{x},\overline{x}}$ are the noise and stimulus covariance matrix only between the estimation times, and β depends only on the GP parameters and covariance matrices. To evaluate the remaining integral, we use a Laplace approximation, allowing us to use gradient descent optimization to solve for $\overline{x}$.

Results

To assess our model, we validate the estimation accuracy both in simulation and on data from macaque V1. We first sample from the process as in Figure 1b and assess the recovery of the latent process as a function of the number of trials available. Our estimation is fairly accurate, as measured by the relative mean-squared error (rMSE) and becomes more accurate with an increased number of trials (Fig. 2).

Finally, we compare the log-normal process model and the negative binomial (NB) model on data from 113 V1 cells collected in macaque V1 by Graf et al. (Graf, Kohn,

Jazayeri, & Movshon, 2011) This dataset contains spike-time data collected from all 113 neurons from 72 different stimulus orientations, each presented 50 times (50 trials). First we fit the negative binomial model as in (R. Goris et al., 2014) by fitting g to have a Gamma hyper-prior distribution. Figure 3 demonstrates that the variance fit to this hyper distribution is not consistent when the data is binned at different bin sizes. To fit the log normal process, we manually set the two parameters p and l in the log-normal process covariance kernel. Figure 3 also displays that the fit log-normal process better captures the data variance as a function of bin size.

Figure 3: Two different V1 neurons (data collected by Graf et al.) demonstrate different cross-bin-size statistics. Each plot shows the relative variance across trials as a function of bin-size (normalized to the Poisson variance). Our model can explain the variance over all time bins, as opposed to the Poisson model and a recent over-dispersion model (R. Goris et al., 2014).
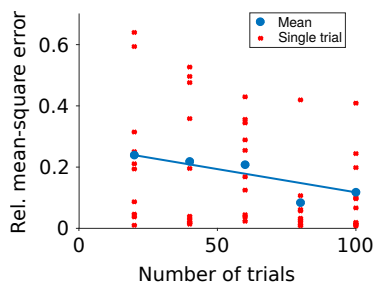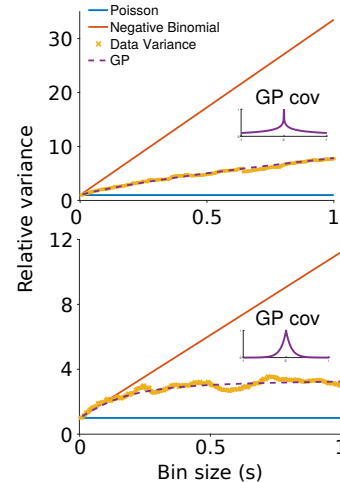

Figure 2: Recovery of spike-rates from simulated data indicates that our method is accurate (most latent rates estimated to within 10% relative mean-squared error), and our estimation accuracy improves with additional data (more trials).

Conclusions

We present here a new model of continuous-time over-dispersion and a related estimation procedure for inferring the latent stimulus-related rate directly from spike-trains with no need for binning. Our model is influenced by recent work in characterizing over-dispersion in neural firing (R. Goris et al., 2014; N. R. R. Goris, Ballé, & Simoncelli, 2015; Rabinowitz, Goris, Cohen, & Simoncelli, 2015; Charles et al., 2018). Specifically, our model can be considered a continuous-time extension of previous work on flexible over-dispersion models for binned data (Charles et al., 2018). The extension to continuous time incurs extra complexity in that enough samples of a Gaussian Process need to be estimated to approximate an integral. While in our method here we use a simple Riemann approximation, other methods such as the quadrature methods (e.g. (Mena & Paninski, 2014)) to approximating the integral could be easily used by modifying a single weight vector. The ability of our method to bypass binning while still accounting for over-dispersion warrants additional inquiry. For example, developing fast tools for inferring the latent rate and understanding the ef-

fects of the GP parameters on the estimates would make this methods more applicable in a wider setting.

Finally, while our model assumes a particular form for the non-stimulus contribution to spiking (i.e. the noise GP), there are many mechanisms that could cause such a contribution. Being able to better partition the two additive components without the confounding factor of binning, and to tune the time-scale of the noise model, can lead to improved understanding of the entire neural circuit in addition to improved rate estimation.

References

Charles, A. S., Park, M., Weller, J. P., Horwitz, G. D., & Pillow, J. W. (2018). Dethroning the fano factor: a flexible, model-based approach to partitioning neural variability. *Neural Computation*, *30*(4), 1012–1045. (Accepted)

Cohen, M., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature neuroscience*, *14*(7), 811–819.

Eden, U., & Kramer, M. (2010). Drawing inferences from fano factor calculations. *Journal of neuroscience methods*, *190*(1), 149–152.

Gao, Y., Busing, L., Shenoy, K., & Cunningham, J. (2015). High-dimensional neural spike train analysis with generalized count linear dynamical systems. In *Advances in neural information processing systems* (pp. 2044–2052).

Geisler, W. S., & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: detection, discrimina-

tion, and identification. *Vis Neurosci*, *14*(5), 897–919.

Gerstein, G. L., & Perkel, D. H. (1969). Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science*, *164*(3881), 828–830.

Goris, N. R. R., Ballé, J., & Simoncelli, E. (2015). A model of sensory neural responses in the presence of unknown modulatory inputs. *arXiv preprint arXiv:1507.01497*.

Goris, R., Movshon, J., & Simoncelli, E. (2014, Jun). Partitioning neuronal variability. *Nat Neurosci*, *17*(6), 858–865.

Graf, A. B., Kohn, A., Jazayeri, M., & Movshon, J. A. (2011). Decoding the activity of neuronal populations in macaque primary visual cortex. *Nature neuroscience*, *14*(2), 239–245.

Mena, G., & Paninski, L. (2014). On quadrature methods for refractory point process likelihoods. *Neural computation*, *26*(12), 2790–2797.

Moshitch, D., & Nelken, I. (2014). Using tweedie distributions for fitting spike count data. *Journal of neuroscience methods*, *225*, 13–28.

Pillow, J. W., & Scott, J. G. (2012). Fully bayesian inference for neural models with negative-binomial spiking. In *Nips* (pp. 1907–1915).

Rabinowitz, N., Goris, R., Cohen, M., & Simoncelli, E. (2015). Attention stabilizes the shared gain of v4 populations. *eLife*, e08998.

Sellers, K. F., Borle, S., & Shmueli, G. (2012). The compoisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry*, *28*(2), 104–116.

Shadlen, M., & Newsome, W. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience*, *18*, 3870–3896.

Shimazaki, H., & Shinomoto, S. (2007). A method for selecting the bin size of a time histogram. *Neural computation*, *19*(6), 1503–1527.

Stevenson, I. H. (2016). Flexible models for spike count data with both over-and under-dispersion. *Journal of computational neuroscience*, *41*(1), 29–43.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). 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.