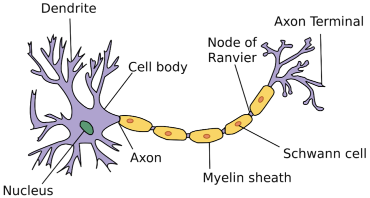
Dynamic modeling of neural spike count data with non-Poisson variability

Ganchao Wei Ian H. Stevenson

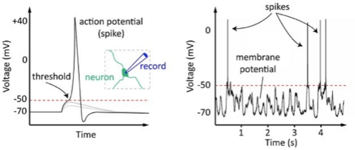
University of Connecticut

October 1, 2021

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



(a) Neuron

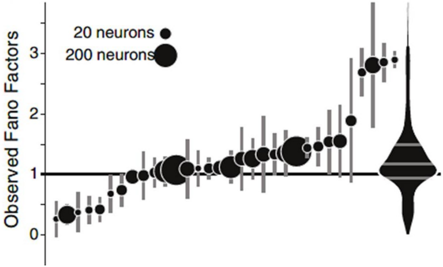


(b) Neural Spikes

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 $x = 0, 1, \ldots$ $Z(\lambda, \nu) = \sum_{y=0}^{\infty} \frac{\lambda^y}{(y!)^\nu}$ is the normalizing constant.
Parameter ν controls the dispersion pattern:

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

Track the Change: State Space Model

Jointly model parameters of the i^{th} neuron at step k: $\log(\boldsymbol{\lambda}_{ik}) = \boldsymbol{X}_{ik}\boldsymbol{\beta}_k$ and $\log(\boldsymbol{\nu}_{ik}) = \boldsymbol{G}_{ik}\gamma_k$. Denote $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k', \gamma_k')'$.

- The prior (state equation): $\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k-1} \sim N(\boldsymbol{F}_k \boldsymbol{\theta}_{k-1}, \boldsymbol{Q}_k)$
- The posterior: $P\left(\boldsymbol{\theta}_k \middle| \boldsymbol{Y}_{[k]}\right) \propto P\left(\boldsymbol{y}_k \middle| \boldsymbol{\theta}_k,\ \boldsymbol{Y}_{[k]}\right) P\left(\boldsymbol{\theta}_k \middle| \boldsymbol{Y}_{[k-1]}\right)$

However, since the likelihood is CMP distributed $\Rightarrow$ no closed posterior. Normal approximation: at recursive prior.

- Assume approximated (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:

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

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

- Posterior:

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

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

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

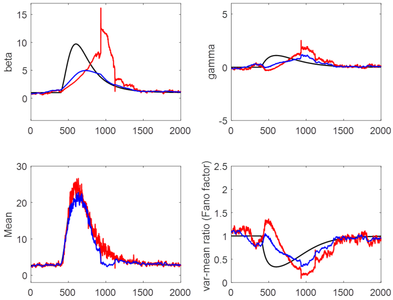
- Hessian is not robust to outliers.
- Evaluate things at recursive prior: may bias too much if true values are far from priors.

Improvement

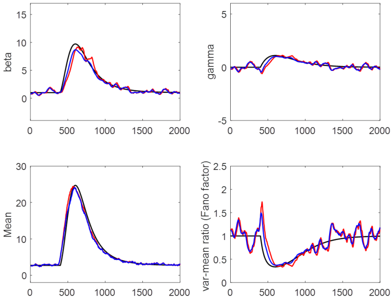
1. Routinely improve filter by backward RTS smoother
2. Ensure positive-definite covariance (robustness): use expected information (like Fisher scoring).
3. Do exact Laplace approximation at the posterior mode: Because of Markovian assumption, the hessian for log-posterior is tri-block diagonal $\Rightarrow$ can update efficiently in $\mathcal{O}(T)$ by Newton-Raphson, starting with smoother estimates.
4. Enlarge the local sample size at each step, by assuming stationary state vectors within a pre-specified window. (select window size by forward-chaining)

Both (3) and (4) have their own strengths. When spikes are sparse, (4) is better.

Simulation



(c) original

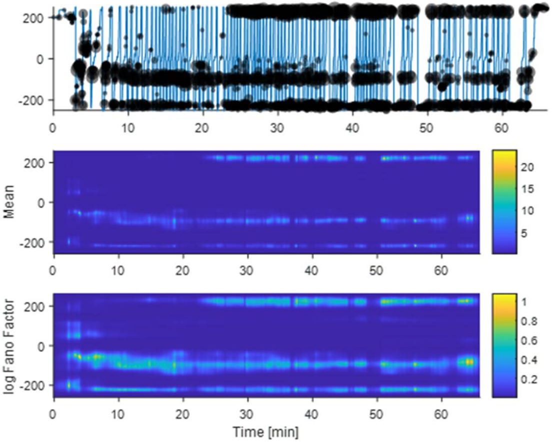


(d) after improvement 1, 2 and 4

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:

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

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

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

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

Appendix 2: Fisher Scoring

gradient and hessian of log-likelihood:

$$\left[\frac{\partial l_k}{\partial \boldsymbol{\theta}_k}\right]_{\boldsymbol{\theta}_{k|k-1}} = \sum_{i=1}^{n_k} \begin{pmatrix} \left(y_{ik} - E\left(Y_{ik}\right)\right) \boldsymbol{x}_{ik} \\ \nu_{ik}\left(E\left(\log Y_{ik}!\right) - \log y_{ik}!\right) \boldsymbol{g}_{ik} \end{pmatrix}_{\boldsymbol{\theta}_{k|k-1}}$$

$$\left[-\frac{\partial^2 l_k}{\partial \theta_k \partial \theta'_k}\right]_{\theta_{k|k-1}} = \sum_{i=1}^{n_k} \begin{pmatrix} A_{ik} & B_{ik} \\ B'_{ik} & C_{ik} \end{pmatrix}$$

, where

$$\begin{aligned} A_{ik} &= Var(Y_{ik}) \boldsymbol{x}_{ik} \boldsymbol{x}'_{ik} \\ B_{ik} &= -\nu_{ik} Cov(Y_{ik}, \log Y_{ik}!) \boldsymbol{x}_{ik} \boldsymbol{g}'_{ik} \\ C_{ik} &= \nu_{ik}(\nu_{ik} Var(\log Y_{ik}!) - E(\log Y_{ik}!) + \log y_{ik}!) \boldsymbol{g}_{ik} \boldsymbol{g}'_{ik} \end{aligned}$$

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

$$C^*_{ik} = \nu^2_{ik} Var(\log Y_{ik}!) \boldsymbol{g}_{ik} \boldsymbol{g}'_{ik}$$

Appendix 3: Tri-block Diagonal Hessian

Hessian for log-posterior:

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

,where

$$\frac{\partial^2 \log P(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1'} = \frac{\partial^2 l_1}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1'} - \boldsymbol{\Sigma}_0^{-1} - \boldsymbol{F}_2' \boldsymbol{Q}_2^{-1} \boldsymbol{F}_2$$

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

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