Dynamic Modeling of Spike Count Data with Conway-Maxwell Poisson Variability

Ganchao Wei*

ganchao.wei@uconn.edu

Department of Statistics, University of Connecticut, Storrs, CT, 06250, USA

Ian H. Stevenson

ian.stevenson@uconn.edu

Department of Psychological Sciences and Department of Biomedical Engineering, University of Connecticut, Storrs, CT, 06250, USA

Abstract

In many areas of the brain, neural spiking activity covaries with features of the external world, such as sensory stimuli or an animal's movement. Experimental findings suggest

---

*Corresponding author

that the variability of neural activity changes over time and may provide information about the external world beyond the information provided by the average neural activity. To flexibly track time-varying neural response properties, here we developed a dynamic model with Conway-Maxwell Poisson (CMP) observations. The CMP distribution can flexibly describe firing patterns that are both under- and over-dispersed relative to the Poisson distribution. Here we track parameters of the CMP distribution as they vary over time. Using simulations, we show that a normal approximation can accurately track dynamics in state vectors for both the centering and shape parameters (λ and ν). We then fit our model to neural data from neurons in primary visual cortex, "place cells" in the hippocampus and a speed-tuned neuron in anterior pretectal nucleus. We find that this method outperforms previous dynamic models based on the Poisson distribution. The dynamic CMP model provides a flexible framework for tracking time-varying non-Poisson count data and may also have applications beyond neuroscience.

1 Introduction

Although many models of neural activity assume that neurons respond with stable responses to external sensory stimuli or movements, there is substantial evidence that neural spiking activity changes over time due to adaptation and plasticity (Brown et al., 2001; Lesica et al., 2007) as well as spontaneously (Rokni et al., 2007; Tomko and Crapper, 1974). At the same time, a neuron's spiking responses on individual trials can be highly variable, even in the controlled settings with constant stimuli. In most previous research, trial-to-trial neural variability is assumed to be Poisson distributed. However, spike count distributions can be substantially more or less variable than Poisson (Maimon and Assad, 2009; Amarasingham et al., 2006; DeWeese et al., 2003; Kara et al., 2000), and that the variability also appears to change over time, in many cases (Churchland et al., 2010, 2011). Here we introduce a dynamic model with Conway-Maxwell Poisson observations that can describe non-Poisson spike statistics and track

changing response properties.

Variability appears to be an increasingly important feature of neural responses and can act as a signature of decision making (Churchland et al., 2011), movement preparation (Churchland et al., 2006), or stimulus onset (Churchland et al., 2010). Although systems neuroscience has a long history of studying how external variables influence mean firing rates, less is known about response variability. Neural activity changes on different timescales, and distinguishing changes in variability from changes in the mean response based on sparse, spike observations is a nontrivial statistical challenge (DeWeese and Zador, 1998). Statistical tools to accurately track the sources of variability within a given experiment may be useful for understanding neural systems. There has been substantial work developing dynamic Poisson models (Brown et al., 2001; Eden et al., 2004), as well as, other Poisson models that can account for fluctuating response properties with latent or observed variables (Czanner et al., 2008; Smith and Brown, 2003). Several models of neural activity with non-Poisson observations have also been described (DeWeese et al., 2003; Gao et al., 2015; Pillow and Scott, 2012), including a static model with Conway-Maxwell Poisson observations (Stevenson, 2016). Each of these models, however, whether static or dynamic, assumes a fixed mean-variance relationship (i.e. fixed dispersion parameters). Here, to flexibly track how neural variability might change over time, we explicitly consider changes in both the mean and dispersion.

Here we develop a dynamic GLM with Conway-Maxwell-Poisson (CMP) observations. The CMP distribution can account for both over- and under-dispersion in spike count data. To get the closed-form posterior for state vectors with CMP likelihood, we fit the model using a global Gaussian approximation (Laplace approximation). Since the state-space of the dynamic model has Markovian structure, inference is efficient with this approximation, and we estimate the process noise by maximizing the predictive likelihood. After illustrating the proposed method in simulations, we apply it to neu-

ral activity from primary visual cortex, place cells in the hippocampus and a speed-tuned neuron in anterior pretectal nucleus. The dynamic CMP model can track changes in both the mean and variance of neural responses and outperforms previous Poisson models.

2   Methods

Here we consider a dynamic GLM with Conway-Maxwell Poisson (CMP) observations to describe time-varying spike counts. We first introduce the model. Although the CMP distribution allows us to flexibly model non-Poisson variability, one major challenge with using this model is that there are no closed-form posteriors for the CMP likelihood. Here, we fit the model using a global Gaussian approximation, and we discuss several additional technical challenges that arise when using the CMP distribution with a dynamic GLM. Code is available at https://github.com/weigcdsb/COM_POISSON.

2.1   Dynamic Conway-Maxwell Poisson Model

A count observation y, such as the spike count for a neuron, is assumed to follow the CMP distribution, with parameters λ and ν. The probability mass function (pmf) of CMP distribution is:

$$P(Y = y|\lambda, \nu) = \frac{\lambda^y}{(y!)^\nu} \cdot \frac{1}{Z(\lambda, \nu)}, \tag{1}$$

where $Z(\lambda, \nu) = \sum_{k=0}^{\infty} \frac{\lambda^k}{(k!)^\nu}$ is the normalizing constant. The shape parameter $\nu \geq 0$ controls different dispersion patterns, i.e. equi- ($\nu = 1$), over- ($0 \leq \nu < 1$) or under-dispersion ($\nu > 1$). Three common distributions occur as special cases: 1) the Poisson ($\nu = 1$), 2) the geometric ($\nu = 0, \lambda < 1$), and 3) the Bernoulli ($\nu \to \infty$).

For multiple observations up to T steps, such as simultaneous spike counts from n neurons, denote the counts at time bin t as $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{nt})'$, for $t = 1, \ldots, T$. The

corresponding CMP parameters at t are $\boldsymbol{\lambda}_t = (\lambda_{1t}, \ldots, \lambda_{nt})'$ and $\boldsymbol{\nu}_t = (\nu_{1t}, \ldots, \nu_{nt})'$. Previous work has examined the CMP-GLM (Chatla and Shmueli, 2018; Sellers and Shmueli, 2010), and here we focus on the dynamic version of this GLM. The CMP parameters at t are modeled by two log-linear models, $\log \boldsymbol{\lambda}_t = \boldsymbol{X}_t \boldsymbol{\beta}_t$ and $\log \boldsymbol{\nu}_t = \boldsymbol{G}_t \boldsymbol{\gamma}_t$, with $\boldsymbol{\beta}_t \in \mathbb{R}^p$ and $\boldsymbol{\gamma}_t \in \mathbb{R}^q$, and $\boldsymbol{X}_t$ and $\boldsymbol{G}_t$ denote known predictors. Under the CMP-GLM, the parameters are static. Here, we assume that they progress linearly with Gaussian noise.

The observations follow conditionally independent CMP distributions, given the state vector $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t', \boldsymbol{\gamma}_t')'$.

$$\boldsymbol{y}_t \sim CMP(\boldsymbol{\lambda}_t, \boldsymbol{\nu}_t),$$
$$\log \boldsymbol{\lambda}_t = \boldsymbol{X}_t \boldsymbol{\beta}_t, \quad \log \boldsymbol{\nu}_t = \boldsymbol{G}_t \boldsymbol{\gamma}_t. \tag{2}$$

While the state vector $\boldsymbol{\theta}_t$ evolves linearly with Gaussian noise:

$$\boldsymbol{\theta}_1 \sim N_{p+q}(\boldsymbol{\theta}_0, \boldsymbol{Q}_0),$$
$$\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1} \sim N_{p+q}(\boldsymbol{F}\boldsymbol{\theta}_{t-1}, \boldsymbol{Q}). \tag{3}$$

Given the initial state mean $\boldsymbol{\theta}_0$, covariance $\boldsymbol{Q}_0$, linear dynamics $\boldsymbol{F}$ and process covariance $\boldsymbol{Q}$.

2.2   Inference by Gaussian approximation

To fit the model to data we need to estimate the time-varying state vector $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_T')$ $\in \mathbb{R}^{(p+q)T}$. In this section, we first assume $\boldsymbol{F}$ and $\boldsymbol{Q}$ are known. Since the observations are CMP distributed, we cannot estimate $\boldsymbol{\Theta}$ in closed form. Instead, here we approximate it by a multivariate Gaussian distribution, $P(\boldsymbol{\Theta}|\boldsymbol{Y}) \approx N_{(p+q)T}(\boldsymbol{\Theta}|, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{Y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_T')'$. The parameters of this Gaussian are found by a global Laplace approximation, i.e. $\boldsymbol{\mu} = \arg\max_{\boldsymbol{\Theta}} P(\boldsymbol{\Theta}|\boldsymbol{Y})$ and $\boldsymbol{\Sigma} = -(\nabla\nabla_{\boldsymbol{\Theta}} \log P(\boldsymbol{\Theta}|\boldsymbol{Y})|_{\boldsymbol{\Theta}-\boldsymbol{\mu}})^{-1}$. The log-posterior is given by:

$$\log P(\boldsymbol{\Theta}|\boldsymbol{Y}) = \sum_{t=1}^{T} l_t - \frac{1}{2}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)' \boldsymbol{Q}_0^{-1}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$$

$$- \frac{1}{2}\sum_{t=2}^{T}(\boldsymbol{\theta}_t - \boldsymbol{F}\boldsymbol{\theta}_{t-1})'\boldsymbol{Q}^{-1}(\boldsymbol{\theta}_t - \boldsymbol{F}\boldsymbol{\theta}_{t-1}), \qquad (4)$$

$$l_t = l(\boldsymbol{\theta}_t) = \log P(\boldsymbol{y}_t|\boldsymbol{\theta}_t) = \sum_{i=1}^{n} y_{it} \log \lambda_{it} - \nu_{it}y_{it}! - \log Z(\lambda_{it}, \nu_{it}),$$

where $l(\cdot)$ is the log-likelihood. The log-posterior is concave (Gupta et al., 2014), and the Markovian structure of the state vector dynamics makes it possible to optimize by Newton-Raphson (NR) in $\mathcal{O}(T)$ time (Paninski et al., 2010).After the Newton update, we can further quantify the uncertainty for the CMP parameters and the underlying rates, as in Appendix A.

There are several technical challenges involved with performing the Newton update with CMP observations. Firstly, in order to find the gradient and Hessian we need to calculate moments of Y_{it} and $\log Y_{it}!$, which have no closed forms (Shmueli et al., 2005). We can calculate these moments by truncated summation. However, when $\lambda \geq 2$ and $\nu \leq 1$, truncated summation is computationally costly since we need many steps for accurate approximation. In this case, we approximate the moments using previous (Chatla and Shmueli, 2018; Gaunt et al., 2019) asymptotic results as in Appendix B. A second challenge is that the Hessian is not robust to outliers. Outliers often result in the Hessian being close to singular or even positive-definite. See details in Appendix C. To ensure robustness, we use Fisher scoring where the observed information is replaced by the expected information. Finally, a third challenge is that the Newton updates take a long time to converge if the initial state estimate is far from the maximum of the posterior, especially when T is large. To resolve this issue, we use a smoothing estimate with local Gaussian approximation as a "warm start". Forward filtering for a dynamic Poisson model has been previously described in Eden et al. (2004), and here we implement CMP filtering following the same rationale. Let $\boldsymbol{\theta}_{t|t-1} = E(\boldsymbol{\theta}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1})$

and $\boldsymbol{\Sigma}_{t|t-1} = Var(\boldsymbol{\theta}_t|\boldsymbol{y}_1, \dots, \boldsymbol{y}_{t-1})$ be the mean and variance for the one-step prediction density and $\boldsymbol{\theta}_{t|t} = E(\boldsymbol{\theta}_t|\boldsymbol{y}_1, \dots, \boldsymbol{y}_t)$ and $\boldsymbol{\Sigma}_{t|t} = Var(\boldsymbol{\theta}_t|\boldsymbol{y}_1, \dots, \boldsymbol{y}_t)$ be mean and variance for the posterior density, then the filtering update for step t is given by

$$
\begin{aligned}
\boldsymbol{\theta}_{t|t-1} &= \boldsymbol{F}\boldsymbol{\theta}_{t-1|t-1}, \\
\boldsymbol{\Sigma}_{t|t-1} &= \boldsymbol{F}\boldsymbol{\Sigma}_{t-1|t-1}\boldsymbol{F}' + \boldsymbol{Q}, \\
\boldsymbol{\theta}_{t|t} &= \boldsymbol{\theta}_{t|t-1} + (\boldsymbol{\Sigma}_{t|t}) \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}_t}\right]_{\boldsymbol{\theta}_{t|t-1}}, \\
\boldsymbol{\Sigma}_{t|t}^{-1} &= \boldsymbol{\Sigma}_{t|t-1}^{-1} - \left[\frac{\partial^2 l_t}{\partial \boldsymbol{\theta}_t \partial \boldsymbol{\theta}_t'}\right]_{\boldsymbol{\theta}_{t|t-1}}.
\end{aligned}
\tag{5}
$$

Here, to again ensure robustness, we use Fisher scoring when updating the state covariance. We then find smoothed estimates using a backward pass (Rauch et al., 1965). Although doing smoothing is fast, the estimates can be inaccurate, especially when there are large changes in the state vector. In the forward filtering stage, the Gaussian approximation at each step t is conducted locally at the recursive prior $\boldsymbol{\theta}_{t|t-1}$. This will be statistically inefficient when the recursive prior is too far away from the posterior mode, or when there is a large change in the state vector. Moreover, Fisher scoring reduces the efficiency of the smoother even further. The smoother provides reasonable initial estimates, but estimation accuracy is substantially improved by using Newton's method to find the global Laplace approximation for the posterior.

2.3 Estimating process noise

For the applications to neural data examined here, we assume that $\boldsymbol{F} = \boldsymbol{I}$. However, we still need to estimate the process noise $\boldsymbol{Q}$. When n is small, especially when $n = 1$, different $\boldsymbol{Q}$ values will have a substantial influence on estimation. One possible way to estimate $\boldsymbol{Q}$ is to use an Expectation Maximization (EM) algorithm as in Macke et al. (2011). However, using the Laplace approximation for $\boldsymbol{\Theta}$ during E-step breaks the usual guarantee of non-decreasing likelihoods in EM, and, hence, may lead to divergence.

To avoid that, we could sample the posterior directly by MCMC. However, the lack of closed-form moments for the CMP distribution makes sampling computationally intensive. Here, to estimate Q robustly and quickly, we instead assume Q is diagonal and estimate it by maximizing the prediction likelihood in the filtering stage, as in Wei and Stevenson (2021).

3   Results

3.1   Tracking the mean and dispersion of spike counts over time

To illustrate how the dynamic CMP model can track both time-varying mean and dispersion, we simulated a neuron with a time-varying tuning curve, where the response to 100 evenly-spaced hypothetical visual stimuli shifts over 100 trials. Here, the neuron's tuning curve is determined by a linear combination of cubic B-spline basis functions with equally-spaced knots. The stimulus that evokes the highest average response – the "preferred orientation" – is initially 80 deg, but shifts over the course of the experiment, and the response amplitude also increases over time (Fig. 1A). Meanwhile, the dispersion pattern also changes: the responses are initially over-dispersed relative to a Poisson distribution and then become under-dispersed (Fig. 1B). Noisy observations are sampled from the Conway-Maxwell Poisson distribution at each time (Fig. 1C), mimicking the types of experimental observations collected during adaptation experiments in primary visual cortex (Dragoi et al., 2000). We then fitted the simulated spike observations using the same predictor variables as the generative model: the covariates for λ_t capture the tuning curve with X_t as a 10-knot cubic spline basis expansion of the orientation, and the covariate for the shape parameter ν_t does not depend on the stimulus orientation G_t. The fitted results match the ground truth well, for both the mean (Fig. 1C) and Fano factors (variance-to-mean ratio, Fig. 1D).

This model-based approach provides estimates of tuning curves and dispersion at each

time point. In cases where the tuning curve and variability change simultaneously, this approach can efficiently track both. By using the model with CMP observations, rather than Poisson or negative-binomial observations, the Fano factor can be both < 1 (under-dispersed) and > 1 (over-dispersed).
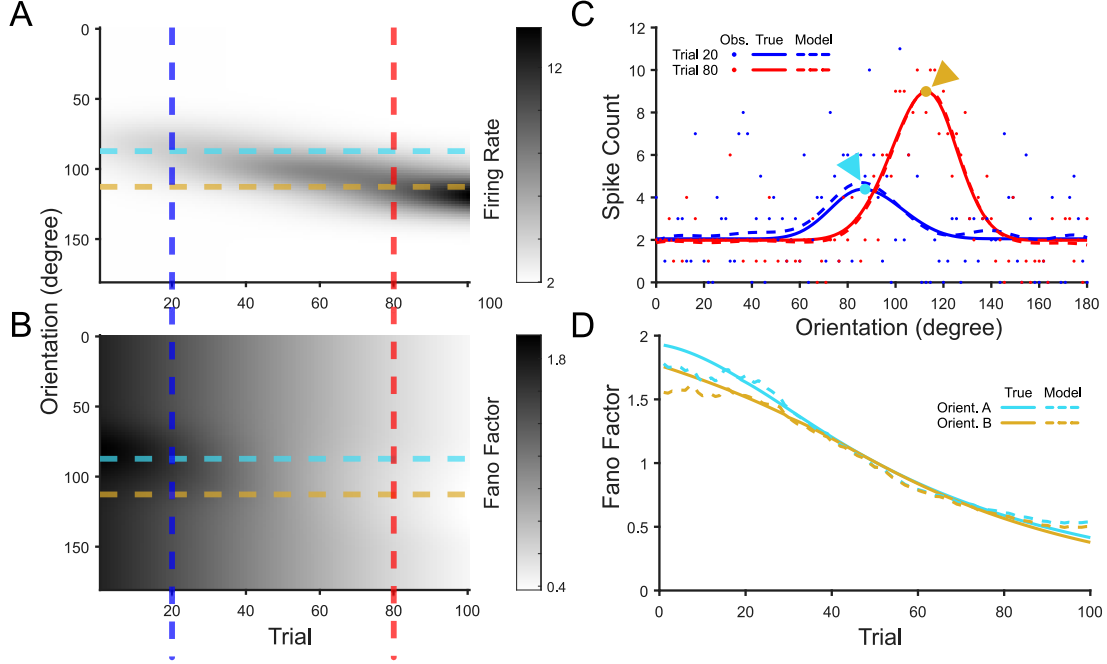

Figure 1: **A simulated neuron with a shifting firing and dispersion pattern.** **(A)** The tuning curve of the neuron shifts over time, with the preferred stimulus orientation changing and the response amplitude increasing. **(B)** At the same time, the variability in spiking changes from being over-dispersed relative to a Poisson distribution to under-dispersed. This leads to a decreasing Fano factor (variance-to-mean ratio) from 1.9 to 0.4 overall. **(C)** To illustrate the shifts, we show the tuning curve at two time points: Trial 20 (blue) and Trial 80 (red). Dots denote observed spike counts. The solid lines are the ground truth in mean firing rate, while the corresponding dashed lines are the fitted values. **(D)** To illustrate the shift in dispersion over time we show the true (solid) and estimated (dashed) Fano factor for two specific stimuli as a function of time. The dispersion for the early preferred orientation is shown in cyan, while the dispersion for the late preferred orientation is shown in yellow.

Changes in tuning have been widely documented in systems neuroscience both due to changing environment and spontaneous nonstationarity. Changes in variability also occur, but have been less well studied. With the CMP model, the mean and dispersion are both tracked and, thus, changes in variability can occur even when the mean is

stable. To illustrate this potential, we simulated a neuron whose mean firing rate is controlled to be constant, but whose Fano factor varies over time. Here X_t is a 5-knot cubic B-spline basis expansion of the orientation and $G_t = 1$. The model recovers the true mean firing rate (Fig. 2A) and capture the fluctuations in variance (Fig. 2B) at the same time. However, the estimated Fano factor is somewhat oversmoothed when the process noise Q is optimized by maximizing the predictive likelihood (see 2.3).
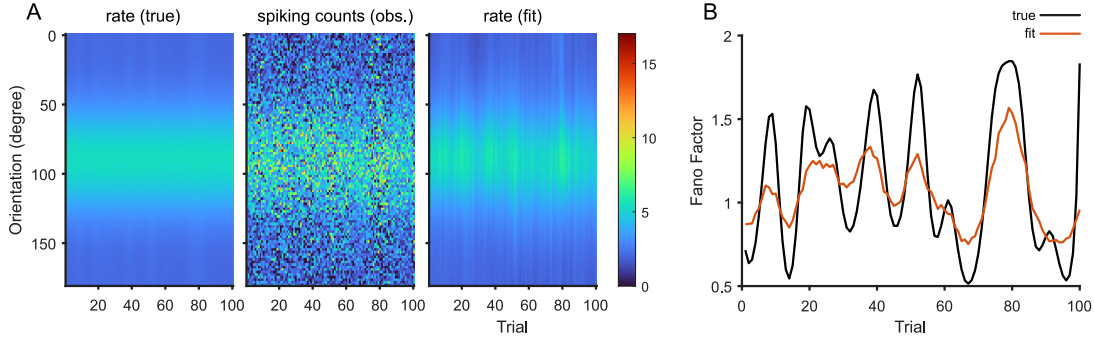

Figure 2: **Constant mean with fluctuations in dispersion.** **(A)** The first two panels show the true mean firing rate and the simulated observations. The last panel show the fitted mean response. **(B)** Although the mean response is constant, the Fano factor varies across the trial (black line). The colored line show the fitted result.

Although dynamic Poisson models have been applied in some neuroscientific settings, when spike counts are not Poisson distributed the model estimates can be biased. Since the dispersion influences estimates of the process noise Q, estimates of the mean in the dynamic Poisson model can be effected by over- or under-dispersion. To illustrate this interaction here we simulate a place cell from the hippocampus whose "place field" drifts over time. The true mean is determined by a Gaussian function where the preferred position varies over time. The spike counts are then generated by CMP distributions, here over-dispersed with constant shape parameter $\nu_t = 0.1$. We fit 1000 observations randomly sampled from 100 "runs" of a linear track. We find that, in this data-limited regime, the dynamic Poisson model and the dynamic CMP model give substantially different estimates of the time-varying place field (Fig. 3A). The dynamic Poisson model, in this case, under-estimates the firing rate at the true preferred position
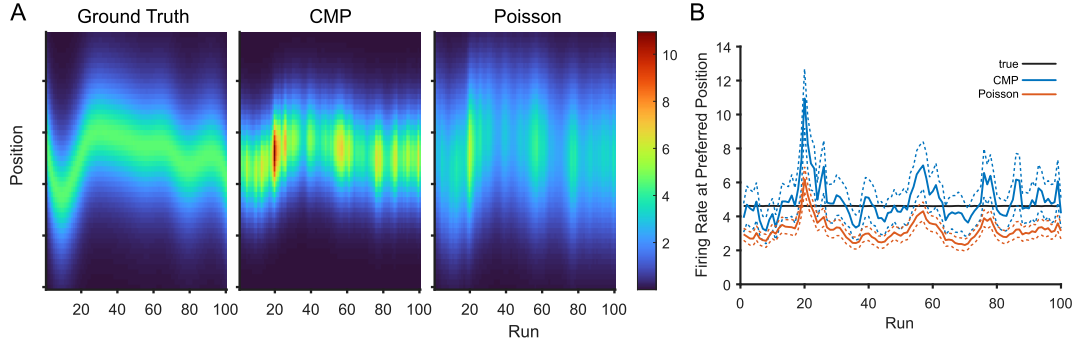
and under-estimates the uncertainty (Fig. 3B).

Figure 3: **Dynamic CMP and dynamic Poisson estimates differ.** Simulated over-dispersed place cell spiking is generated by the dynamic CMP model with $\nu_t = 0.1$. **(A)** We then fit dynamic CMP and dynamic Poisson models with $\boldsymbol{X}_t$, a 2-knot B-spline expansion for position ($\boldsymbol{G}_t = 1$ for the CMP). **(B)** When evaluating the response at the true preferred position for each run, the dynamic Poisson estimates are biased (under-estimated) and the uncertainty is also underestimated. The solid line gives the MAP estimates of mean firing rate, and the dashed lines show one S.D. credible intervals. The standard deviations of dynamic CMP estimates are calculated using the truncated summations (see details in Appendix B), while the standard deviations for the dynamic Poisson model are from a log-normal distribution.

3.2 Application to Experimental Data

We next applied our method to three publicly available datasets of extracellular spike recordings: 1) Utah array recordings of visually evoked activity from anesthetized macaque primary visual cortex ("V1 data"), 2) multi-shank silicon probe recordings from hippocampus of a rat running back-and-forth on a linear maze ("HC data") and 3) a speed-tuned neuron recorded from the anterior pretectal nucleus (APN) of an awake mouse ("APN neuron")

3.2.1 V1 Data

In the V1 dataset. CRCNS pvc-11 (Kohn and Smith, 2016), anesthetized macaque monkeys viewed full-field sinusoidal grating movies while neural activity was recorded by a 96-channel "Utah" array. Extracellular spiking activity was recorded on each electrode, and spike waveform segments were sorted by hand with modified competitive

mixture decomposition methods (Shoham et al., 2003). Here we use data from one animal (Monkey 1) viewing a movie of drifting sinusoidal gratings with 100 different drift directions presented in pseudorandom order (300ms each, 30s movie in total), and the movie was repeated 120 times. Here we analyze spike counts following each stimulus presentation from the period 50-350ms after stimulus onset, considering the response delay. For further details on how the data were obtained, see Kelly et al. (2010); Smith and Kohn (2008).

As with many neurons in visual cortex, the responses of the neurons in this dataset are tuned to the stimulus direction. Neurons respond to some directions of stimuli more than others, but the spike counts from trial to trial are highly variable. Here, we are specifically interested in tracking changes tuning curves and changes in variability over time. Fig. 4A shows responses from one example neuron with a preferred direction around 240 deg. This neuron is somewhat direction insensitive, and also responds with increased spiking to stimuli moving in the opposite direction, around 70 deg. After fitting the dynamic CMP model to these data, we find that the tuning curve itself appears mostly stable, but the overall firing rate increases over the course of the recording (Fig. 4B). At the same time, the Fano factor decreases over the course of the recording (Fig. 4C).

Although the data here is structured in 120 "trials" the data are collected sequentially, and we model nonstationary at the level of individual observations. For the predictors X_t and G_t we use cubic B-spline basis functions with periodic boundary conditions over the grating directions. Results for the example neuron use 5 and 3 equally-spaced knots for X_t and G_t, respectively. Fitting the model with half of the data (in a speckled hold-out pattern) gives patterns for the mean response (Fig. 4B) and Fano factor (Fig. 4C) that are similar to those using the full data. However, since the model-based approach provides a continuous estimate of the state vectors, the Fano factor estimated by the dynamic CMP model differs from a simple estimate of the Fano factor calculated

using a sliding window (Fig. 4C).

We then compare the performance of multiple models on data from all 74 neurons in this recording (Fig. 4D). We assess four dynamic models: (1) dynamic CMP, with 5 knots for X_t and 3 knots for G_t, denoted as dCMP-(5,3); (2) dynamic CMP with $G_t = 1$, dCMP-(5,1); (3) dynamic CMP with constant ν_t, dCMP-(5)-ν (fit by coordinate descent) and (4) a dynamic Poisson model, dPoi-(5). Additionally, we assess three static models: (1) static CMP, sCMP-(5,3); (2) static CMP with $G_t = 1$, sCMP-(5,1) and (3) static Poisson, sPoi-(5). The held-out log-likelihoods relative to a homogeneous static Poisson model show that the CMP-based models, both dynamic and static, outperform the Poisson-based models (Fig. 4D). The dynamic models perform slightly better than the corresponding static models, on average. The best performance on test data comes from modeling nonstationarity and stimulus-dependence with the full dynamic CMP model dCMP-(5,3). However, the benefit of adding nonstationary shape parameter (dCMP-(5)-ν vs. dCMP-(5,1)) and of adding stimulus-dependent shape parameter (dCMP-(5,1) vs dCMP-(5,3)) tend to be small for these data.

3.2.2 HC Data

In the HC dataset, CRCNS hc-3 (Mizuseki et al., 2013), a rat was running back and forth along a 250cm linear track. Extra cellular spiking activity was recorded in dorsal hippocampus using multi-shank silicon probes. Spikes were automatically sorting using KlustaKwik followed by manual adjustment (Rossant et al., 2016). Here we use data from one 66 min recording session (ec014-468) and analyze spike counts in 200ms bins. For further details on how the data were obtained, see Mizuseki et al. (2014)

As with many neurons in hippocampus, the responses of the neurons in this dataset are tuned to the rat's position along the track. Neurons spike at specific locations, but the place fields can also shift over time and the spike counts from run to run are highly variable. Fig. 5A shows an example from one neuron with two place fields where
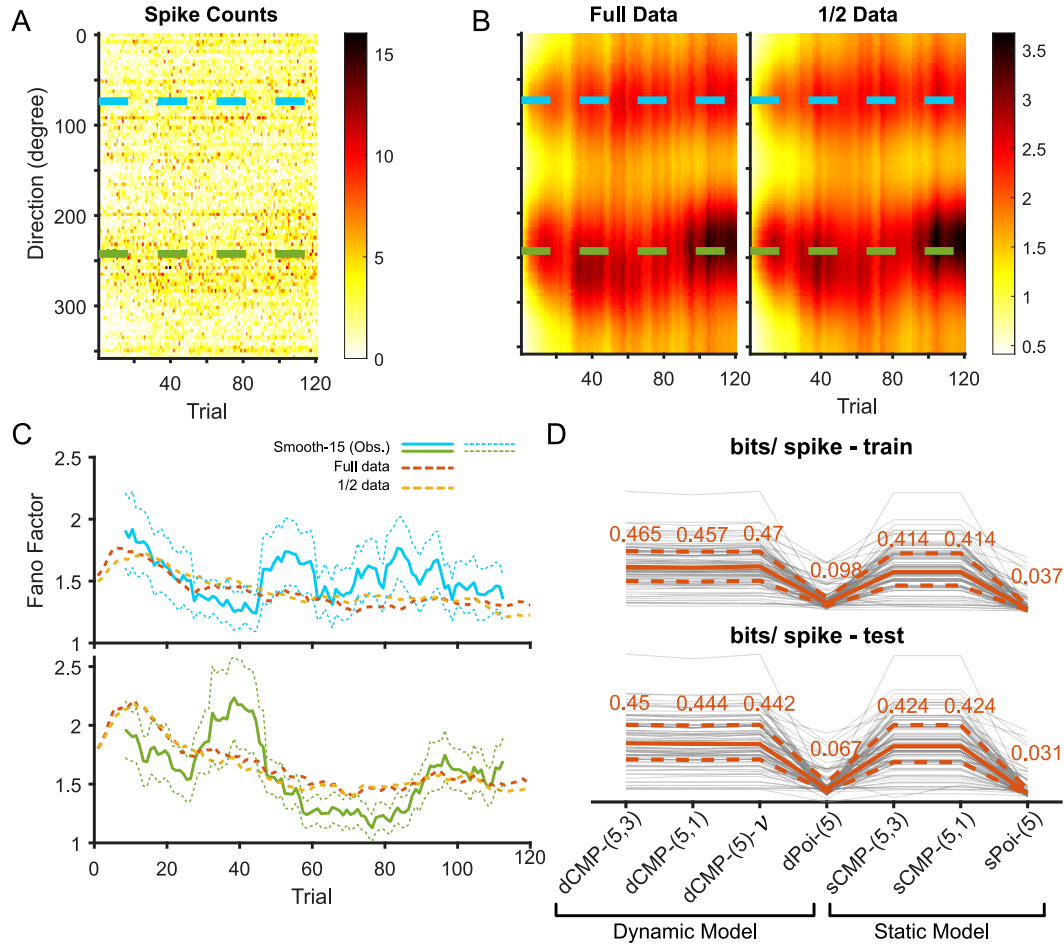
Figure 4: **Modeling nonstationary spiking from visual cortex. (A)** The spike counts of one example neuron from V1 in response to drifting grating stimuli with different drift directions presented over 120 trials. Two preferred directions (estimated by the CMP model) are marked by the dashed lines. **(B)** Estimated mean for the dynamic CMP model dCMP-(5,3) when fit to all the data and only half of the observations (held out in a speckled pattern). **(C)** Fano factor estimates for the two models, along with a direct estimate from 15-trial sliding windows, at the two preferred directions. Dashed lines denote ±1 standard deviation around the window estimates, obtained by Bayesian bootstrapping. **(D)** Model comparison for all 74 neurons in the V1 dataset. In these models, 4 are dynamic and the remaining 3 are static, with different noise distributions (Poisson vs CMP) and bases. The training and test log-likelihood ratios (bits/spike) with respect to a homogeneous static Poisson model are shown for all neurons in grey lines. The solid orange lines and numbers denote the medians, and the dashed lines show the first and third quartiles.

the location and firing within the place field vary over the course of the recording.

Compared to the data from V1, neural responses of place cells in hippocampus tend to be sparser and more selective. Many place cells also tend to be direction tuned – spiking

14

only when the animal is running in one direction down the track but not the other. We, thus, fit the data using a dynamic CMP model with 12 equally-spaced knots for $\boldsymbol{X}_t$ with a circular representation of position, and let $\boldsymbol{G}_t = 1$.

For this example neuron, the dynamic CMP model accurately tracks the time-varying place field (Fig. 5B). We then evaluate the fitted Fano factors at the peaks of the two place fields (Fig. 5C). Compared to example from the V1 data, the spiking of this example place cell is much more highly dispersed. The Fano factors vary over time and are also specific in both position and running direction. We then compare model performance on 78 neurons from this recording (19 neurons were excluded due to sparse spiking patterns). In these data, the dynamic models are generally better than the static models (Fig. 5D). Within groups of dynamic or static models, CMP-based models are consistently better than the Poisson-based models.

3.2.3 APN Neuron

To illustrate a case where the dynamic CMP provides a qualitatively better description of neural activity compared to previous models we show results from one neuron recorded from the Allen Institute Visual Coding Neuropixels dataset. See detailed data description in Siegle et al. (2021). Here, when examining tuning to running speed we found a neuron in APN, whose responses were speed tuned - increasing firing with increasing running speed (Fig. 6A), but also highly under-dispersed relative to a Poisson distribution (Fano factor < 1). We analyzed spike counts in 200ms bins for the 160 min recording (ecephys_session_id 719161530, unit_id 950917034), and fit the spiking activity for the whole session using the dynamic CMP model. Here we use a nonlinear function of the running speed v_t for the $\boldsymbol{X}_t = [1\ f(v_t)]$ ($f(v) = v/(1 + 0.1v)$) and $\boldsymbol{G}_t = [1\ v_t]$.

Evaluating all of the data and averaging over time, the neuron is significantly under-dispersed (Fano factors less than 1), and both the mean response and dispersion appear
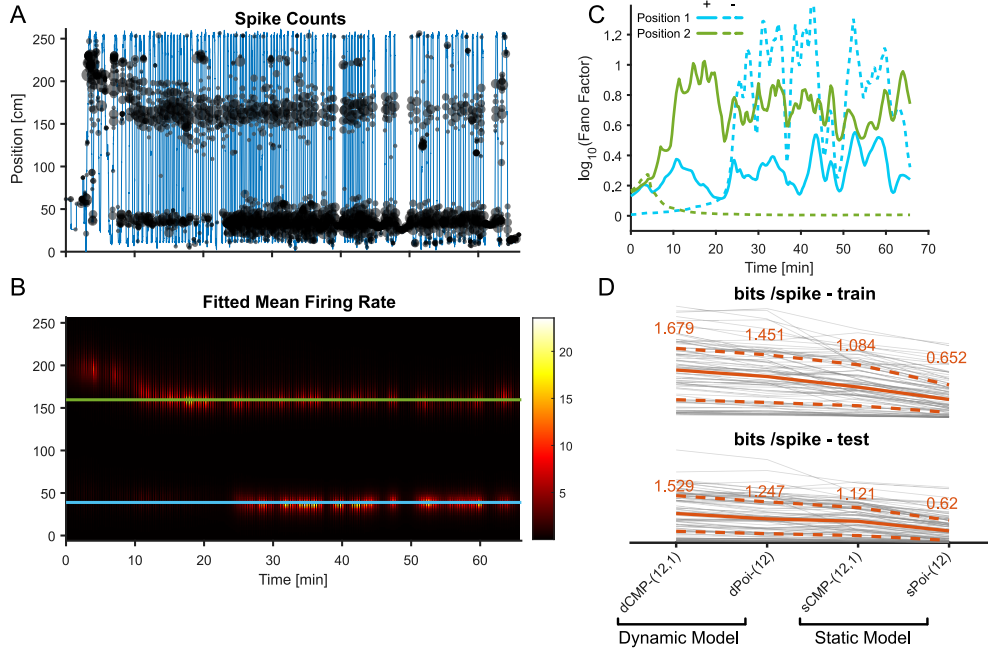
Figure 5: **Modeling nonstationarity in hippocampal spiking activity.** (**A**) Spike counts of an example neuron from the hippocampus recorded while a rat was running back-and-forth on a linear maze. The blue lines show the animal's position over time. The black circles denote spike counts with the radius of each circle corresponding to number of spikes. (**B**) The heatmap shows fitted mean firing rate for the dynamic CMP model dCMP-(12,1). The colored lines show peaks for two place fields, chosen based on the model fit. (**C**) The estimated Fano factors at the two place field peaks, with each running direction (+ vs. -) shown separately. (**D**) The training and test log-likelihood ratios (bits/spike) with respect to a homogeneous static Poisson model for 78 neurons (gray lines). Here the test set log-likelihood ratios are calculated using 5% of the data held-out in a speckled pattern. The red solid lines and numbers denote median values, while the dashed lines show the first and third quartiles.

tuned to running speed (Fig. 6B). However, we found that this neuron's speed tuning is somewhat nonstationary, with the baseline firing rate shifting over time (Fig. 6C). Within individual segments of the recording the Fano factor is much lower than the Fano factor evaluated for the entire recording. Moreover, the Fano factor within each chunk doesn't show a strong relationship to running speed (Fig. 6D). This suggests that the apparent relationship between running speed and the overall Fano factor is a byproduct of the nonstationarity. Using the static CMP model may be able to describe the underdispersion of these responses, but would miss this key feature of the data by assuming that the tuning curve is static. The dynamic Poisson model, on the other

16

hand, is able to describe the nonstationary in the tuning curve, but cannot describe the underdispersion of these data, since the Poisson model always assumes Fano factor = 1.
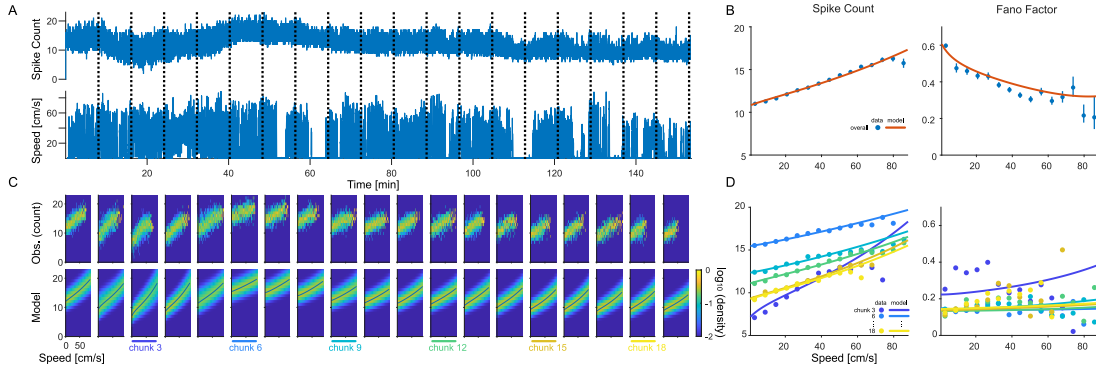

Figure 6: **Modeling a speed-tuned neuron in mouse anterior pretectal nucleus with the dynamic CMP model** Here we analyze a 160 min recording of a neuron from APN, with spike counts calculated in 200ms bins. The whole session was evenly partitioned into chunks for visualization. **(A)** Spike counts and mouse running speed for the recording session. **(B)** The spiking activity and Fano factor as a function of speed for the whole session. Dots denote results from observed spike counts. Errorbars denote 95% confidence intervals from Bayesian bootstrapping. Red lines denote fits from the dynamic CMP model averaged over time. **(C)** The top row shows 2D histograms for spiking counts vs. running speed for each chunk. The second row shows corresponding model fitted densities (averages within each chunk). The lines show predicted average spike counts (tuning curves) for each chunk. **(D)** There is substantial variation in the tuning curves across chunks, and the Fano factor tends to be smaller on individual chunks than in the overall recording.

4 Discussion

Here we introduced a dynamic CMP model to track changes in both the mean and dispersion of neural spikes over time. A global Laplace approximation with a smoothing-based initialization can provide accurate and computationally efficient model estimates. In both simulations and applications with experimental data we find that this model outperforms previous static and dynamic Poisson models, and may, thus, be a useful tool for understanding the role of variability in neural systems. While many studies have characterized noise (DeWeese et al., 2003; Deweese and Zador, 2004; Taouali et al., 2016) or non-stationarity (Tomko and Crapper, 1974; Wu and Hatsopoulos, 2008) sep-

arately, modeling changes in the mean-variance relationship directly may allow us to more closely examine the role of variability in the brain.

The original motivation for the development the CMP distribution is from industrial research on queuing theory where production resources are allocated in a state-dependent way to meet demand (Conway and Maxwell, 1962). Although spiking neurons may not be exactly analogous to production queues, they are both multi-input, single-output, resource-limited nonlinear systems. Previous studies have proposed that resources in neurons may be dynamically allocated across dendrites (Acharya et al., 2022) and that neural networks may share some features with queued, packet-switching systems (Luczak et al., 2015).

The extent to which the dynamic CMP model can predict neural responses more accurately than the dynamic Poisson model or static non-Poisson models depends on the neural activity itself. Here with the V1 data we found a 6% improvement in test log-likelihood ratios between the dynamic and static CMP, while in the HC data there is a 35% improvement. The extent of spontaneous changes in neural responses is somewhat unclear, some evidence suggests that neurons can be relatively stable in some circumstances (Chestek et al., 2007; Stevenson et al., 2011; Dickey et al., 2009), but variability appears to differ across brain areas (Mochizuki et al., 2016). More accurate spike sorting may account for some degree of instability (Steinmetz et al., 2021), and the degree of spontaneous changes may also depends on the brain area (Rule et al., 2019). However, neurons do clearly change both their average responses and dispersion in many situations.

Although the current model works well for fitting neural spikes, there are some potential improvements. First, the state transition matrix F is currently assumed to be known and is fixed to I for convenience in our simulations and experimental analysis. This doesn't allow for interactions between state vectors and may limit the usage in some situations. However, when using a Laplace approximation for the state vectors, F can

be estimated using the EM algorithm as in Macke et al. (2011). Secondly, although the CMP distribution can flexibly model over- and under-dispersed data, the assumed mean-variance relationship may not be appropriate in some cases. To more flexibly model the dispersion for single observations, it may be useful to instead consider the generalized count (GC) distribution (del Castillo and Pérez-Casany, 2005), which includes the CMP distribution as a special case. This model has been applied in the context of linear dynamical systems, similar to the dynamical factor analysis model, with a fixed dispersion function (Gao et al., 2015). To track fluctuations in dispersion more flexibly, it could be useful to allow the function to vary dynamically similar to ν_t in the dynamic CMP model here.

The best modeling strategy also likely depends on what researchers want to know about the variability. Omitted variables (Stevenson, 2018; Goris et al., 2014) and history effects (Uzzell and Chichilnisky, 2004) can alter apparent observation noise. For example, in the hippocampus, place cell firing is highly variable on different passes through the field (Fenton and Muller, 1998). This may be partially due to joint selectivity to position, speed, and head direction, as well as the influence of local field potentials. Here, rather than model these distinct covariates assuming Poisson observations, we allow the variability to be non-Poisson and introduce a dynamic GLM with CMP observations. However, doubly stochastic Poisson models (Barbieri et al., 2001) or latent variable models with fixed mean-variance relationships (Gao et al., 2015; Pillow and Scott, 2012) may also be able to account for some differences in the variance over time. Nonetheless, the dynamic CMP model may provide a useful tool for neuroscientists to study the role of variance directly. Since the static CMP model can improve decoding of external variables in some cases (Ghanbari et al., 2019), the dynamic CMP may lead to further improvements in decoding by tracking nonstationarity in neural response properties.

Acknowledgments

Appendices

A   Quantifying Uncertainties

After convergence, we have an approximation of the log-posterior $P(\boldsymbol{\theta}_t|\boldsymbol{Y}) \approx N(\boldsymbol{\theta}_t|\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$, and we can use this approximation to quantify the uncertainty about the CMP parameters, as well as about the mean rate at each time.

The CMP parameters are log-normal distributed. Let $\boldsymbol{Z}_{it} = \begin{pmatrix} \boldsymbol{x}'_{it} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{g}'_{it} \end{pmatrix}$, then $(\lambda_{it}, \nu_{it})' = \exp(\boldsymbol{Z}_{it}\boldsymbol{\theta}_t) \sim \text{Lognormal}_2(\boldsymbol{Z}_{it}\boldsymbol{\mu}_t, \boldsymbol{Z}_{it}\boldsymbol{\Sigma}_t\boldsymbol{Z}'_{it})$. Denote the variance of CMP parameters as $\boldsymbol{V}_{it}$, where for $\boldsymbol{a} = \boldsymbol{Z}\boldsymbol{\mu}$ and $\boldsymbol{S} = \boldsymbol{Z}\boldsymbol{\Sigma}\boldsymbol{Z}'$, $[V]_{mn} = e^{a_m + a_n + \frac{1}{2}(S_{mm} + S_{nn})}(e^{S_{mn}} - 1)$

The conditional mean firing rate is $\delta_{it} = E(Y_{it})$, whose variance can be calculated by the Delta method:

$$
\begin{aligned}
\widehat{Var}(\delta_{it}) &= \begin{pmatrix} \frac{\partial \delta_{it}}{\partial \lambda_{it}} & \frac{\partial \delta_{it}}{\partial \nu_{it}} \end{pmatrix} \boldsymbol{V}_{it} \begin{pmatrix} \frac{\partial \delta_{it}}{\partial \lambda_{it}} \\ \frac{\partial \delta_{it}}{\partial \nu_{it}} \end{pmatrix}, \\
\frac{\partial \delta_{it}}{\partial \lambda_{it}} &= \frac{\partial^2 \log Z_{it}}{\partial \log \lambda_{it} \partial \lambda_{it}} = \frac{Var(Y_{it})}{\lambda_{it}}, \\
\frac{\partial \delta_{it}}{\partial \nu_{it}} &= \frac{\partial^2 \log Z_{it}}{\partial \log \lambda_{it} \partial \nu_{it}} = -Cov(Y_{it}, \log Y_{it}!).
\end{aligned}
\tag{6}
$$

We can calculate the moments as in Appendix B, or we can use simpler approximations $E(Y) = \lambda^{1/\nu} - \frac{\nu - 1}{2\nu}$ when $\nu \leq 1$ or $\lambda > 10^\nu$. Then $\frac{\partial \delta_{it}}{\partial \lambda_{it}} \approx \frac{1}{\nu_{it}}\lambda_{it}^{1/\nu_{it} - 1}$ and $\frac{\delta_{it}}{\nu_{it}} \approx$

$$-\frac{\lambda_{it}^{1/\nu_{it}} \log \lambda_{it}}{\nu_{it}^2} - \frac{1}{2\nu_{it}^2}.$$

B   Moments approximation for Conway-Maxwell Poisson distribution

To estimate the state-vector for the dynamic CMP model, we need to find first and second moments for Y and $\log Y!$. For $Y \sim CMP(\lambda, \nu)$,

$$
\begin{aligned}
Z(\lambda, \nu) &= \sum_{k=0}^{\infty} \frac{\lambda^k}{(k!)^\nu}, \\
E(Y) &= \frac{\partial \log Z}{\partial \log \lambda} = \frac{1}{Z} \sum_{k=0}^{\infty} \frac{k\lambda^k}{(k!)^\nu}, \\
Var(Y) &= \frac{\partial^2 \log Z}{\partial (\log \lambda)^2} = \frac{1}{Z} \sum_{k=0}^{\infty} \frac{k^2 \lambda^k}{(k!)^\nu} - E^2(Y) \\
E(\log Y!) &= -\frac{\partial \log Z}{\partial \nu} = \frac{1}{Z} \sum_{k=0}^{\infty} \frac{(\log k!)\lambda^k}{(k!)^\nu}, \\
Var(\log Y!) &= \frac{\partial^2 \log Z}{\partial \nu^2} = \frac{1}{Z} \sum_{k=0}^{\infty} \frac{(\log k!)^2 \lambda^k}{(k!)^\nu} - E^2(\log Y!), \\
Cov(Y, \log Y!) &= -\frac{\partial^2 \log Z}{\partial \log \lambda \partial \nu} = \frac{1}{Z} \sum_{k=0}^{\infty} \frac{(\log k!)k\lambda^k}{(k!)^\nu} - E(\log Y!)E(Y).
\end{aligned}
\tag{7}
$$

Generally, these moments can be calculated by truncated summation. However, when $\lambda \geq 2$ and $\nu \leq 1$, we need many steps for accurate approximation. In this case, we make use of a previous asymptotic results (Chatla and Shmueli (2018); Gaunt et al. (2019)) for efficient calculation. Let $\alpha = \lambda^{1/\nu}$, $c_1 = \frac{\nu^2-1}{24}$ and $c_2 = \frac{\nu^2-1}{48} + \frac{c_1^2}{2}$,

$$Z(\lambda, \nu) = \frac{e^{\nu\alpha}}{\lambda^{\frac{\nu-1}{2\nu}}(2\pi)^{\frac{\nu-1}{2}}\sqrt{\nu}}(1 + c_1(\nu\alpha)^{-1} + c_2(\nu\alpha)^{-2} + \mathcal{O}(\lambda^{-3/\nu})). \tag{8}$$

Then the moments can be calculated as follows:

$$E(Y) = \alpha - \frac{\nu - 1}{2\nu} - \frac{\nu^2 - 1}{24\nu^2}\alpha^{-1} - \frac{\nu^2 - 1}{24\nu^3}\alpha^{-2} + \mathcal{O}(\alpha^{-3}),$$

$$Var(Y) = \frac{\alpha}{\nu} + \frac{\nu^2 - 1}{24\nu^3}\alpha^{-1} + \frac{\nu^2 - 1}{12\nu^4}\aleph^{-2} + \mathcal{O}(\alpha^{-3}),$$

$$E(\log Y!) = \alpha\left(\frac{\log \lambda}{\nu} - 1\right) + \frac{\log \lambda}{2\nu^2} + \frac{1}{2\nu} + \frac{\log 2\pi}{2}$$
$$- \frac{\alpha^{-1}}{24}\left(1 + \frac{1}{\nu^2} + \frac{\log \lambda}{\nu} - \frac{\log \lambda}{\nu^3}\right)$$
$$- \frac{\alpha^{-2}}{24}\left(\frac{1}{\nu^3} + \frac{\log \lambda}{\nu^2} - \frac{\log \lambda}{\nu^4}\right) + \mathcal{O}(\alpha^{-3}),$$

$$Var(\log Y!) = \frac{\alpha(\log \lambda)^2}{\nu^3} + \frac{\log \lambda}{\nu^3} + \frac{1}{2\nu^2} \tag{9}$$
$$+ \frac{\alpha^{-1}}{24\nu^5}[-2\nu^2 + 4\nu \log \lambda + (-1 + \nu^2)(\log \lambda)^2]$$
$$+ \frac{\alpha^{-2}}{24\nu^6}[-3\nu^2 - 2\nu(-3 + \nu^2)\log \lambda + 2(-1 + \nu^2)(\log \lambda)^2]$$
$$+ \mathcal{O}(\alpha^{-3}),$$

$$Cov(Y, \log Y!) = \frac{\alpha \log \lambda}{\nu^2} + \frac{1}{2\nu^2} + \frac{\alpha^{-1}}{24}\left(\frac{2}{\nu^3} + \frac{\log \lambda}{\nu^2} - \frac{\log \lambda}{\nu^4}\right)$$
$$- \frac{1}{24\alpha^2}\left(\frac{1}{\nu^2} - \frac{3}{\nu^4} - \frac{2\log \lambda}{\nu^3} + \frac{2\log \lambda}{\nu^5}\right) + \mathcal{O}(\alpha^{-3}).$$

C   Gradient and Hessian of the log-posterior

We estimate the state vector by maximizing the log-posterior with Newton-Raphson updates. Denote $f = P(\Theta|Y)$, the $(k + 1)$-th update of NR algorithm is $\Theta^{(k+1)} =$

$\boldsymbol{\Theta}^{(k)} + [\nabla\nabla_{\boldsymbol{\Theta}^{(k)}} f]^{-1} \nabla_{\boldsymbol{\Theta}^{(k)}} f$. The gradient is:

$$
\begin{aligned}
\nabla_{\boldsymbol{\Theta}} f &= \left[\left(\frac{\partial f}{\partial\boldsymbol{\theta}_1}\right)', \ldots, \left(\frac{\partial f}{\partial\boldsymbol{\theta}_T}\right)'\right]', \\
\frac{\partial f}{\partial\boldsymbol{\theta}_1} &= \frac{\partial l_1}{\partial\boldsymbol{\theta}_1} - \boldsymbol{Q}_0^{-1}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0) + \boldsymbol{F}'\boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2 - \boldsymbol{F}\boldsymbol{\theta}_1), \\
\frac{\partial f}{\partial\boldsymbol{\theta}_t} &= \frac{\partial l_t}{\partial\boldsymbol{\theta}_t} - \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_t - \boldsymbol{F}\boldsymbol{\theta}_{t-1}) + \boldsymbol{F}'\boldsymbol{Q}^{-1}(\boldsymbol{\theta}_{t+1} - \boldsymbol{F}\boldsymbol{\theta}_t), \\
\frac{\partial f}{\partial\boldsymbol{\theta}_T} &= \frac{\partial l_T}{\partial\boldsymbol{\theta}_T} - \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_T - \boldsymbol{F}\boldsymbol{\theta}_{T-1}), \\
\frac{\partial l_t}{\partial\boldsymbol{\theta}_t} &= \sum_{i=1}^{n} \begin{pmatrix} (y_{it} - E(Y_{it}))\,\boldsymbol{x}_{it} \\ \nu_{it}(E(\log Y_{it}!) - \log y_{it}!)\,\boldsymbol{g}_{it} \end{pmatrix}.
\end{aligned}
\tag{10}
$$

The Hessian matrix is:

$$
\nabla\nabla_{\boldsymbol{\Theta}} f = \begin{pmatrix}
\frac{\partial^2 f}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'} & \boldsymbol{F}'\boldsymbol{Q}^{-1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{Q}^{-1}\boldsymbol{F} & \frac{\partial^2 f}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2'} & \boldsymbol{F}'\boldsymbol{Q}^{-1} & \cdots & \vdots \\
\boldsymbol{0} & \boldsymbol{Q}^{-1}\boldsymbol{F} & \frac{\partial^2 f}{\partial\boldsymbol{\theta}_3\partial\boldsymbol{\theta}_3'} & \cdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\boldsymbol{0} & \cdots & \cdots & \cdots & \frac{\partial^2 f}{\partial\boldsymbol{\theta}_T\partial\boldsymbol{\theta}_T'}
\end{pmatrix},
\tag{11}
$$

$$
\begin{aligned}
\frac{\partial^2 f}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'} &= \frac{\partial^2 l_1}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'} - \boldsymbol{Q}_0^{-1} - \boldsymbol{F}'\boldsymbol{Q}^{-1}\boldsymbol{F}, \\
\frac{\partial^2 f}{\partial\boldsymbol{\theta}_t\partial\boldsymbol{\theta}_t'} &= \frac{\partial^2 l_t}{\partial\boldsymbol{\theta}_t\partial\boldsymbol{\theta}_t'} - \boldsymbol{Q}^{-1} - \boldsymbol{F}'\boldsymbol{Q}^{-1}\boldsymbol{F}, \\
\frac{\partial^2 f}{\partial\boldsymbol{\theta}_T\partial\boldsymbol{\theta}_T'} &= \frac{\partial^2 l_T}{\partial\boldsymbol{\theta}_T\partial\boldsymbol{\theta}_T'} - \boldsymbol{Q}^{-1},
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial^2 l_t}{\partial\boldsymbol{\theta}_t\partial\boldsymbol{\theta}_t'} &= \sum_{i=1}^{n} \begin{pmatrix} \boldsymbol{A}_{it} & \boldsymbol{B}_{it} \\ \boldsymbol{B}_{it}' & \boldsymbol{C}_{it} \end{pmatrix}, \\
\boldsymbol{A}_{it} &= Var(Y_{it})\boldsymbol{x}_{it}\boldsymbol{x}_{it}', \\
\boldsymbol{B}_{it} &= -\nu_{it}Cov(Y_{it}, \log Y_{it}!)\boldsymbol{x}_{it}\boldsymbol{g}_{it}', \\
\boldsymbol{C}_{it} &= \nu_{it}[\nu_{it}Var(\log Y_{it}) - E(\log Y_{it}!) + \log y_{it}!]\boldsymbol{g}_{it}\boldsymbol{g}_{it}'.
\end{aligned}
\tag{12}
$$

When $\log y_{it}! \ll E(\log Y_{it}!)$, the Hessian may be ill-conditioned or even positive-definite. To ensure the robustness, do Fisher scoring, i.e. replace the observed information $-\nabla\nabla_{\Theta} f$ by the expected information $E(-\nabla\nabla_{\Theta} f)$, so that $\boldsymbol{C}_{it} = \nu_{it}^2 Var(\log Y_{it}!)\boldsymbol{g}_{it}\boldsymbol{g}_{it}'$.

The observed information may provide more accurate estimates of the variance at the MAP estimate (Efron and Hinkley, 1978), but these estimates may not be robust. Robustness to outliers is especially necessary for convergence during optimization of CMP models (Green, 1984; Chatla and Shmueli, 2018).

References

Acharya, J., Basu, A., Legenstein, R., Limbacher, T., Poirazi, P., and Wu, X. (2022). Dendritic Computing: Branching Deeper into Machine Learning. *Neuroscience*, 489:275–289.

Amarasingham, A., Chen, T. L., Geman, S., Harrison, M. T., and Sheinberg, D. L. (2006). Spike Count Reliability and the Poisson Hypothesis. *Journal of Neuroscience*, 26(3):801–809.

Barbieri, R., Quirk, M. C., Frank, L. M., Wilson, M. A., and Brown, E. N. (2001). Construction and analysis of non-Poisson stimulus-response models of neural spiking activity. *Journal of Neuroscience Methods*, 105(1):25–37.

Brown, E. N., Nguyen, D. P., Frank, L. M., Wilson, M. A., and Solo, V. (2001). An analysis of neural receptive field plasticity by point process adaptive filtering. *Proceedings of the National Academy of Sciences*, 98(21):12261–12266.

Chatla, S. B. and Shmueli, G. (2018). Efficient estimation of COM–Poisson regression and a generalized additive model. *Computational Statistics and Data Analysis*, 121:71–88.

Chestek, C. A., Batista, A. P., Santhanam, G., Yu, B. M., Afshar, A., Cunningham,

J. P., Gilja, V., Ryu, S. I., Churchland, M. M., and Shenoy, K. V. (2007). Single-neuron stability during repeated reaching in macaque premotor cortex. *Journal of Neuroscience*, 27(40):10742–10750.

Churchland, A. K., Kiani, R., Chaudhuri, R., Wang, X. J., Pouget, A., and Shadlen, M. N. (2011). Variance as a Signature of Neural Computations during Decision Making. *Neuron*, 69(4):818–831.

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., Newsome, W. T., Clark, A. M., Hosseini, P., Scott, B. B., Bradley, D. C., Smith, M. a., Kohn, A., Movshon, J. A., Armstrong, K. M., Moore, T., Chang, S. W., Snyder, L. H., Lisberger, S. G., Priebe, N. J., Finn, I. M., Ferster, D., Ryu, S. I., Santhanam, G., Sahani, M., and Shenoy, K. V. (2010). Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nature Neuroscience*, 13(3):369–378.

Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G., and Shenoy, K. V. (2006). Neural Variability in Premotor Cortex Provides a Signature of Motor Preparation. *Journal of Neuroscience*, 26(14):3697.

Conway, R. W. and Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, 12(2):132–136.

Czanner, G., Eden, U. T., Wirth, S., Yanike, M., Suzuki, W. A., and Brown, E. N. (2008). Analysis of between-trial and within-trial neural spiking dynamics. *Journal of Neurophysiology*, 99(5):2672–2693.

del Castillo, J. and Pérez-Casany, M. (2005). Overdispersed and underdispersed Poisson generalizations. *Journal of Statistical Planning and Inference*, 134(2):486–500.

DeWeese, M. and Zador, A. (1998). Asymmetric Dynamics in Optimal Variance Adaptation. *Neural Computation*, 10(5):1179–1202.

DeWeese, M. R., Wehr, M., and Zador, A. M. (2003). Binary spiking in auditory cortex.

The Journal of neuroscience : the official journal of the Society for Neuroscience, 23(21):7940–7949.

Deweese, M. R. and Zador, A. M. (2004). Shared and private variability in the auditory cortex. *Journal of neurophysiology*, 92(3):1840–1855.

Dickey, A. S., Suminski, A., Amit, Y., and Hatsopoulos, N. G. (2009). Single-unit stability using chronically implanted multielectrode arrays. *Journal of Neurophysiology*, 102(2):1331–1339.

Dragoi, V., Sharma, J., and Sur, M. (2000). Adaptation-Induced Plasticity of Orientation Tuning in Adult Visual Cortex. *Neuron*, 28(1):287–298.

Eden, U. T., Frank, L. M., Barbieri, R., Solo, V., and Brown, E. N. (2004). Dynamic Analysis of Neural Encoding by Point Process Adaptive Filtering. *Neural Computation*, 16(5):971–998.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 65(3):457–483.

Fenton, A. A. and Muller, R. U. (1998). Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proceedings of the National Academy of Sciences of the United States of America*, 95(6):3182–3187.

Gao, Y., Buesing, L., Shenoy, K. V., and Cunningham, J. P. (2015). High-dimensional neural spike train analysis with generalized count linear dynamical systems. Technical report.

Gaunt, R. E., Iyengar, S., Olde Daalhuis, A. B., and Simsek, B. (2019). An asymptotic expansion for the normalizing constant of the Conway-Maxwell-Poisson distribution. *Ann Inst Stat Math*, 71:163–180.

Ghanbari, A., Lee, C. M., Read, H. L., and Stevenson, I. H. (2019). Modeling stimulus-

dependent variability improves decoding of population neural responses. *Journal of Neural Engineering*, 16(6).

Goris, R. L., Movshon, J. A., and Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience 2014 17:6*, 17(6):858–865.

Green, P. J. (1984). Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(2):149–192.

Gupta, R. C., Sim, S. Z., and Ong, S. H. (2014). Analysis of discrete data by Conway–Maxwell Poisson distribution. *AStA Advances in Statistical Analysis*, 98(4):327–343.

Kara, P., Reinagel, P., and Reid, R. C. (2000). Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron*, 27(3):635–646.

Kelly, R. C., Smith, M. A., Kass, R. E., and Lee, T. S. (2010). Local field potentials indicate network state and account for neuronal response variability. *Journal of Computational Neuroscience*, 29(3):567–579.

Kohn, A. and Smith, M. (2016). Utah array extracellular recordings of spontaneous and visually evoked activity from anesthetized macaque primary visual cortex (V1).

Lesica, N. A., Jin, J., Weng, C., Yeh, C. I., Butts, D. A., Stanley, G. B., and Alonso, J. M. (2007). Adaptation to Stimulus Contrast and Correlations during Natural Visual Stimulation. *Neuron*, 55(3):479–491.

Luczak, A., McNaughton, B. L., and Harris, K. D. (2015). Packet-based communication in the cortex. *Nature Reviews Neuroscience*, 16(12):745–755.

Macke, J. H., Buesing, L., Cunningham, J. P., Yu, B. M., Shenoy, K. V., and Sahani, M. (2011). Empirical models of spiking in neural populations. *Advances in Neural Information Processing Systems*, 24.

Maimon, G. and Assad, J. a. (2009). Beyond Poisson: Increased Spike-Time Regularity across Primate Parietal Cortex. *Neuron*, 62(3):426–440.

Mizuseki, K., Diba, K., Pastalkova, E., Teeters, J., Sirota, A., and Buzsáki, G. (2014). Neurosharing: large-scale data sets (spike, LFP) recorded from the hippocampal-entorhinal system in behaving rats. *F1000Research 2014 3:98*, 3:98.

Mizuseki, K., Sirota, A., Pastalkova, E., Diba, K., and Buzsáki, G. (2013). Multiple single unit recordings from different rat hippocampal and entorhinal regions while the animals were performing multiple behavioral tasks. *CRCNS.org*.

Mochizuki, Y., Onaga, T., Shimazaki, H., Shimokawa, T., Tsubo, Y., Kimura, R., Saiki, A., Sakai, Y., Isomura, Y., Fujisawa, S., Shibata, K. I., Hirai, D., Furuta, T., Kaneko, T., Takahashi, S., Nakazono, T., Ishino, S., Sakurai, Y., Kitsukawa, T., Lee, J. W., Lee, H., Jung, M. W., Babul, C., Maldonado, P. E., Takahashi, K., Arce-McShane, F. I., Ross, C. F., Sessle, B. J., Hatsopoulos, N. G., Brochier, T., Riehle, A., Chorley, P., Grün, S., Nishijo, H., Ichihara-Takeda, S., Funahashi, S., Shima, K., Mushiake, H., Yamane, Y., Tamura, H., Fujita, I., Inaba, N., Kawano, K., Kurkin, S., Fukushima, K., Kurata, K., Taira, M., Tsutsui, K. I., Ogawa, T., Komatsu, H., Koida, K., Toyama, K., Richmond, B. J., and Shinomoto, S. (2016). Similarity in Neuronal Firing Regimes across Mammalian Species. *Journal of Neuroscience*, 36(21):5736–5747.

Paninski, L., Ahmadian, Y., Ferreira, D. G., Koyama, S., Rahnama Rad, K., Vidne, M., Vogelstein, J., and Wu, W. (2010). A new look at state-space models for neural data. *J. Comput. Neurosci.*, 29(1–2):107–126.

Pillow, J. W. and Scott, J. G. (2012). Fully Bayesian inference for neural models with negative-binomial spiking. Technical report.

Rauch, H. E., Tung, F., and Striebel, C. T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445–1450.

Rokni, U., Richardson, A. G., Bizzi, E., and Seung, H. S. (2007). Motor learning with unstable neural representations. *Neuron*, 54(4):653–666.

Rossant, C., Kadir, S. N., Goodman, D. F., Schulman, J., Hunter, M. L., Saleem, A. B., Grosmark, A., Belluscio, M., Denfield, G. H., Ecker, A. S., Tolias, A. S., Solomon, S., Buzski, G., Carandini, M., and Harris, K. D. (2016). Spike sorting for large, dense electrode arrays. *Nature Neuroscience 2016 19:4*, 19(4):634–641.

Rule, M. E., O'Leary, T., and Harvey, C. D. (2019). Causes and consequences of representational drift. *Current Opinion in Neurobiology*, 58:141–147.

Sellers, K. F. and Shmueli, G. (2010). A flexible regression model for count data. *The Annals of Applied Statistics*, 4(2):943 – 961.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):127–142.

Shoham, S., Fellows, M. R., and Normann, R. A. (2003). Robust, automatic spike sorting using mixtures of multivariate t-distributions. *Journal of neuroscience methods*, 127(2):111–122.

Siegle, J. H., Jia, X., Durand, S., et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature 2021 592:7852*, 592(7852):86–92.

Smith, A. C. and Brown, E. N. (2003). Estimating a state-space model from point process observations. *Neural Computation*, 15(5):965–991.

Smith, M. A. and Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *Journal of Neuroscience*, 28(48):12591–12603.

Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., Chen, S., Colonell, J., Gardner, R. J., Karsh, B.,

Kloosterman, F., Kostadinov, D., Mora-Lopez, C., O'Callaghan, J., Park, J., Putzeys, J., Sauerbrei, B., van Daal, R. J., Vollan, A. Z., Wang, S., Welkenhuysen, M., Ye, Z., Dudman, J. T., Dutta, B., Hantman, A. W., Harris, K. D., Lee, A. K., Moser, E. I., O'Keefe, J., Renart, A., Svoboda, K., Häusser, M., Haesler, S., Carandini, M., and Harris, T. D. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*, 372(6539).

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

Stevenson, I. H. (2018). Omitted Variable Bias in GLMs of Neural Spiking Activity. *Neural Computation*, 30(12):3227–3258.

Stevenson, I. H., Cherian, A., London, B. M., Sachs, N. A., Lindberg, E., Reimer, J., Slutzky, M. W., Hatsopoulos, N. G., Miller, L. E., and Kording, K. P. (2011). Statistical assessment of the stability of neural movement representations. *Journal of Neurophysiology*, 106(2):764–774.

Taouali, W., Benvenuti, G., Wallisch, P., Chavane, F., and Perrinet, L. U. (2016). Testing the odds of inherent vs. observed overdispersion in neural spike counts. *Journal of neurophysiology*, 115(1):434–44.

Tomko, G. J. and Crapper, D. R. (1974). Neuronal variability: non-stationary responses to identical visual stimuli. *Brain Research*, 79(3):405–418.

Uzzell, V. J. and Chichilnisky, E. J. (2004). Precision of spike trains in primate retinal ganglion cells. *Journal of Neurophysiology*, 92(2):780–789.

Wei, G. and Stevenson, I. H. (2021). Tracking Fast and Slow Changes in Synaptic Weights From Simultaneously Observed Pre- and Postsynaptic Spiking. *Neural computation*, 33(10):2682–2709.

Wu, W. and Hatsopoulos, N. G. (2008). Real-time decoding of nonstationary neu-

ral activity in motor cortex. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 16(3):213–222.