

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

Ganchao Wei ¹
ganchao.wei@uconn.edu

Ian H. Stevenson ^{2, 3}
ian.stevenson@uconn.edu

¹ Department of Statistics, University of Connecticut

²Department of Psychological Sciences, University of Connecticut

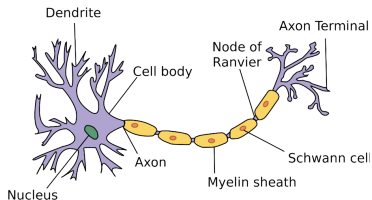
³Department of Biomedical Engineering, University of Connecticut

October 1, 2021

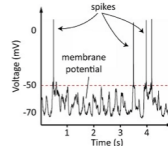
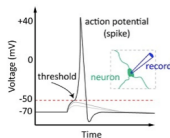
Introduction: Neuron and Neural Spikes

Neuron and Neural Spikes

Spikes for single neuron: $\{y_k\}_{k=0}^T$ for $y_k \in \mathbb{N}_{\geq 0}$



(a) Neuron



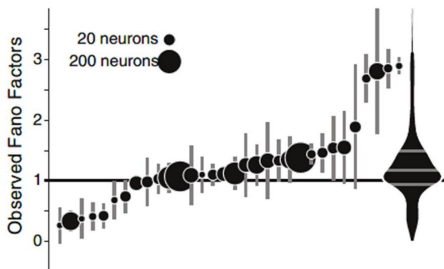
(b) Neural Spikes

Naturally, the spiking pattern (e.g. spiking rate & variance) will change along the time. **These changes tell us how brain process information!**

Introduction: Neuron and Neural Spikes

Poisson (mean = Var)? NB (mean \leq Var)? **Unrealistic!**

Fano factor = variance-to-mean ratio (σ^2/μ)



In summary, neural spikes:

- pattern (both mean & variance) changes along the time.
- non-Poisson, can be both over- and under-dispersed.

Non-Poisson Counts: Conway-Maxwell Distribution

Goal: describe over- and under dispersed count data flexibly.

Conway-Maxwell Poisson (CMP, jointly model **mean** & **variance**)

p.m.f. ($\lambda, \nu > 0$ or $0 < \lambda < 1, \nu = 0$)

$$P(X = x) = \frac{\lambda^x}{(x!)^\nu} \cdot \frac{1}{Z(\lambda, \nu)}$$

, for $x = 0, 1, \dots$ $Z(\lambda, \nu) = \sum_{y=0}^{\infty} \frac{\lambda^y}{(y!)^\nu}$ is the normalizing constant.

Parameter ν controls the dispersion pattern:

- $\nu = 1$: Poisson
- $\nu < 1$: over-dispersed ($\nu = 0$ Geometric)
- $\nu > 1$: under-dispersed ($\nu \rightarrow \infty$ Bernoulli)

Track the Change: State Space Model

Jointly model parameters of the i^{th} neuron at step k : $\log(\lambda_{ik}) = \mathbf{X}_{ik}\beta_k$ and $\log(\nu_{ik}) = \mathbf{G}_{ik}\gamma_k$. Denote $\theta_k = (\beta_k', \gamma_k')'$.

- The prior (state equation): $\theta_k | \theta_{k-1} \sim N(\mathbf{F}_k \theta_{k-1}, \mathbf{Q}_k)$
- The posterior: $P(\theta_k | \mathbf{Y}_{[k]}) \propto P(\mathbf{y}_k | \theta_k, \mathbf{Y}_{[k]}) P(\theta_k | \mathbf{Y}_{[k-1]})$

However, since the likelihood is CMP distributed \Rightarrow no closed posterior.

Normal approximation: at recursive prior.

- Assume approximated (log-posterior) gradient & hessian are the equal to true values.
- fast: one-time calculation.

Estimation: Filter by Normal Approximation (Original)

Normal approximation at recursive prior

- Prior:

$$\boldsymbol{\theta}_{k|k-1} = \mathbf{F}_{k-1}\boldsymbol{\theta}_{k-1|k-1}$$

$$\boldsymbol{\Sigma}_{k|k-1} = \mathbf{F}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\mathbf{F}_{k-1}' + \mathbf{Q}_k$$

- Posterior:

$$\boldsymbol{\theta}_{k|k} = \boldsymbol{\theta}_{k|k-1} + (\boldsymbol{\Sigma}_{k|k}) \left[\frac{\partial l_k}{\partial \boldsymbol{\theta}_k} \right]_{\boldsymbol{\theta}_{k|k-1}}$$

$$(\boldsymbol{\Sigma}_{k|k})^{-1} = (\boldsymbol{\Sigma}_{k|k-1})^{-1} - \left[\frac{\partial^2 l_k}{\partial \boldsymbol{\theta}_k \partial \boldsymbol{\theta}_k'} \right]_{\boldsymbol{\theta}_{k|k-1}}$$

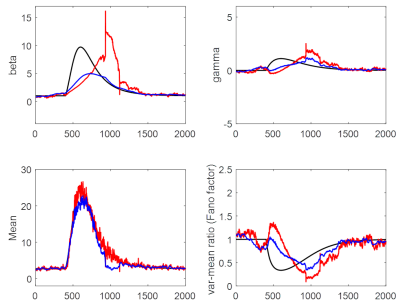
Great, fast 1-time calculation, but...

- Hessian is not robust to outliers.
- Evaluate things at recursive prior: may bias too much if true values are far from priors.

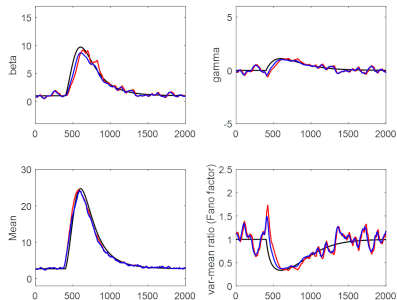
Improvement

- 1 Routinely improve filter by backward RTS smoother
- 2 Ensure positive-definite covariance (robustness): use expected information (like Fisher scoring).
- 3 Do exact Laplace approximation at the posterior mode: Because of Markovian assumption, the hessian for log-posterior is tri-block diagonal \Rightarrow can update efficiently in $\mathcal{O}(T)$ by Newton-Raphson, starting with smoother estimates.

Simulation



(c) original

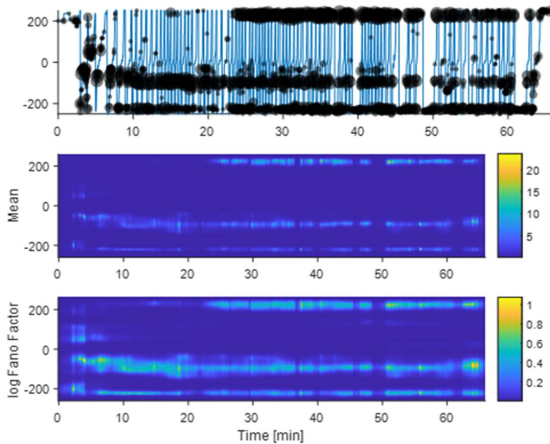


(d) after improvement 1, 2 and 4

black = true, **red** = filter, **blue** = smoother

Application: "place cell" in hippocampus

Experiment: a rat run back and forth along the linear track.



Appendix 1: Evaluation at Recursive Prior

Posterior:

$$\begin{aligned}P(\boldsymbol{\theta}_k | \mathbf{Y}_{[k]}) &\propto L_k \cdot P(\boldsymbol{\theta}_k | \mathbf{Y}_{[k-1]}) \\&= L_k \cdot \exp \left(-\frac{1}{2} (\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k|k-1})' (\boldsymbol{\Sigma}_{k|k-1})^{-1} (\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k|k-1}) \right) \\&\propto \exp \left(-\frac{1}{2} (\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k|k})' (\boldsymbol{\Sigma}_{k|k})^{-1} (\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k|k}) \right)\end{aligned}$$

Take 1st and 2nd derivatives, w.r.t. $\boldsymbol{\theta}_k$:

$$\begin{aligned}\left(\frac{\partial l_k}{\partial \boldsymbol{\theta}_k} \right)' - (\boldsymbol{\Sigma}_{k|k-1})^{-1} (\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k|k-1}) &= -(\boldsymbol{\Sigma}_{k|k})^{-1} (\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k|k}) \\ \frac{\partial^2 l_k}{\partial \boldsymbol{\theta}_k \partial \boldsymbol{\theta}_k'} - (\boldsymbol{\Sigma}_{k|k-1})^{-1} &= -(\boldsymbol{\Sigma}_{k|k})^{-1}\end{aligned}$$

Appendix 2: Fisher Scoring

gradient and hessian of log-likelihood:

$$\begin{aligned}\left[\frac{\partial l_k}{\partial \theta_k}\right]_{\theta_{k|k-1}} &= \sum_{i=1}^{n_k} \left(\frac{(y_{ik} - E(Y_{ik})) \mathbf{x}_{ik}}{\nu_{ik} (E(\log Y_{ik}!) - \log y_{ik}!)} \mathbf{g}_{ik} \right)_{\theta_{k|k-1}} \\ \left[-\frac{\partial^2 l_k}{\partial \theta_k \partial \theta_k'}\right]_{\theta_{k|k-1}} &= \sum_{i=1}^{n_k} \begin{pmatrix} A_{ik} & B_{ik} \\ B_{ik}' & C_{ik} \end{pmatrix}\end{aligned}$$

, where

$$A_{ik} = \text{Var}(Y_{ik}) \mathbf{x}_{ik} \mathbf{x}_{ik}'$$

$$B_{ik} = -\nu_{ik} \text{Cov}(Y_{ik}, \log Y_{ik}!) \mathbf{x}_{ik} \mathbf{g}_{ik}'$$

$$C_{ik} = \nu_{ik} (\nu_{ik} \text{Var}(\log Y_{ik}!) - E(\log Y_{ik}!) + \log y_{ik}!) \mathbf{g}_{ik} \mathbf{g}_{ik}'$$

C_{ik} is not robust to outliers. Replace it by the expected value:

$$C_{ik}^* = \nu_{ik}^2 \text{Var}(\log Y_{ik}!) \mathbf{g}_{ik} \mathbf{g}_{ik}'$$

Appendix 3: Tri-block Diagonal Hessian

Hessian for log-posterior:

$$H = \frac{\partial^2 \log P(\theta|\mathbf{Y})}{\partial \theta \partial \theta'} = \begin{pmatrix} \frac{\partial^2 \log P(\theta|\mathbf{Y})}{\partial \theta_1 \partial \theta'_1} & \mathbf{F}'_2 \mathbf{Q}_2^{-1} & \dots & 0 \\ \mathbf{Q}_2^{-1} \mathbf{F}_2 & \frac{\partial^2 \log P(\theta|\mathbf{Y})}{\partial \theta_2 \partial \theta'_2} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \frac{\partial^2 \log P(\theta|\mathbf{Y})}{\partial \theta_T \partial \theta'_T} \end{pmatrix}$$

,where

$$\frac{\partial^2 \log P(\theta|\mathbf{Y})}{\partial \theta_1 \partial \theta'_1} = \frac{\partial^2 l_1}{\partial \theta_1 \partial \theta'_1} - \Sigma_0^{-1} - \mathbf{F}'_2 \mathbf{Q}_2^{-1} \mathbf{F}_2$$

$$\frac{\partial^2 \log P(\theta|\mathbf{Y})}{\partial \theta_k \partial \theta'_k} = \frac{\partial^2 l_k}{\partial \theta_k \partial \theta'_k} - \mathbf{Q}_k^{-1} - \mathbf{F}'_{k+1} \mathbf{Q}_{k+1}^{-1} \mathbf{F}_{k+1}$$

$$\frac{\partial^2 \log P(\theta|\mathbf{Y})}{\partial \theta_T \partial \theta'_T} = \frac{\partial^2 l_T}{\partial \theta_T \partial \theta'_T} - \mathbf{Q}_T^{-1}$$