**Dynamic modeling of spike count data with non-Poisson variability**

Ganchao Wei1\* and Ian H. Stevenson2-4

*1 Department of Statistics, University of Connecticut*

*2 Department of Psychological Sciences, University of Connecticut*

*3 Department of Biomedical Engineering, University of Connecticut*

*4 Connecticut Institute for Brain and Cognitive Science, University of Connecticut*

*\* Corresponding Author:* ganchao.wei@uconn.edu

# Abstract

# Introduction

Drift…

(Rokni et al. 2007)

(Chestek et al. 2007)

(Stevenson et al. 2011)

(Tomko and Crapper 1974)

Stability…

(Dickey et al. 2009)

(Steinmetz et al. 2021) [shows that correcting for electrode drift reduces some instability]

[some amount of drift is functional] adaptation/plasticity…

(Lesica et al. 2007)

Variability…

(Ghanbari et al. 2019)

(Stevenson 2016)

(Fenton and Muller 1998)

(Barbieri et al. 2001)

(Maimon and Assad 2009)

(Churchland et al. 2010)

(Churchland et al. 2011)

(Eden et al. 2004)

(Brown et al. 2001)

(DeWeese and Zador 1998)

# Methods

## Dynamic Conway-Maxwell Poisson Model

Based on Friday’s talk, it seems you may be uncomfortable to let . However, I prefer to keep writing things in this general form. It’s always used in statistics textbook & Kalman filter/ DLM in wiki. Moreover, even in Eden’s “adaptive filtering” paper, the derivation is conducted for ensemble neurons in appendix. The benefit from recursive prior in DGLM is significant for small , but not confined to only. A more important reason: it suggests that we can use the model to analyze the neural population. Although it’s trivial in math, it makes things a bit more interesting for scientific application.

In V1 application, the reason I accept to do single neuron observation is that it takes a long time to conduct a trial, and assuming within-trial shift of state vector is appropriate. However, if the experiment for different stimuli is conducted quickly or even (impossibly) simultaneously, we may stack the observations within trial together. Otherwise, we lose information/ efficiency.

If the above arguments still cannot persuade you, I’m totally fine to write things in , since the extension to is obvious.

Denote the spike count of neuron at time bin as , for The spikes follow Conway-Maxwell Poisson (CMP) distributions component-wise, with parameters and . The probability mass function (pmf) of CMP is

, where is the normalizing constant. The parameter controls different dispersion patterns, i.e. equi- (), over- () or under-dispersion (), linking to three common distributions as special cases: (1) the Poisson (), (2) geometric () and (3) Bernoulli ().

The CMP parameters at are modeled by two log-linear models, and , with and .The state vector is assumed to progress linearly with a Gaussian noise, .

In summary, the neurons follow CMP distributions independently, conditioning on state vector .

The state vector evolves linearly:

In this paper, we set for convenience.

## Estimate the model by Gaussian approximation

In this section, we assume **F** and **Q** are known. Since the spikes are CMP distributed, we cannot estimate in closed form, with . Here we approximate it by a multivariate Gaussian distribution, , with . The parameters of this Gaussian found by a global Laplace approximation, i.e., and . The log-posterior is given by:

, where is the log-likelihood. The log-posterior is concave [cite CMP paper], and the Markovian structure of the state vector dynamics makes it possible to optimize by Newton-Raphson (NR) in time [cite Liam et al.]. After the Newton update, we can further quantify the uncertainties for CMP parameters and mean firing rate. See details in the appendix.

However, there are several issues for the Newton update. Firstly, to get gradient and hessian we need to calculate moments of and , which have no closed forms and are generally calculated by truncated summation. When and , we will need a large step size for accurate approximation, and this makes update cumbersome. However, we can approximate them in closed forms by some asymptotic results [cite approximations]. The details of the moment approximation can be found in the appendix.

Secondly, the hessian is not robust to outliers, which is discussed in detail in the appendix. After using the Fisher scoring, i.e., replacing the observed information by expected information, we can ensure the robustness by sacrificing the efficiency a bit.

Thirdly, the Newton update may take a long time for bad initials, especially when is large. To resolve that, we use the smoothing estimate with local Gaussian approximation as a warm start. The forward filtering stage for Poisson case is derived in the appendix of [cite Eden et al.]. Although doing smoothing is fast, the estimations will be biased too much for large dynamics. In the forward filtering stage, the Gaussian approximation at each step is conducted locally at the recursive prior . This will be statistically inefficient when recursive prior is too far away from the posterior mode, or when there is a huge “jump” in the state vector. Moreover, using Fisher scoring for robustness makes the algorithm even less efficient. In this case, it is essential to do global Laplace approximation by NR.

## Robust and quick estimation

When is small, different values will influence the estimation a lot. One possible way to estimate the is to use the EM algorithm. However, using the Laplace approximation for at E-step breaks the usual guarantee of non-decreasing likelihoods in EM, and hence may lead to divergence. This problem is even exacerbated for CMP distribution. To avoid that, we can sample posteriors directly by MCMC. However, non-closed CMP moments makes sampling cumbersome. To estimate the robustly and quickly, we assume is diagonal and estimate it by maximizing the prediction likelihood in the filtering stage, as in [cite our smooth-GBLM].

# Results

## Figure 1

Code: [figure1\_singleNu\_dirShift.m](https://github.com/weigcdsb/COM_POISSON/blob/main/demo/figure1/figure1_singleNu_dirShift.m)



Simulation settings: 100 directions from 0 to 1 rad. 100 trials. is modeled by with 10 knots, , i.e., over- to under- dispersion.



## Simulation

TBD…

Q optimization (Do we need this? It has been illustrated in SMOOTH-GBLM, and it’s shouldn’t be the focus of this paper)

|  |  |  |  |
| --- | --- | --- | --- |
| Heatmap | Q\_lam  Y = estimated log10(Q\_lam)  X = true log10(Q\_lam)  Color: yellow = small Q\_nu; black = large Q\_nu | Q\_nu  Y = estimated log10(Q\_nu)  X = true log10(Q\_nu)  Color: purple = small Q\_lam; black = large Q\_lam | fitting |
| Chart  Description automatically generated | Chart, scatter chart  Description automatically generated | Chart, scatter chart  Description automatically generated | Graphical user interface  Description automatically generated with low confidence |

## Application

V1 data

Code: [v1\_comparison\_nan.m](https://github.com/weigcdsb/COM_POISSON/blob/main/demo/v1/v1_comparison_nan.m); [compare\_all\_na.m](https://github.com/weigcdsb/COM_POISSON/blob/main/demo/v1/compare_all_na.m); [demo\_v1\_pd.m](https://github.com/weigcdsb/COM_POISSON/blob/main/demo/v1/demo_v1_pd.m)



Hippocampus data

Code: [hc\_comparison\_v2.m](https://github.com/weigcdsb/COM_POISSON/blob/main/demo/hc/hc_comparison_v2.m);



# Discussion

Omitted variables can increase the apparent variability of observations via the law of total variance. For example, in the hippocampus, place cell firing is highly variable on different passes through the field (cite Fenton). This may be partially due to joint selectivity to position, speed, and head direction, as well as the influence of theta phase. Here, rather than model these distinct covariates assuming Poisson observations, we allow the variability to be non-Poisson.

# Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1931249.

# References

**Barbieri R**, **Quirk MC**, **Frank LM**, **Wilson MA**, **Brown EN**. Construction and analysis of non-Poisson stimulus-response models of neural spiking activity. *J Neurosci Methods* 105: 25–37, 2001.

**Brown EN**, **Nguyen DP**, **Frank LM**, **Wilson MA**, **Solo V**. An analysis of neural receptive field plasticity by point process adaptive filtering. *Proc Natl Acad Sci* 98: 12261–12266, 2001.

**Chatla SB**, **Shmueli G**. Efficient estimation of COM–Poisson regression and a generalized additive model. *Comput Stat Data Anal* 121: 71–88, 2018.

**Chestek CA**, **Batista AP**, **Santhanam G**, **Yu BM**, **Afshar A**, **Cunningham JP**, **Gilja V**, **Ryu SI**, **Churchland MM**, **Shenoy K V**. Single-neuron stability during repeated reaching in macaque premotor cortex. *J Neurosci* 27: 10742–10750, 2007.

**Churchland AK**, **Kiani R**, **Chaudhuri R**, **Wang XJ**, **Pouget A**, **Shadlen MN**. Variance as a Signature of Neural Computations during Decision Making. *Neuron* 69: 818–831, 2011.

**Churchland MM**, **Yu BM**, **Cunningham JP**, **Sugrue LP**, **Cohen MR**, **Corrado GS**, **Newsome WT**, **Clark AM**, **Hosseini P**, **Scott BB**, **Bradley DC**, **Smith M a**, **Kohn A**, **Movshon JA**, **Armstrong KM**, **Moore T**, **Chang SW**, **Snyder LH**, **Lisberger SG**, **Priebe NJ**, **Finn IM**, **Ferster D**, **Ryu SI**, **Santhanam G**, **Sahani M**, **Shenoy K V**. Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nat Neurosci* 13: 369–378, 2010.

**DeWeese M**, **Zador A**. Asymmetric Dynamics in Optimal Variance Adaptation. *Neural Comput* 10: 1179–1202, 1998.

**Dickey AS**, **Suminski A**, **Amit Y**, **Hatsopoulos NG**. Single-unit stability using chronically implanted multielectrode arrays. *J Neurophysiol* 102: 1331–1339, 2009.

**Eden UT**, **Frank LM**, **Barbieri R**, **Solo V**, **Brown EN**. Dynamic Analysis of Neural Encoding by Point Process Adaptive Filtering. *Neural Comput* 16: 971–998, 2004.

**Fenton AA**, **Muller RU**. Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proc Natl Acad Sci U S A* 95: 3182–3187, 1998.

**Ghanbari A**, **Lee CM**, **Read HL**, **Stevenson IH**. Modeling stimulus-dependent variability improves decoding of population neural responses. *J Neural Eng* 16, 2019.

**Lesica NA**, **Jin J**, **Weng C**, **Yeh CI**, **Butts DA**, **Stanley GB**, **Alonso JM**. Adaptation to Stimulus Contrast and Correlations during Natural Visual Stimulation. *Neuron* 55: 479–491, 2007.

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

**Rauch HE**, **Tung F**, **Striebel CT**. Maximum likelihood estimates of linear dynamic systems. *AIAA J* 3: 1445–1450, 1965.

**Rokni U**, **Richardson AG**, **Bizzi E**, **Seung HS**. Motor learning with unstable neural representations. *Neuron* 54: 653–666, 2007.

**Steinmetz NA**, **Aydin C**, **Lebedeva A**, **Okun M**, **Pachitariu M**, **Bauza M**, **Beau M**, **Bhagat J**, **Böhm C**, **Broux M**, **Chen S**, **Colonell J**, **Gardner RJ**, **Karsh B**, **Kloosterman F**, **Kostadinov D**, **Mora-Lopez C**, **O’Callaghan J**, **Park J**, **Putzeys J**, **Sauerbrei B**, **van Daal RJJ**, **Vollan AZ**, **Wang S**, **Welkenhuysen M**, **Ye Z**, **Dudman JT**, **Dutta B**, **Hantman AW**, **Harris KD**, **Lee AK**, **Moser EI**, **O’Keefe J**, **Renart A**, **Svoboda K**, **Häusser M**, **Haesler S**, **Carandini M**, **Harris TD**. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science (80- )* 372, 2021.

**Stevenson IH**. Flexible models for spike count data with both over- and under- dispersion. *J Comput Neurosci* 41: 29–43, 2016.

**Stevenson IH**, **Cherian A**, **London BM**, **Sachs NA**, **Lindberg E**, **Reimer J**, **Slutzky MW**, **Hatsopoulos NG**, **Miller LE**, **Kording KP**. Statistical assessment of the stability of neural movement representations. *J Neurophysiol* 106: 764–774, 2011.

**Tomko GJ**, **Crapper DR**. Neuronal Variability - Nonstationary Responses to Identical Visual-Stimuli. *Brain Res* 79: 405–418, 1974.

# Appendix

## Moments approximation for Conway-Maxwell Poisson distribution

Assume and , let , and

Then the moments are (The highlights can be found in the reference, the covariance is calculated by myself):

## Gradient and hessian for log-posterior

Denote and . The gradient:

, where , and .

The hessian:

, where

When , the hessian may be ill-conditioned or even positive-definite. Because of the factorial, the “outliers” are common. To ensure the robustness, do fisher scoring, i.e. replace the observed information by the expected information , so that .

## Quantifying Uncertainties

After the Newton update, we get . Then . The covariance matrix of two CMP parameters can be calculated by property of multivariate log-normal distribution.

The conditional mean firing rate, is , whose variance can be calculated by the Delta method:

We can calculate the moments as in the first section of the appendix, or we can use a simpler approximation when or . Then and .