Testing Between Different Types of Poisson Mixtures with Applications to Neuroscience

by

Yunran Chen

Department of Statistical Science
Duke University

Date: _____

Approved:

_____
Surya Tokdar, Supervisor

_____
Alexander Volfovsky

_____
Mike West

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Department of Statistical Science
in the Graduate School of
Duke University

2019

ABSTRACT

Testing Between Different Types of Poisson Mixtures with Applications to Neuroscience

by

Yunran Chen

Department of Statistical Science
Duke University

Date: _____

Approved:

_____
Surya Tokdar, Supervisor

_____
Alexander Volfovsky

_____
Mike West

Abstract

We propose a hypothesis testing for different types of stochastic order of mixture distributions (PRML classifier) and a hypothesis testing for screening out data with mixture distributions (PRML filter), in a Bayesian framework using a recursive algorithm called predictive recursion marginal likelihood (PRML) algorithm. Of particular interest is the special case of testing between different types of Poisson mixtures and testing Poisson distribution versus Poisson mixtures. The first testing procedure applies Laplace approximation coupled with optimization algorithm. This testing helps neuroscientists to classify the activation patterns that a single neuron exhibits when preserving information from multiple stimuli. The second testing aims to screen out over-dispersed data to boost the scientific information. Simulation shows the new classifier and filter outperform the previous testing especially for over-dispersed data. We apply the PRML classifier on the analysis of inferior colliculus neurons filtered by PRML filter. We show the PRML classifier emphasizes second order stochasticity. We present empirical evidence that the PRML filter contributes to avoid mistaking trial-to-trial variation as second order stochasticity.

Acknowledgements

I would first like to express my sincere gratitude to my thesis advisor, Dr. Surya Tokdar, for his guidance and continuous support during the research. Surya taught me a great deal about both scientific research and collaboration with researchers from other fields, and also provided me invaluable advice on my career paths. Without his guidance, I would not able to make such progress.

I would also like to thank Dr. Mike West for serving as my thesis committee and providing insightful comments to incent me to have a comprehensive understanding of the research. My sincere thanks also goes to Dr. Alexander Volfovsky for his enthusiasm and encouragement, which stimulated my interests in research, as well as his support and understanding.

I am grateful to all of those with whom I have had pleasure to work during this research project. Also, I would like to thank my family, my friends and Arashi for supporting me spiritually.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

An important issue in neuroscience is the question of how a single neuron preserves information from multiple stimuli. For simplicity consider the case where two stimuli, namely A and B, may be presented simultaneously, and neural activity is recorded by the count of spikes emitted by a neuron within a given response window. Let P^A, P^B and P^{AB} denote the probability distribution functions of spike counts under the three possible scenarios: (A) only stimulus A is presented, (B) only stimulus B is presented, and (AB) stimuli A and B are presented together.

Conventional analysis of neural activity recording under a single stimulus typically works under an assumption of first order stochasticity, where the recorded spike count is seen as a noisy realization of an expected count. In this case, the spike count distribution is often well represented by a Poisson distribution (Ventura et al., 2002; Kass et al., 2005).

While such a Poisson assumption may be appropriate for P^A and P^B, Caruso et al. (2018); Glynn et al. (2019) argue that the possibility of a second order stochasticity may give P^{AB} interesting non-Poisson shapes. Second order stochasticity is taken to be information-encoding random variation across trials that is instrumental to preserve information on each constituent member of the stimuli set.

For example, when the neuron is believed to randomly switch across AB trials between its A and B firing patterns, one may represent P^{AB} as a mixture distribution

$$\text{(Mixture)} \quad P^{AB} = \alpha P^A + (1-\alpha)P^B, \tag{1.1}$$

for some unknown mixing proportion $\alpha \in (0, 1)$. In addition to this 'Mixture' type, several other biologically meaningful types of second order stochasticity may manifest, e.g.,

$$(\text{Intermediate}) \quad P^{\text{A}} < P^{\text{AB}} < P^{\text{B}} \text{ or } P^{\text{B}} < P^{\text{AB}} < P^{\text{A}}, \tag{1.2}$$

$$(\text{Outside}) \quad P^{\text{AB}} < P^{\text{A}}, P^{\text{B}} \text{ or } P^{\text{AB}} > P^{\text{A}}, P^{\text{B}}, \tag{1.3}$$

$$(\text{Single}) \quad P^{\text{AB}} = P^{\text{A}} \text{ or } P^{\text{AB}} = P^{\text{B}}. \tag{1.4}$$

where a stochastic order $P^{\text{A}} < P^{\text{B}}$ indicates random variable Y^{A} is less than Y^{B} stochastically. Specifically, for any $x \in (-\infty, \infty)$, we have $Pr(Y^{\text{A}} > x) < Pr(Y^{\text{B}} > x)$.

The 'Intermediate' hypothesis is appropriate when the neuron, under condition AB, behaves as the 'Mixture' type but switches between A and B patterns at a faster time scale that is shorter than the response window, or responds with firing rates that are intermediate between its A and B firing patterns. If the neuron under AB trial acts as a summation-type of signal A and B, or in contrast, is inhibited by the presence of both A and B, P^{AB} may be outside both P^{A} and P^{B}, called the 'Outside' type. Otherwise, if the neuron under AB trials acts similar to either A or B firing patterns, we call it 'Single' type.

Caruso et al. (2018) use a Poisson assumption for $P^{\text{A}} = \text{Poi}(\mu^{\text{A}})$ and $P^{\text{B}} = \text{Poi}(\mu^{\text{B}})$, and also for $P^{\text{AB}} = \text{Poi}(\mu^{\text{AB}})$ in the case of 'Intermediate' and 'Outside'. Specifically, Caruso et al. (2018) assume $P^{\text{AB}} = \text{Poi}(\mu^{\text{AB}})$ where the rate μ^{AB} lies between μ^{A} and μ^{B} for 'Intermediate', and is outside μ^{A} and μ^{B} for 'Outside' hypothesis. Such assumption ignores the full scale of second order stochasticity as represented by the stochastic ordering relations in equation 1.2 and 1.3. An important question is whether efficient tests can be designed while replacing the Poisson assumption on

P^{AB} under 'Intermediate' and 'Outside' hypotheses with more flexible distributional models that allow for possible overdispersion due to second order stochasticity.

We achieve this by extending the Poisson assumption to a Poisson mixtures assumption $P^{\mathrm{AB}} = \int Poi(\mu^{\mathrm{AB}})dF(\mu^{\mathrm{AB}})$ under 'Intermediate' and 'Outside' case. And thus, μ^{AB} may vary across trials within a range between μ^{A} and μ^{B} for 'Intermediate', and outside μ^{A} and μ^{B} for 'Outside'. The across-trial variation of μ^{AB} is measured by an unknown distribution $F(\mu^{\mathrm{AB}})$, which captures overdispersion due to second order stochasticity.

With specified distribution for single-stimulus and dual-stimuli spike counts, we construct hypothesis testing based on Bayesian framework and choose the model with the highest posterior probability. In order to estimate the marginal likelihood to further obtain posterior probability, we apply Predictive Recursion Marginal Likelihood (PRML) algorithm (Newton et al., 1998; Newton, 2002; Martin and Tokdar, 2011) coupled with optimization algorithm. Martin and Tokdar (2011) shows the marginal likelihood estimation from PRML algorithm can be considered as an online filtering approximation to a Dirichlet process mixture marginal likelihood computation; and thus, can be applied within the Bayesian framework. Furthermore, we can also obtain a density estimation of the mixing density from the PRML algorithm to unveil the underlying mechanism.

In real data analysis, before testing between different hypotheses shown in equations 1.1 to 1.4, Caruso et al. (2018) apply Pearson's Chi-squared testing to screen out the data with non-Poisson-like single-stimulus spike counts. In our case, we need to screen out overdispersed Poisson distributed single-stimulus data. The filtering procedure is necessary to avoid introducing trial-to-trial variation caused by unmeasured extrinsic conditions in single-stimulus trials into the second order stochasticity in dual-stimuli activation patterns. Therefore, we apply PRML algorithm to con-

struct a hypothesis testing, called PRML filter, to test Poisson distribution versus Poisson mixtures based on Bayes factor. Simulation shows the traditional Pearson's Chi-squared test is less powerful than PRML filter for screening out overdispersed data.

In the following chapter, we first introduce PRML algorithm and PRML gradient (PRMLG) algorithm. Then, we present how to apply the two algorithms within Bayesian framework to construct PRML classsifier, modified PRML classifier and PRML filter. In Chapter 3 and Chapter 4, we compare our new classifier and filter to the ones proposed by Caruso et al. (2018) on simulation datasets and real datasets respectively. In Chapter 5, we conclude with a discussion. The Appendix includes several technical details on algorithm, related tables on simulation and an additional figure of real data analysis.

Chapter 2

Hypothesis Testing

2.1 Predictive Recursion Marginal Likelihood Algorithm

Consider data $Y_1, ..., Y_n$ come from mixture distribution with the following density function independently:

$$m_f(y) = \int p(y|u)f(u)d\mu(u) \tag{2.1}$$

where $(y, u) \mapsto p(y|u)$ is a known kernel on $\mathcal{Y} \times \mathcal{U}$ and f is an unknown mixing density in $\mathbb{F}$, which is the set of densities with respect to a σ-definite Borel measure μ on $\mathcal{U}$.

Newton et al. (1998) propose predictive recursion (PR) algorithm to estimate mixing densities f and density of mixture distributions $m_f(y)$. It shows great computation efficiency and flexibility for allowing any specified $\mathcal{U}$. Tokdar et al. (2009) shows the consistency of the resulting estimation of mixing density f_n. Furthermore, Martin and Tokdar (2011) introduce marginal likelihood estimation $L_n = \prod_{i=1}^{n} m_{i-1}(Y_i)$ as a byproduct of PR algorithm, which can be seen as Dirichlet process mixtures model; and thus, we can apply the marginal likelihood estimator within the Bayesian framework. The Predictive Recursion Marginal likelihood algorithm (PRML) proposed by Martin and Tokdar (2011) works as Algorithm 1 shown.

Furthermore, Martin and Tokdar (2011) extend PRML algorithm to allow extra

Input: i.i.d observations $Y_1, ..., Y_n$
Output: marginal likelihood $L_n = \prod_{i=1}^{n} m_{i-1}(Y_i)$, estimation on mixing
  density f_n
Initialize: $f_0 \in \mathbb{F}$ of f(usually uniform);
weights $w_1, .., w_n \in (0, 1)$(usually $w_i = (i+1)^{-\gamma}$)
for $i = 1,...,n$ **do**

$$m_{i-1}(Y_i) = \int p(Y_i|u')f_{i-1}(u')d\mu(u') \tag{2.2}$$

$$f_i(u) = (1 - w_i)f_{i-1}(u) + w_i \frac{p(Y_i|u)f_{i-1}(u)}{m_{i-1}(Y_i)} \tag{2.3}$$

end

Algorithm 1: Pseudocode for PRML Algorithm

unknown parameters $\theta \in \Theta$ in the kernel $p(y|u, \theta)$ of the mixture distribution, and develop PRML gradient (PRMLG) algorithm (See Algorithm 2) to provide gradient of PR marginal likelihood as a byproduct to facilitate the computation.

It is worth noting that the weights w_i should decay with i increase, satisfying $\sum_{i=1}^{\infty} w_i = \infty, \sum_{i=1}^{\infty} w_i^2 < \infty$. Therefore, according to equation 2.3 2.5, the data point entering later makes less influence on the estimation. To avoid the case that estimator may depend on the order of the data, we consider a permutation invariant version $L_p = \frac{1}{n!} \sum_{s \in S_n} L$, where S_n is the permutation group on $\{1, ..., n\}$.

2.2 PRML Classifier

Denote the three experiment conditions as triplet $\{A, B, AB\}$. Under the experiment conditions $\{A, B, AB\}$, let $\{n_A, n_B, n_{AB}\}$, $\{Y_j^A, Y_j^B, Y_j^{AB}\}$ and $\{\mu^A, \mu^B, \mu^{AB}\}$ represent the number of repeated trials, the spike counts under trial j and Poisson rate respectively. Without loss of generosity, consider $\mu^A < \mu^B$. We restate the testing in Chapter 1 as follows:

1. Consider $Y_j^A \overset{i.i.d}{\sim} Poi(\mu^A)$, $Y_j^B \overset{i.i.d}{\sim} Poi(\mu^B)$ for unknown μ^A, μ^B

6

Input: i.i.d observations $Y_1, ..., Y_n$
Output: marginal likelihood $L_n(\theta) = \prod_{i=1}^{n} m_{i,\theta}(Y_i)$, gradient $\nabla \log L_n(\theta)$ and
 mixing density $f_{n,\theta}$
Initialize: $f_{0,\theta} \in \mathbb{F}$ of f (usually uniform);compute $\nabla f_{0,\theta}(u)$
weights $w_1, .., w_n \in (0, 1)$ (usually $w_i = (i+1)^{-\gamma}$)
for $i = 1,...,n$ **do**

$$m_{i,\theta}(Y_i) = \int p(Y_i|\theta, u')f_{i-1,\theta}(u')d\mu(u') \qquad (2.4)$$

$$f_{i,\theta}(u) = (1 - w_i)f_{i-1,\theta}(u) + w_i\frac{p(Y_i|\theta, u)f_{i-1,\theta}(u)}{m_{i,\theta}(Y_i)} \qquad (2.5)$$

For $\nabla f_{i,\theta}(u)$:

$$G(\theta, u) = p(Y_i|\theta, u)\nabla f_{i-1,\theta}(u) + \nabla p(Y_i|\theta, u)f_{i-1,\theta}(u) \qquad (2.6)$$

$$\nabla \log m_{i,\theta}(Y_i) = \frac{\int G(\theta, u)d\mu(u)}{m_{i,\theta}(Y_i)} \qquad (2.7)$$

$$\nabla f_{i,\theta}(u) = (1-w_i)\nabla f_{i-1,\theta}(u)+w_i\{\frac{G(\theta, u) - p(Y_i|\theta, u)f_{i-1,\theta}(u)\nabla \log m_{i,\theta}(Y_i)}{m_{i,\theta}(Y_i)}\} \qquad (2.8)$$

end

Algorithm 2: Pseudocode for PRMLG Algorithm

2. Consider $Y_j^{\mathrm{AB}} \overset{i.i.d}{\sim} \int Poi(\mu^{\mathrm{AB}}) f(\mu^{\mathrm{AB}}) d\mu^{\mathrm{AB}}$ with four competing scenarios with different support of f:

   (a) Mixture: for unknown $\mu^{\mathrm{AB}} \in \{\mu^{\mathrm{A}}, \mu^{\mathrm{B}}\}$

   (b) Intermediate: for unknown $\mu^{\mathrm{AB}} \in (\mu^{\mathrm{A}}, \mu^{\mathrm{B}})$

   (c) Outside: for unknown $\mu^{\mathrm{AB}} \in [\mu_{\mathrm{L}}, \mu^{\mathrm{A}})$ or $\mu^{\mathrm{AB}} \in (\mu^{\mathrm{B}}, \mu_{\mathrm{U}}]$, where known $\mu_{\mathrm{L}}, \mu_{\mathrm{U}}$ indicate the lower bound and upper bound of μ^{AB}.

   (d) Single: for $\mu^{\mathrm{AB}} = \mu^{\mathrm{A}}$ or $\mu^{\mathrm{AB}} = \mu^{\mathrm{B}}$

We choose the model with the highest posterior probability. By assuming uniform prior on the four models, the core problem is to compare the corresponding marginal likelihood $p(Y^{AB}|Y^{\mathrm{A}}, Y^{\mathrm{B}})$. Here we pick the 'Intermediate' hypothesis as an example.

Consider marginal likelihood as a joint density function with all the related parameters integrated out, we have

$$
\begin{aligned}
p(Y^{\mathrm{AB}}|Y^{\mathrm{A}}, Y^{\mathrm{B}}) &= \int\int p(Y^{\mathrm{AB}}|\mu^{\mathrm{A}}, \mu^{\mathrm{B}}) p(\mu^{\mathrm{A}}, \mu^{\mathrm{B}}|Y^{\mathrm{A}}, Y^{\mathrm{B}}) d\mu^{\mathrm{A}} d\mu^{\mathrm{B}} \\
&= \int\int \int_{\mu^{\mathrm{A}}}^{\mu^{\mathrm{B}}} p(Y^{\mathrm{AB}}|\mu^{\mathrm{AB}}) f(\mu^{\mathrm{AB}}) d\mu^{\mathrm{AB}} p(\mu^{\mathrm{A}}, \mu^{\mathrm{B}}|Y^{\mathrm{A}}, Y^{\mathrm{B}}) d\mu^{\mathrm{A}} d\mu^{\mathrm{B}}
\end{aligned}
$$

We apply Laplace approximation then obtain marginal likelihood estimator as follows:

$$
p(Y^{\mathrm{AB}}|Y^{\mathrm{A}}, Y^{\mathrm{B}}) \approx (2\pi)^{k/2} |-H|^{-1/2} e^{l(\hat{\mu}^{\mathrm{A}}, \hat{\mu}^{\mathrm{B}})} \tag{2.9}
$$

where $l(\mu^{\mathrm{A}}, \mu^{\mathrm{B}}) = \log p(Y^{\mathrm{AB}}|\mu^{\mathrm{A}}, \mu^{\mathrm{B}}) p(\mu^{\mathrm{A}}, \mu^{\mathrm{B}}|Y^{\mathrm{A}}, Y^{\mathrm{B}})$, $H = \nabla^2 l(\mu^{\mathrm{A}}, \mu^{\mathrm{B}})|_{\hat{\mu}^{\mathrm{A}}, \hat{\mu}^{\mathrm{B}}}$, $k = \dim((\mu^{\mathrm{A}}, \mu^{\mathrm{B}})) = 2$, and $(\hat{\mu}^{\mathrm{A}}, \hat{\mu}^{\mathrm{B}}) = \arg\max_{\mu^{\mathrm{A}}, \mu^{\mathrm{B}}} \log p(Y^{\mathrm{AB}}|\mu^{\mathrm{A}}, \mu^{\mathrm{B}}) p(\mu^{\mathrm{A}}, \mu^{\mathrm{B}}|Y^{\mathrm{A}}, Y^{\mathrm{B}})$.

By applying Laplace approximation, density estimation problem is transferred to an optimization problem with objection function $l(\mu^{\mathrm{A}}, \mu^{\mathrm{B}})$. To evaluate $l(\mu^{\mathrm{A}}, \mu^{\mathrm{B}}) = \log p(Y^{AB}|\mu^{\mathrm{A}}, \mu^{\mathrm{B}})p(\mu^{\mathrm{A}}, \mu^{\mathrm{B}}|Y^{\mathrm{A}}, Y^{\mathrm{B}})$, we assume μ^{A} and μ^{B} are independent and set the Jeffereys' prior for μ^{A} and μ^{B} to get the posterior density $p(\mu^{\mathrm{A}}, \mu^{\mathrm{B}}|Y^{\mathrm{A}}, Y^{\mathrm{B}})$. For the mixture distribution with unknown parameter in the domain of f as follows,

$$p(Y^{\mathrm{AB}}|\mu^{\mathrm{A}}, \mu^{\mathrm{B}}) = \int_{\mu^{\mathrm{A}}}^{\mu^{\mathrm{B}}} p(Y^{\mathrm{AB}}|\mu^{\mathrm{AB}})f(\mu^{\mathrm{AB}})d\mu^{\mathrm{AB}} \qquad (2.10)$$

in order to apply PRMLG algorithm, we need to transfer the unknown parameter from the domain of f to the kernel by reparametrization.

Consider

$$\mu^{AB} = h(z) = \mu^{\mathrm{A}} + z(\mu_{\mathrm{B}} - \mu_{\mathrm{A}}) \qquad (2.11)$$

, which is a linear mapping $h : z \mapsto \mu^{AB}$, from $[0, 1]$ to $(\mu^{\mathrm{A}}, \mu^{\mathrm{B}})$. We then represent the integral in equation 2.10 with respect to z as follows:

$$p(Y^{\mathrm{AB}}|\mu^{\mathrm{A}}, \mu^{\mathrm{B}}) = \int_0^1 p(Y^{\mathrm{AB}}|h(z))\tilde{f}(z|\mu^{\mathrm{A}}, \mu^{\mathrm{B}})dz \qquad (2.12)$$

We can apply PRMLG algorithm to estimate the marginal likelihood of mixture distribution shown in equation 2.12. We consider the range of μ^{AB} in original equation 2.10 to be an interval between μ^{A} and μ^{B}, and thus consider a uniform initialization $f_0 = \mathrm{Unif}(\mu^{\mathrm{A}}, \mu^{\mathrm{B}})$. According to parameterization (see equation 2.11), in Algorithm 2, we set $\tilde{f}_0 = \mathrm{Unif}(0, 1)$, because $h(z)$ is a linear transformation; we calculate $m_{i,\theta}(Y_i)$ as an integral on $(0, 1)$.

Coupled with any optimization algorithm, we now can obtain the estimated parameters $(\hat{\mu}^{\mathrm{A}}, \hat{\mu}^{\mathrm{B}})$, object function $l(\hat{\mu}^{\mathrm{A}}, \hat{\mu}^{\mathrm{B}})$ and corresponding Hessian matrix $H|_{\mu^{\mathrm{A}}=\hat{\mu}^{\mathrm{A}}, \mu^{\mathrm{B}}=\hat{\mu}^{\mathrm{B}}}$ as output. It is worth noting that it is a constrained optimization

problem for we assume $0 < \mu^{\mathrm{A}} < \mu^{\mathrm{B}}$. In order to remove the restriction to further ease the computation, we apply log transformation

$$\psi^{\mathrm{A}} = log(\mu^{\mathrm{A}}), \quad \psi^{\mathrm{B}} = log(\mu^{\mathrm{B}}) \tag{2.13}$$

to remove the restriction, and sort the $(\mu^{\mathrm{A}}, \mu^{\mathrm{B}})$ within each iteration. A Jacobian multiplier is added to the gradient and Hessian matrix in the unconstrained optimization procedure.

Under other hypotheses, the Poisson mixtures model shown in equation 2.10 is different at the support and continuity of mixing density f, leading to different specification on initial value f_0 and calculation of integral $m_{i,\theta}(Y_i)$ (See Table 2.1). Similarly, the linear mapping $h(\cdot)$ shown in equation 2.11 differs for different hypothesis as shown in Table 2.2. The transformation shown in equation 2.13 differs for different model assumption as shown in Table 2.3.

Table 2.1: PRML setting under different model assumptions before parametrization

Model	Support	f_0	$m_f(y)$
Mixture	$\{\mu^{\mathrm{A}}, \mu^{\mathrm{B}}\}$	$(0.5, 0.5)$	$\sum_{A,B} p(Y_i^{AB}|\mu^{AB}) f_{i-1}(\mu')$
Intermediate	$(\mu^{\mathrm{A}}, \mu^{\mathrm{B}})$	$\mathrm{Unif}(\mu^{\mathrm{A}}, \mu^{\mathrm{B}})$	$\int_{\mu^{\mathrm{A}}}^{\mu^{\mathrm{B}}} p(Y_i^{AB}|\mu') f_{i-1}(\mu') d\mu'$
OutsideA	$(\mu_{\mathrm{L}}, \mu^{\mathrm{A}})$	$\mathrm{Unif}(\mu_{\mathrm{L}}, \mu^{\mathrm{A}})$	$\int_{\mu_{\mathrm{L}}}^{\mu^{\mathrm{A}}} p(Y_i^{AB}|\mu') f_{i-1}(\mu') d\mu'$
OutsideB	$(\mu^{\mathrm{B}}, \mu_{\mathrm{U}})$	$\mathrm{Unif}(\mu^{\mathrm{A}}, \mu_{\mathrm{U}})$	$\int_{\mu^{\mathrm{B}}}^{\mu_{\mathrm{U}}} p(Y_i^{AB}|\mu') f_{i-1}(\mu') d\mu'$

Table 2.2: Reparameterization to facilitate the calculation of the gradient

Model	Reparameterization	Support
Mixture	$\mu^{AB} = h(z) = \mu^{\mathrm{A}} + z(\mu^{\mathrm{B}} - \mu^{\mathrm{A}})$	$\{0, 1\}$
Intermediate	$\mu^{AB} = h(z) = \mu^{\mathrm{A}} + z(\mu^{\mathrm{B}} - \mu^{\mathrm{A}})$	$[0, 1]$
OutsideA	$\mu^{AB} = h(z) = \mu_{\mathrm{L}} + z(\mu^{\mathrm{A}} - \mu_{\mathrm{L}})$	$[0, 1]$
OutsideB	$\mu^{AB} = h(z) = \mu^{\mathrm{B}} + z(\mu_{\mathrm{U}} - \mu_B)$	$[0, 1]$

Table 2.3: Restriction and corresponding transformation to free from restriction

Model	Restriction	Reparameterization
Mixture	$\mu^{\mathrm{A}}, \mu^{\mathrm{B}} > 0$	$\psi = (\log(\mu^{\mathrm{A}}), \log(\mu^{\mathrm{B}}))$
Intermediate	$0 < \mu^{\mathrm{A}} < \mu^{\mathrm{B}}$	$\psi = (\log(\mu^{\mathrm{A}}), \log(\mu^{\mathrm{B}}))$
Outside(left)	$0 < \mu_{\mathrm{L}} < \mu^{\mathrm{A}}$	$\psi = \log(\mu^{\mathrm{A}} - \mu_{\mathrm{L}})$
Outside(right)	$0 < \mu^{\mathrm{B}} < \mu_{\mathrm{U}}$	$\psi = \mathrm{logit}(\frac{\mu^{\mathrm{B}}}{\mu_{\mathrm{U}}})$

2.3 Modified PRML classifier

Due to the continuity of the domain of f, the 'Single' hypothesis can be considered as a special case of the 'Mixture' hypothesis. The 'Mixture' hypothesis can also be considered as a special case of the 'Intermediate' hypothesis. The vague separation between the 'Single', 'Mixture' and the 'Intermediate' hypotheses may lead to a relative less posterior probability of the 'Single' and 'Mixture' hypotheses. Therefore, we create a gap between the 'Intermediate' and 'Mixture' hypotheses on the domain of f, a gap between the 'Mixture' and 'Single' hypotheses on the measurement of f.

To separate the 'Intermediate' and 'Mixture' hypotheses, we introduce a hyperparameter e, and restrict $z \in [e, 1-e]$ with $e \in [0,1]$. For separation between the 'Mixture' hypothesis and the 'Single' hypothesis, we consider a separation on the measurement, which is $f(z) \in \{e, 1-e\}$. In other words, we only consider data comes from the 'Mixture' hypothesis only when the smallest mixing proportion is greater than e. Since the PRMLG algorithm is to estimate mixture density $m_f(y) = \int p(y|u) f(u) d\mu(u)$ with known kernel $p(y|u)$ and unknown mixing density f, we apply a reparametrization on f to transfer the constraint effect on f to a new defined kernel $p^e(y|u)$. Denote the original model as $\int p(y|z) f(z) dz$ where $z \in \{0,1\}$, $f(0) + f(1) = 1$ and with constraint $f(z) \in [e, 1-e]$. We consider a mapping $f(z) = e + (1-2e)g(z)$, which leads to a mixture density with new defined kernel as a function of e, which is $p^e(y|z) = e \times (p(y|0) + p(y|1)) + (1-2e)p(y|z)$. Therefore

11

we apply PRMLG algorithm to estimate the mixture density $\int p^e(y|z)g(z)dz$ where $z \in \{0,1\}$, $g(0) + g(1) = 1$, and $g(z) \in [0,1]$ is free of restriction. The choice of e depends on the researchers' needs. How the reparametrization works is as follows:

$$\int p(y|z)f(z)dz = p(y|0)f(0) + p(y|1)f(1)$$
$$= e \times (p(y|0) + p(y|1)) + (1 - 2e)(p(y|0)g(0) + p(y|1)g(1))$$
$$= \int [e \times (p(y|0) + p(y|1)) + (1 - 2e)p(y|z)]g(z)dz$$
$$= \int p^e(y|z)g(z)dz$$

2.4   PRML Filter

We construct the following hypothesis testing, with Poisson as the null hypothesis, Poisson mixtures as the alternative hypothesis.

Consider the spike counts data Y_i for $i = 1, ..., n$,

$H_0 : Y_i \overset{i.i.d}{\sim} Poi(\mu)$

$H_1 : Y_i \overset{i.i.d}{\sim} \int_{\mu_{\mathrm{L}}}^{\mu_{\mathrm{U}}} Poi(\mu)f(\mu)d\mu$

where μ_{L}, μ_{U} is unknown lower bound and upper bound for μ, which can be estimated using robust estimator or specified according to experience.

A Bayes factor is constructed as the ratio of the marginal likelihood based on the models assumption $\mathrm{BF}_{01} = p(Y|M_0)/p(Y|M_1)$. For the Poisson assumption, we can set a prior on $\mu \sim f$ then integrate out the parameter μ to obtain the marginal likelihood $p(Y|M_0) = \int_{\mu_{\mathrm{L}}}^{\mu_{\mathrm{U}}} \prod_{i=1}^{n} p(Y_i|\mu)f(\mu)d\mu$. f can be conjugate prior, Jeffreys' prior or uniform distribution. For Poisson mixture assumption, it is a mixture distribution without extra parameter as shown in equation 2.3. We can apply predictive recursion marginal likelihood (PRML) algorithm (see Algorithm 1) to estimate the marginal

likelihood. Similar as the 'Intermediate' hypothesis, we set $f_0 = \mathrm{Unif}(\mu_\mathrm{L}, \mu_\mathrm{U})$, and calculate $m_i(Y_i)$ as an integral on $(\mu_\mathrm{L}, \mu_\mathrm{U})$.

Chapter 3

Simulation Study

3.1 Comparison between filters

We generate datasets with size $N = 200$ and sample size $n = 25, 50, 100$ for each dataset. Half of each dataset comes from Poisson distribution $\text{Poi}(240)$, while others come from Poisson mixtures (Negative Binomial distribution) $\int_{150}^{300} \text{Poi}(\mu)\text{Ga}_{[150,300]}(\mu|480, 2)d\mu$. We compare PRML filter to traditional Chi-squared test under different sample size $n = 25, 50, 100$. We use the ROC-AUC to evaluate the performance.

For PRML filter, we estimate the bound based on quantiles:

$$\hat{\mu}_{\text{L}} = Y_{0.25} - \alpha \times \text{IQR}, \quad \hat{\mu}_{\text{U}} = Y_{0.75} + \alpha \times \text{IQR}$$

where $\text{IQR} = Y_{0.75} - Y_{0.25}$ means the interquartile range, and $Y_{0.25}, Y_{0.75}$ indicate the 25% and 75% quantiles respectively; the parameter α controls the range of μ. We set $\alpha = 0.5$.

For PRML filter, we test Poisson versus Poisson mixtures. We estimate marginal likelihood under null assumption $p(Y|M_0)$ by setting uniform prior $\text{Unif}[\hat{\mu}_{\text{L}}, \hat{\mu}_{\text{U}}]$. For $p(Y|M_1)$, we apply PRML algorithm with setting $w_i = (1 + i)^{-1}$, $f = \text{Unif}[\hat{\mu}_{\text{L}}, \hat{\mu}_{\text{U}}]$. Then we calculate Bayes factor as a ratio of marginal likelihood $\text{BF}_{01} = p(Y|M_0)/p(Y|M_1)$. For permutation PRML filter, we consider 100 permutations.

For traditional Chi-squared test, we test Poisson versus non-Poisson distribution. We first construct equal probability bins by setting quantiles of the estimated Poisson

distribution as the separator, and make sure the expected count in each bin is at least five. Based on this, a lack-of-fit statistic is calculated as follows:

$$X = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i} \tag{3.1}$$

where O_i, E_i represent observed and expected bin counts respectively, m is the number of bins. We calculate p-value via Monte Carlo simulation. Specifically, we generate 10,000 Monte Carlo samples of Poisson counts from the estimated Poisson distribution with each sample size same as the observed sample size. Calculate lack-of-fit statistics for each sample. P-value is the proportion of these statistics larger than the statistic from the observed data.
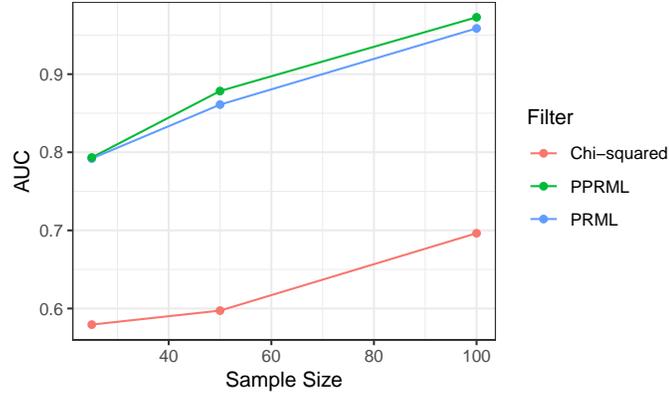
Figure 3.1: Testing Poisson distribution versus Negative Binomial distribution. PPRML and PRML filters are more powerful than Chi-squared testing. The performance of both PPRML and PRML filters improve significant than Chi-squared test does from sample size 25 to 50.

Figure 3.1 shows the AUC of testing on distinguishing Poisson and Poisson mixtures under the different methods and sample sizes. PRML filter and permutation PRML filter perform much better than the traditional testing procedure. Even with a small size, the AUC of PRML and permutation PRML (PPRML) filter can achieve around 0.8. As sample size n increase from 25 to 50, the performances of PRML and

PPRML filters improve significantly especially compared to Chi-squared test. With a large enough sample size 100, the AUC of PRML and permutation PRML filter can achieve over 0.95 while the traditional test can only reach around 0.7. This is because the traditional test can only test Poisson versus non-Poisson distribution, which poses a too general alternative assumption, leading to a loss of statistical power. Our proposed PRML filter specifies the alternative model assumption to Poisson mixtures instead of using a general non-Poisson assumption, which improves sensitivity to the Poisson mixtures. Futhermore, permutation PRML filter performs slightly better than PRML and also more stable due to the "smoothing" effect of averaging. It is worth noting that both the PRML filter and permutation version are not sensitive to the choice of parameter α (see Table 5.1).

We evaluate the performance of PRML filter and Chi-squared testing on distinguishing Poisson from Poisson mixed with other distribution. Explicitly, we generate datasets with size $N = 200$. Half of the dataset come from Poisson distribution Poi(240), while others come from Poisson mixed with normal distribution $0.9\text{Poi}(240) + 0.1N(240, \sigma^2)$ with $\sigma = 5, 10, 15, 20, 25$. All other setting remains the same.

Figure 3.2 shows the AUC of testing on distinguishing Poisson versus Poisson mixed with normal distribution under different methods and different sample size n. With $\sigma < \sqrt{240}(\approx 15.5)$, the data comes from under-dispersion model, whose variance is less than its mean. Otherwise, $\sigma > \sqrt{240}(\approx 15.5)$, the data comes from over-dispersion model. As the Figure 3.2 shows, when $\sigma > 15$, PRML and permutation PRML filter perform better than traditional Chi-squared testing. For $\sigma \leq 15$, the PRML and permutation PRML testing perform even worse than flipping a coin. This is due to the misspecification of alternative hypothesis. Notice Poisson mixture captures over-dispersion on Poisson model. Therefore, under the under-dispersion
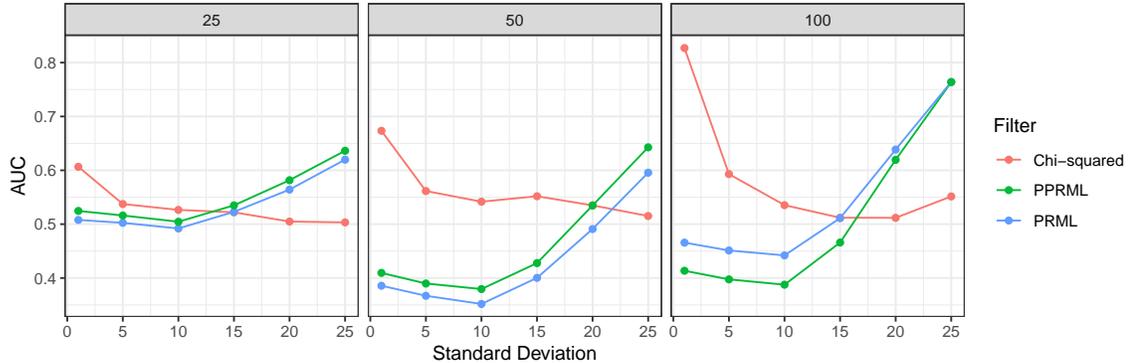
16

Figure 3.2: Testing Poisson versus Poisson mixed with normal distribution. Different panels indicate different sample size $n = 25, 50, 100$. PRML and PPRML filter is more powerful than Chi-squared test when data is over-dispersed (standard deviation equals to 20 or 25), less powerful than Chi-squared test when data is under-dispersed (standard deviation equals to 5, 10 or 15). See Table 5.2 for more details.

model, the likelihood of the alternative model would be no more than the likelihood under the null model assumption, leading to a loss of sensitivity. Moreover, such a problem caused by misspecification cannot be solved by increasing the sample size.

The simulation results suggest that the traditional testing procedure may identify Poisson mixtures as Poisson distribution. Our proposed testing performs much better than the traditional testing procedure and can be used to filter out Poisson mixtures behavior datasets. For the under-dispersion model assumptions, PRML filter and PPRML filter are still favored based on the context. The traditional Chi-squared testing procedure tends to screen out more informative data following under-dispersed Poisson distributions, leading to inefficient use of data.

3.2 Comparison between classifiers

We generate datasets with sample size $n_A = n_B = n_{AB} = 25$, and first generate $Y^A \overset{i.i.d}{\sim} \text{Poi}(50)$, $Y^B \overset{i.i.d}{\sim} \text{Poi}(80)$. For the 'Single' hypothesis, generate Y^{AB} from

Poi(50) and Poi(80) separately. For the 'Mixture' hypothesis, consider $Y^{AB} \overset{i.i.d}{\sim}$ 0.4Poi(50) + 0.6Poi(80). For the 'Intermediate' and 'Outside' hypotheses, we consider Poisson model and Poisson mixtures model with truncated gamma distribution as prior. Specifically, for the 'Intermediate' hypothesis, consider Poi(70) for Poisson model, a Poisson mixtures with light over-dispersion $\int \text{Poi}(\mu)\text{Ga}_{[50,80]}(\mu|700, 10)d\mu$ and a Poisson mixtures with extreme over-dispersion $\int \text{Poi}(\mu)\text{Ga}_{[50,80]}(\mu|70, 1)d\mu$. For the 'Outside' hypothesis, consider Poi(90) for Poisson model, a Poisson mixtures with light over-dispersion $\int \text{Poi}(\mu)\text{Ga}_{[80,180]}(\mu|900, 10)d\mu$ and a Poisson mixtures with extreme over-dispersion $\int \text{Poi}(\mu)\text{Ga}_{[80,180]}(\mu|90, 1)d\mu$. Generate $N = 100$ samples from each model under seed 123.

We are interested in comparing the performance of our PRML classifier and Poisson classifier proposed by Caruso et al. (2018) when the truth comes from the permutation combination of the four scenarios ('Single', 'mixture', 'intermediate' and 'Outside' hypotheses) and different degree of over-dispersion (Poisson, light and extreme over-dispersion). We apply both classifiers to each dataset for each model assumption. We calculate the Bayes factor as a ratio of the posterior probability of data comes from the true model versus the data comes from all other three models, which is the sum of posterior probability under all other three model assumptions.

We consider PRML classifier ($e = 0$) and modified PRML classifier ($e = 0.2$), and average the estimation of marginal likelihood among 100 permutations. For lower bound and upper bound for the 'Outside' hypotheses, we consider the maximum of two standard deviation outside maximum of the triplet $\max_{j=A,B,AB}(\max(Y_j) + 2\text{std}(Y_j))$ and maximum of 0 and the minimum of two standard deviation outside the minimum of the triplet $\min(0, \min_{j=A,B,AB}(\min(Y_j) - 2\text{std}(Y_j)))$ as the upper bound and lower bound respectively. We consider Jeffereys' prior for μ^A, μ^B and use maximum likelihood estimator based on data Y^A, Y^B as the initialization μ_0^A, μ_0^B for the optimization
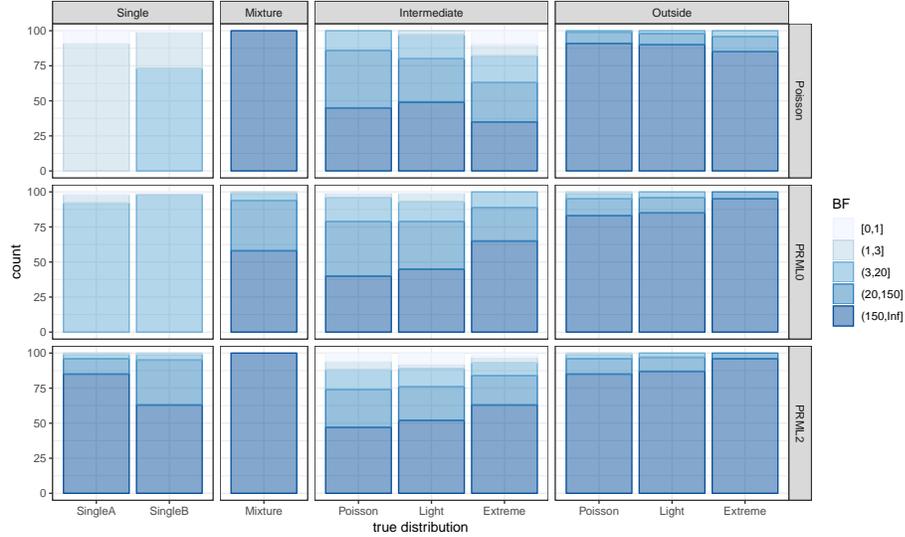
18

Figure 3.3: Summary of Bayes factor under different classifiers and true hypotheses. PRML classifier performs better than Poisson classifier when the truth is 'Single', extreme over-dispersed Poisson under 'Intermediate' and 'Outside' hypotheses. Modified PRML classifier almost outperform Poisson distribution under all hypotheses.

algorithm.

For the Poisson classifier proposed by Caruso et al. (2018), they set truncated Jeffereys' priors on suitable range under different hypotheses for μ^{AB} and integrate out the parameter to obtain the marginal likelihood estimation. For the 'Mixture' hypothesis shown in equation 1.1, they set a uniform prior for the mixing parameter α.

Figure 3.3 shows the proportion of Bayes factor fall into each categories under different testing procedures and different true distributions. Each panel represents one of four different model assumptions. Each bar indicates different models. The color indicates the interval that the Bayes factor falls into. The bar with large proportion of dark color indicates good performance.

PRML classifier performs better than Poisson classifier when the truth is 'Single' model and Poisson mixtures model with extreme over-dispersion. Based on the

PRML classifier, the 'Intermediate' and 'Outside' hypotheses are more complicated than those of Poisson classifier. Therefore, the Bayes factor will add more penalty implicitly to these two hypotheses and favor simple 'Single' hypothesis in PRML classifier. With a clear separation between 'Single', 'Mixture' and 'Intermediate' assumption, the performance of modified PRML classifier($e = 0.2$) can be even better. PRML classifier performs slightly worse than Poisson test when the truth is Poisson 'Intermediate' and 'Outside'. However, such advantage may decline as the sample size increases. Since the Poisson distribution acts as the boundary of Poisson mixtures, as the sample size large enough, the estimation on the mixing density f would have a high peak at single point, leading to results from PRML classifier similar to Poisson classifier. PRML classifier performs worse than Poisson classifier when the truth is mixture model but modified PRML test can perform as well as Poisson test.

These simulation results suggest that our proposed PRML classifier is suitable for the cases where the underlying models may be over-dispersion models. And the flexibility of the Poisson mixtures can make efficient use of the information of the data, leading to a better performance of PRML classifier on identifying the 'Single' hypothesis with strong support. Furthermore, a synthetic separation among the 'Single', 'Mixture' and 'Intermediate' hypotheses is needed to better distinguish these three hypotheses.

Chapter 4

Application to Neuroscience Data

We apply our proposed testing procedures to explore how a single neuron encodes information from two simultaneous stimulus based on the dataset provided by Caruso et al. (2018). Two monkeys were trained to make saccades towards single-sound or dual-sound location. In the dual-sound trials, a sound with 742 Hz frequency and another sound from one of the frequencies (500, 609, 903, 1100 Hz) were displayed at two location (24 degrees and -6 degrees, or -24 degrees and 6 degrees) simultaneously. In the single-sound trials, a single sound is presented with the location and frequency drawn from the same set in dual-sound trials. The trial under different conditions appeared randomly. The single neuron activities from 3 cells in an auditory region (inferior colliculus) are recorded and we analyze the spike trains during the sound displaying period (600ms for cell 1 and 2, 1000ms for cell 3). We obtain 808 triplets $\{Y^A, Y^B, Y^{AB}\}$ where the sample sizes $\{n_A, n_B, n_{AB}\}$ are all at least 5 and Y^A and Y^B are well-separated [1].

4.1 Comparison between different filters

For Chi-squared filter, we keep the triplets with both the p-values of single-trial data Y^A and Y^B greater than 0.1. For PRML filter, we keep the triplets with both the inverse of Bayes factors of single-trial data Y^A and Y^B less than 20, indicating there is no strong evidence showing the data comes from Poisson mixtures instead of

---

[1] Logarithm of the intrinsic Bayes factor is greater or equal to 3 for the hypothesis testing $\mu^A = \mu^B$ versus $\mu^A \neq \mu^B$ (see Caruso et al. (2018))

Poisson distribution. Based on Chi-squared filter and PRML filter, 363 triplets and 352 triplets remain respectively.

We first verify that the PRML filter is more sensitive to Poisson mixtures measured by distribution of fano factors. Fano factor is the ratio of variance to expectation. The fano factor greater than 1 indicates over-dispersion. Large value of fano factor indicates large degree of over-dispersion. Figure 4.1 shows the density function of the fano factors of single-trial spike count data under different filters. The maximum fano factors of the spike count data from single-sound trials remaining in Chi-squared filter but filtered out by PRML filter (the third panel) are all over one and much larger than those remains in double filter and PRML filter. This indicates PRML filter screens out more triplets with over-dispersed single trials.
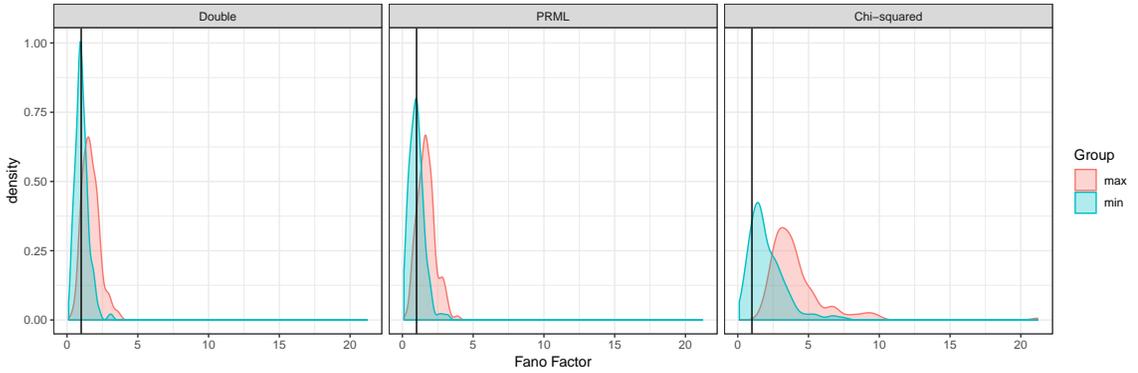

Figure 4.1: Density of fano factors under different filtering procedures. From left to right, different panels indicate the fano factors of the triplets remain in double filtering process, remain in PRML filter but not in Chi-squared filter, and remain in Chi-squared filter but not in PRML filter. Red and blue color indicate the maximum and the minimum value of the fano factors of Y^A and Y^B respectively. Triplets filtered by PRML filter but remaining in Chi-squared filter are more likely to have over-dispersed single trial data.
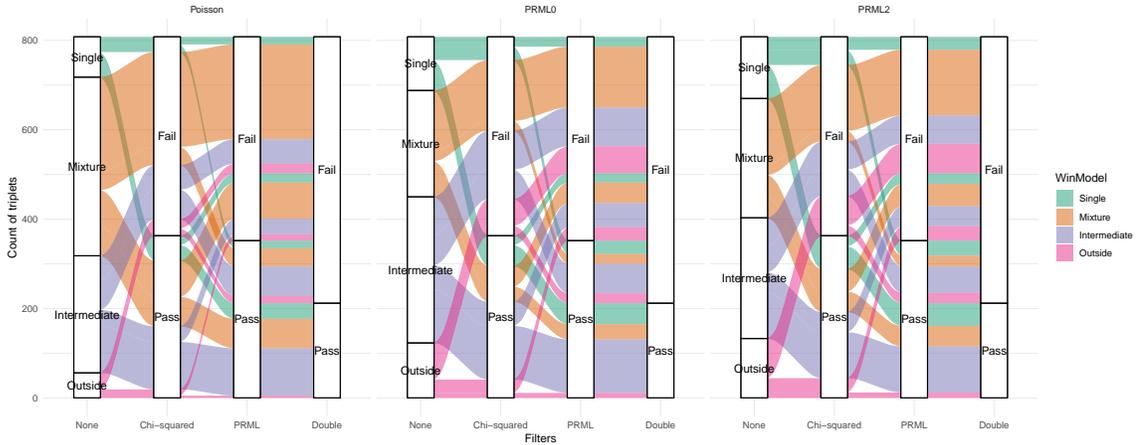
Figure 4.2: Comparison between different filters colored by different classifiers. Different panels indicate different classifiers coloring the flows. The state indicates whether the triplet pass the different filters shown on x-axis, where passing the double filter indicates the triplets pass both Chi-squared and PRML filters. The filtering processes are distinct between Chi-squared filters and PRML filters, and within the four classification.

Figure 4.2 shows the comparison between different filters colored by different classifiers. We see the two filters agree at around 60% triplets, indicating the existence of a clear distinct difference between the characteristics of the triplets passing the two different filters. Within each classification, the proportion of the agreement triplets are different, indicating a correlation exists between the filtering result and classification result. Therefore, the filters play an important role on shaping the result of the classification (see Figure 4.5 also). For example, PRML filter includes more 'Intermediate' and screens out more 'Mixture' behaved triplets compared to Chi-squared filter, which will favor more on 'Intermediate' and less on 'Mixture' on the classification result.

It is worth noting that only a small number of triplets classified as 'Mixture' pass the PRML filter, showing over-dispersion behavior of the corresponding single-trial data of the 'Mixture' triplets. This suggests a second order stochasticity in the dual-trial data may actually come from the trial-to-trial variation from the single-trial
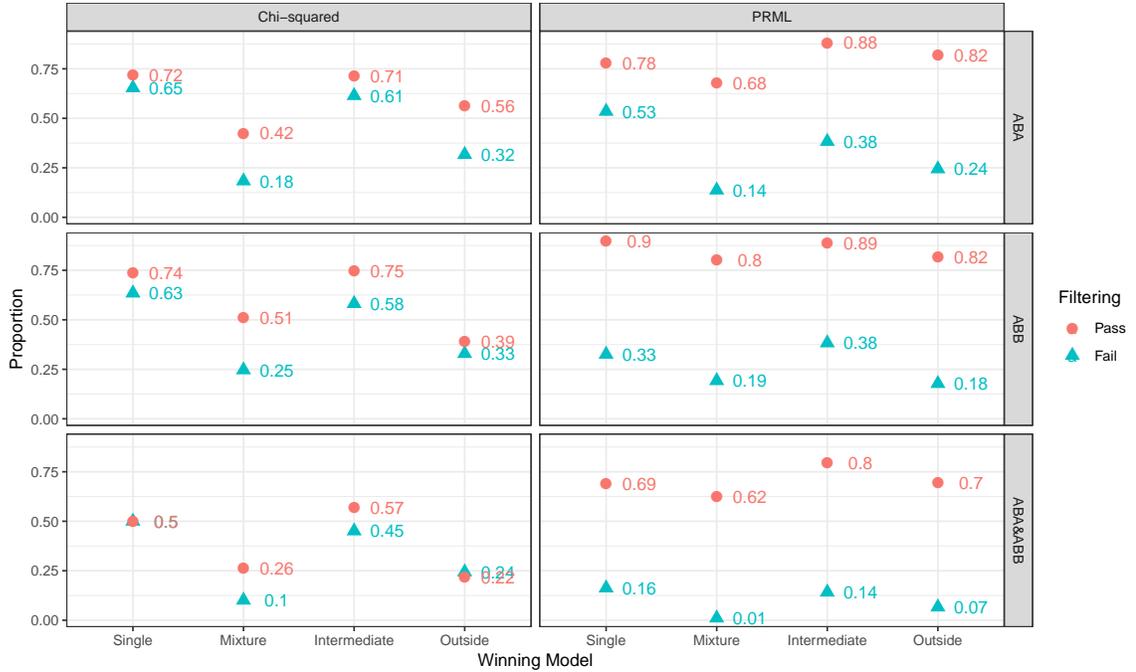
Figure 4.3: Proportion of "valuable" triplets under different filters and conditions. From top to bottom, different panels indicate the proportion of triplets classified as 'Single' in dataset $\{Y^A, Y^B, Y^A\}$, $\{Y^A, Y^B, Y^B\}$ respectively and classified as 'Single' in both datasets. PRML filter improve the proportion of "valuable" triplets significantly.

data.

To evaluate the performance of a filter on screening out single-trial data containing unignorable trial-to-trial variation, we define "valuable" triplets. We create two fake datasets $\{Y^A, Y^B, Y^A\}$ and $\{Y^A, Y^B, Y^B\}$ by setting $Y^{AB} = Y^A$ and $Y^{AB} = Y^B$ respectively. We then apply PRML classifier with $e = 0$ and define the triplets that are classified to the 'Single' hypothesis as "valuable" triplets under condition ABA and ABB. For the triplets that are "valuable" under both condition, we denote them as "valuable" triplets under condition ABA&ABB.

Figure 4.3 shows the proportion of "valuable" triplets within the groups that pass and fail the different filters under different conditions. The proportion of "valuable" triplets within the group passing the PRML filter is significant higher than those

24

within the group failing the PRML filter under all three conditions. And the gap between these two group is much greater than that under the Chi-squared filter. This suggests the PRML filter contributes to ameliorating the performance of the classifier, better than Chi-squared filter does.

4.2   Comparison between different classifiers

Figure 4.4 depicts how the classification of triplets change under different classifiers based on the filtered triplets by the PRML filter. From Poisson classifier to PRML classifier, around 34% of 'Mixture' triplets go to the 'Intermediate' and around 9 % 'Mixture' and 10% 'Single' go to the 'Outside'. This is due to the extension of 'Intermediate' and 'Outside' hypotheses. Around 13% of 'Intermediate' triplets go to the 'Single'. Since Bayes factor will add implicit penalty to complex model, the extension of 'Intermediate' hypothesis will be penalized more compared to Poisson assumption for the increment of parameters. From PRML classifier with $e = 0$ to modified PRML classifier with $e = 0.2$, we expect 'Intermediate' and 'Mixture' flow into 'Single' since the space of 'Intermediate' and 'Mixture' hypothesis is shrinkaged. Also, we observe a relative large flows from 'Intermediate' to 'Mixture', indicating considerable amount of triplets may be Poisson mixtures with a U-shape mixing density (see Figure 5.1 in Appendix).

Figure 4.5 shows the summary of winning model under different classifiers on different datasets filtered by different filters. We consider the probability itself as an indicator for confidence. Three levels of the confidence with thresholds 0.95, 0.5, 0.25 are indicated by the grayscale.

According to the PRML filter and classifier (with $e = 0$), the 'Intermediate' hypothesis have the strongest support then follows 'Mixtures', 'Single' and 'Outside'.
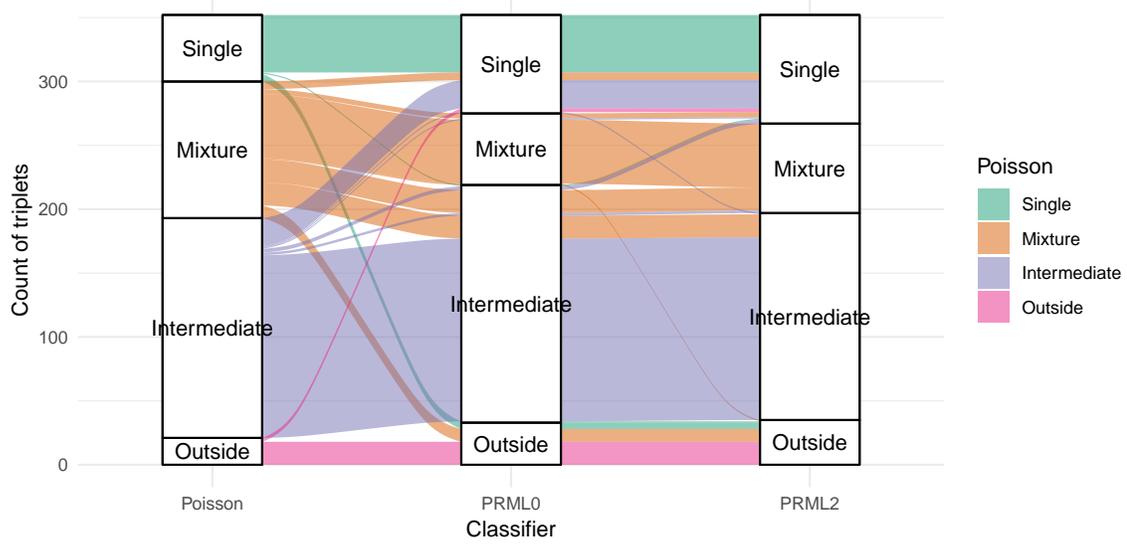
Figure 4.4: Comparison between different classifiers on PRML filtered triplets
The flows are colored by the classification under Poisson classifier. PRML0 classifier
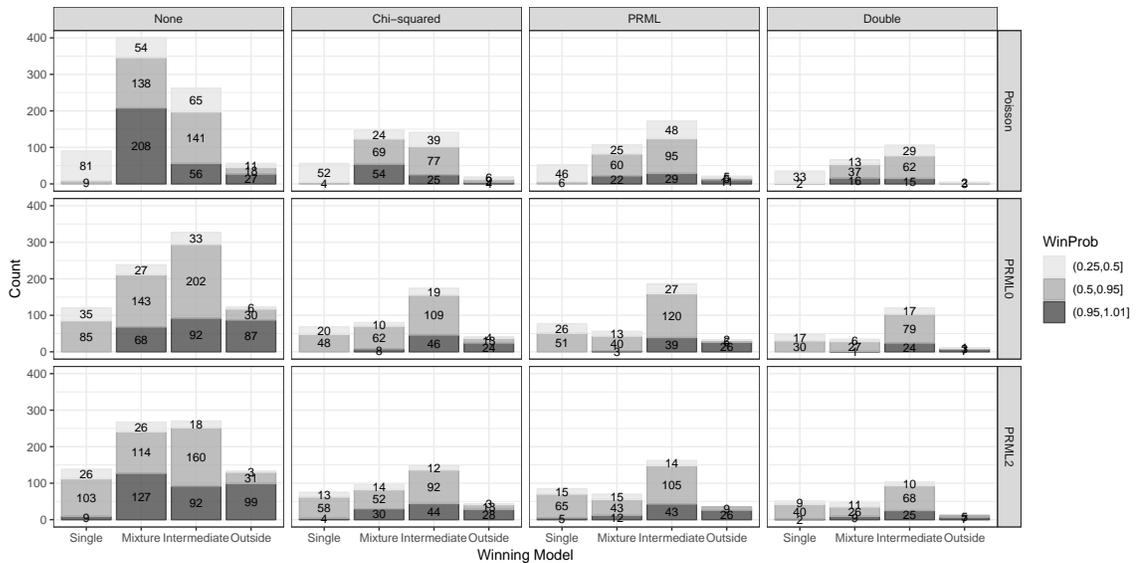favors 'Single' and 'Intermediate' against 'Mixture' compared to Poisson classifier.

Figure 4.5: Comparison among the classification based on different filtered datasets.
'Intermediate' play a dominant role under PRML filter and classifier. In contrast,
'Mixture' and closely followed by 'Intermediate' have the strongest support under
the old filter and classifier proposed by Caruso et al. (2018).

More 'Mixture' and 'Single' hypotheses are favored if we set $e = 0.2$ due to the shrinkage on the space of 'Intermediate' and 'Mixture' hypotheses. In contrast, for the result based on the Chi-squared filter and Poisson classifier, 'Mixture' is the most common model then closely followed by 'Intermediate', then followed by 'Single' with relative small confidence and then 'Outside'. The difference is due to the joint effect of the filter and the classifier mentioned previously. It is worth noting that PRML classifier is more robust, compared to the Poisson classifier, for the histogram remains similar shape across different filters. Similarly, we can see PRML filter is more robust than Chi-squared filter.

Chapter 5

Conclusion

We have presented a new testing procedure for identifying different types of stochastic order of mixture distributions. The Bayes factor provides an efficient framework, and the PRML algorithm enables marginal likelihood estimation of mixture distributions. For mixture density with extra unknown parameters, we can apply Laplace approximation together with any optimization algorithm, and adopt developed PRML gradient algorithm to provide gradient to facilitate the calculation.

Our emphasis has been on testing between different types of Poisson mixtures in the context of neuroscience. For better performance of the new classifier, we proposed a filter to test Poisson distribution versus Poisson mixtures to screen out over-dispersed data.

Based on the context, new proposed filter is favored without any shortcomings. Although PRML filter is less sensitive to under-dispersed data, the under-dispersed data provide more information and will benefit to the afterward analysis. However, when the alternative hypothesis is unknown, the traditional testing can always construct a test statistics and calculate a p-value to make a judgement. In contrast, we cannot calculate the Bayes factor without model assumptions. In this case, using the traditional testing procedure may be a better choice, though the meaning of the p-value here remains debatable (Berger and Delampady, 1987).

Our new proposed classifier extend the 'Intermediate' and 'Outside' hypotheses to capture second order stochasticity. However, the extended framework does not pose a clear separation between 'Single', 'Mixture' and 'Intermediate' hypotheses.

An extra hyper-parameter e needs to be specified by researchers for the 'Mixture' and 'Intermediate' hypotheses. The choice of the e still remains unclear and need further exploration. Similar problems appear when the true distribution is 'Intermediate' or 'Outside' Poisson-like distribution. A well-specified range of the support of mixing density is needed. And a well-specified e also contributes to boost the winning probability implicitly by suppressing the posterior probability of 'Mixture' and 'Intermediate' hypotheses. For the relative small sample size, the hyper-parameter settings for e and range of interval play a more important role.

It is worth noting that PRML algorithm provides not only estimation of the marginal likelihood but also the density estimation of the mixing densities, which will help us explore more on the encoding process. For example, Figure 5.1 in Appendix shows the typical U-shape for density estimation of triplets which are classified as 'Mixture', 'Intermediate', 'Mixture' by Poisson classifier, PRML classifier and modified PRML classifier ($e = 0.2$) respectively. We can make use of the density information to explore the homogeneity and heterogeneity across trials, which may unveil more information beyond the four hypotheses.

A future step is to relax the assumption on single-trial data. We assume Poisson distribution for single-trial data and apply PRML filter to screen out over 50% triplets. An extension on single-trial to Poisson mixtures may contribute to make more efficient use of the data. However, under such extension, new hypotheses needs to be defined. More generally, extend the model assumption for single-trial data and dual-trial data to the general stochastic order may be the future direction.

Appendix

Gaussian quadrature

For the integral calculation, we approximate the integral $\int_{\Theta} \cdot d\theta$ with Gaussian quadrature. The basic idea is to approximate integral with a weighted sum of function values at specified points within the domain of integration. Here we set the points and weights according to Legendre polynomials proposed by Abramowitz and Stegun (1965).

$$\int_{-1}^{1} f(x)\, dx = \sum_{i=1}^{n} w_i f(x_i).$$

And change the interval from $[-1, 1]$ to $[a, b]$ according to

$$\int_{a}^{b} f(x)\, dx = \frac{b-a}{2} \int_{-1}^{1} f\left(\frac{b-a}{2}x + \frac{a+b}{2}\right) dx.$$

, which is

$$\int_{a}^{b} f(x)\, dx \approx \frac{b-a}{2} \sum_{i=1}^{n} w_i f\left(\frac{b-a}{2}x_i + \frac{a+b}{2}\right).$$

For two-dimension Gaussian quadrature, the calculation is similar. Here we also need to tranform the interval from $[-1, 1] \times [-1, 1]$ to $[a, b] \times [c, d]$

$$\int_{-1}^{1} \int_{-1}^{1} f(\xi, \eta) d\xi d\eta \approx \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j f(\xi_i, \eta_j)$$

Laplace approximation

According to Laplace approximation, approximate the posterior distribution with normal distribution.

$$\pi(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{p(y)} \approx N(\theta|\hat{\theta}, \hat{\Sigma})$$

where

$$\hat{\theta} = \arg\max_{\theta} \log \pi(\theta|y) = \arg\max_{\theta} \log p(y|\theta)\pi(\theta)$$

$$\hat{\Sigma} = \{-\nabla^2 \log \pi(\theta|y)|_{\theta=\hat{\theta}}\}^{-1} = \{-\nabla^2 \log p(y|\theta)\pi(\theta)|_{\theta=\hat{\theta}}\}^{-1} = \{-H\}^{-1}$$

So we have the estimation for marginal likelihood valued at $\hat{\theta}$:

$$p(y) \approx (2\pi)^{k/2}|\hat{\Sigma}|^{1/2}p(y|\theta=\hat{\theta})\pi(\theta=\hat{\theta})$$

where $k = dim(\theta)$, H is Hessian matrix. So approximation on marginal likelihood is transformed to an optimization problem:

$$p(y) \approx (2\pi)^{k/2}|H|^{-1/2}e^{l(\hat{\theta})}$$

where $\hat{\theta} = \underset{\theta}{argmax} l(\theta)$, H is corresponding Hessian matrix.

Tables and figures

Table **5.1**: AUC of testing on Poisson versus Poisson mixture (Negative Binomial)

Method	α	Sample size		
		25	50	100
Chi-squared		0.58	0.60	0.70
PPRML	0.4	0.80	0.88	0.97
	0.5	0.79	0.88	0.97
	0.8	0.79	0.87	0.97
	1	0.79	0.87	0.97
PRML	0.4	0.79	0.87	0.96
	0.5	0.79	0.86	0.96
	0.8	0.78	0.85	0.95
	1	0.78	0.84	0.94

Table 5.2: AUC of testing on Poisson versus Poisson mixed with normal distribution

| (Method)	Sample size(std)																	
α	25						50						100					
	1	5	10	15	20	25	1	5	10	15	20	25	1	5	10	15	20	25
(Chisq)	0.61	0.54	0.53	0.52	0.51	0.51	0.67	0.56	0.54	0.55	0.54	0.52	0.82	0.58	0.48	0.49	0.51	0.54
(PPRML)																		
0.4	0.50	0.50	0.50	0.53	0.58	0.63	0.60	0.62	0.62	0.57	0.53	0.64	0.59	0.61	0.62	0.54	0.62	0.76
0.5	0.48	0.48	0.50	0.53	0.58	0.64	0.59	0.61	0.62	0.57	0.53	0.64	0.59	0.60	0.61	0.53	0.62	0.76
0.8	0.56	0.54	0.48	0.54	0.58	0.64	0.57	0.60	0.61	0.57	0.54	0.65	0.59	0.61	0.61	0.54	0.62	0.77
1	0.56	0.55	0.48	0.54	0.58	0.64	0.58	0.60	0.62	0.57	0.54	0.65	0.60	0.62	0.61	0.53	0.63	0.77
(PRML)																		
0.4	0.51	0.51	0.51	0.52	0.56	0.62	0.62	0.64	0.65	0.60	0.49	0.60	0.55	0.56	0.57	0.49	0.63	0.76
0.5	0.49	0.50	0.51	0.52	0.56	0.62	0.61	0.63	0.65	0.60	0.51	0.60	0.53	0.55	0.56	0.49	0.64	0.76
0.8	0.54	0.53	0.49	0.53	0.57	0.63	0.61	0.63	0.65	0.61	0.52	0.58	0.51	0.53	0.54	0.53	0.65	0.77
1	0.55	0.54	0.49	0.53	0.57	0.63	0.61	0.64	0.66	0.62	0.53	0.58	0.51	0.52	0.53	0.54	0.66	0.77
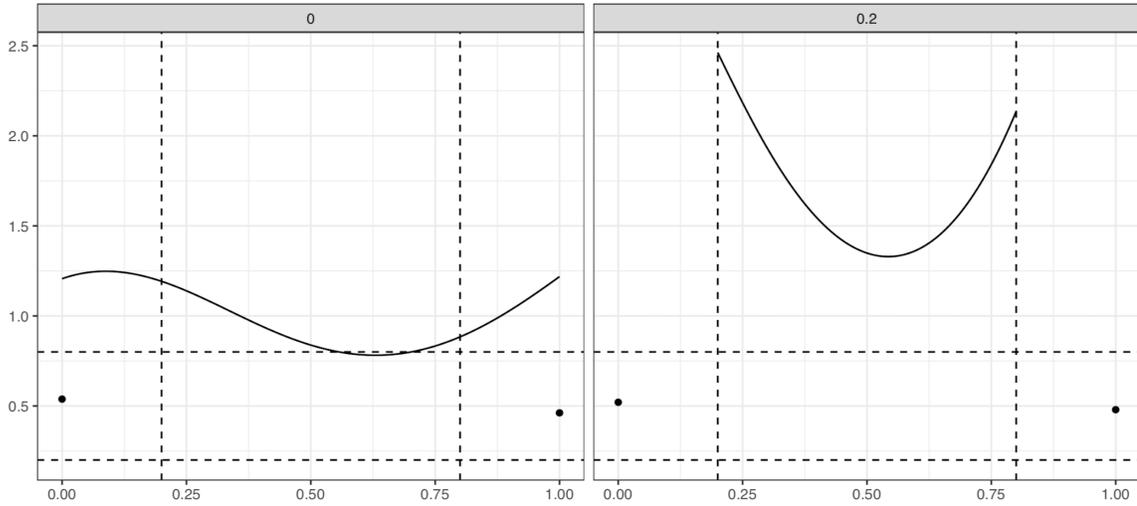
Figure 5.1: Density estimation of mixing density $\tilde{f}(z)$ of a triplet
The triplet is classified as 'Intermediate' by PRML ($e = 0$) classifier but classified as 'Mixture' by PRML ($e = 0.2$) classifier. The points represent the probability estimation under 'Mixture' hypothesis. Both panels present a U-shape.

Bibliography

Abramowitz, M. and I. A. Stegun (1965). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Volume 55. Courier Corporation.

Berger, J. O. and M. Delampady (1987). Testing precise hypotheses. *Statistical Science*, 317–335.

Caruso, V. C., J. T. Mohl, C. Glynn, J. Lee, S. M. Willett, A. Zaman, A. F. Ebihara, R. Estrada, W. A. Freiwald, S. T. Tokdar, et al. (2018). Single neurons may encode simultaneous stimuli by switching between activity patterns. *Nature communications 9*(1), 2715.

Glynn, C. D., S. T. Tokdar, A. Zaman, V. C. Caruso, J. T. Mohl, S. M. Willett, and J. M. Groh (2019). Analyzing second order stochasticity of neural spiking under stimuli-bundle exposure. Submitted.

Kass, R. E., V. Ventura, and E. N. Brown (2005). Statistical issues in the analysis of neuronal data. *Journal of neurophysiology 94*(1), 8–25.

Martin, R. and S. T. Tokdar (2011). Semiparametric inference in mixture models with predictive recursion marginal likelihood. *Biometrika 98*(3), 567–582.

Newton, M. A. (2002). On a nonparametric recursive estimator of the mixing distribution. *Sankhyā: The Indian Journal of Statistics, Series A*, 306–322.

Newton, M. A., F. A. Quintana, and Y. Zhang (1998). Nonparametric bayes methods using predictive updating. In *Practical nonparametric and semiparametric Bayesian statistics*, pp. 45–61. Springer.

Tokdar, S. T., R. Martin, J. K. Ghosh, et al. (2009). Consistency of a recursive estimate of mixing distributions. *The Annals of Statistics 37*(5A), 2502–2522.

Ventura, V., R. Carta, R. E. Kass, S. N. Gettner, and C. R. Olson (2002). Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics 3*(1), 1–20.