A simple two parameter distribution for modelling neuronal activity and capturing neuronal association

Thomas Delaney[1] and Cian O'Donnell[1]

[1]*School of Computer Science, Electrical and Electronic Engineering, and Engineering Mathematics, University of Bristol, Bristol, United Kingdom.*

Abstract

Recent developments in electrophysiological technology have lead to an increase in the size of electrophysiological datasets. Consequently, there is a requirement for new analysis techniques that can make use of these new datasets, while remaining easy to use in practice. In this work, we fit the Conway-Maxwell-binomial distribution to spiking data read from a mouse exposed to visual stimuli.

1 Introduction

Motivate by pointing out how much computational power it can require to calculate nth order correlations.

Point out that we don't necessarily need to measure correlations anyway.

2 Results

3 Discussion

4 Data

We used data collected by Nick Steinmetz and his lab 'CortexLab at UCL' [9]. The data can be found online [1] and are free to use for research purposes.

Two 'Phase3' Neuropixels [4] electrode arrays were inserted into the brain of an awake, head-fixed mouse for about an hour and a half. These electrode arrays recorded 384 channels of neural data each at 30kHz and less than 7μV RMS noise levels. The sites are densely spaced in a 'continuous tetrode'-like arrangement, and a whole array records from a 3.8mm span of the brain. One array recorded from visual cortex, hippocampus, and thalamus, the other array recorded from motor cortex and striatum. The data were spike-sorted automatically by Kilosort and manually by N. Steinmetz using Phy. In total 831 well-isolated individual neurons were identified.

4.1 Experimental protocol

The mouse was shown a visual stimulus on three monitors placed around the mouse at right angles to each other, covering about ± 135 degrees azimuth and ± 35 degrees elevation.

The stimulus consisted of sine-wave modulated full-field drifting gratings of 16 drift directions $(0°, 22.5°, \ldots, 337.5°)$ with 2Hz temporal frequency and 0.08 cycles/degree spatial frequency displayed for 2 seconds plus a blank condition. Each of these 17 conditions were presented 10 times in a random order across 170 different trials. There were therefore 160 trials with a visual stimulus present, and 10 trials without a visual stimulus.

5 Methods

Details about all kinds of things here.

[1] `http://data.cortexlab.net/dualPhase3/`

5.1 Binning data

We converted the spike times for each cell into spike counts by putting the spike times into time bins of a given 'width' (in seconds). We used time bins of 1ms, 5ms, and 10ms. We used different time bin widths to assess the impact of choosing a bin width.

5.2 Number of *active* neurons

To count the number of active neurons in each neuronal ensemble, we split the time interval for each trial into bins of a given width. We counted the number of spikes fired by each cell in each bin. If a cell fired *at least* one spike in a given bin, we regarded that cell as active in that bin. We recorded the number of active cells in every bin, and we recorded each cell's individual spike counts.

It should be noted that when we used a bin width of 1ms, the maximum number of spikes in any bin was 1. For the wider time bins, some bins had spike counts greater than 1. Consequently when using a bin width of 1ms, the number of active neurons and the total spike count of a given bin were identical. But for wider bin widths, the total spike count was greater than the number of active neurons.

So for the 1ms bin width, the activity of a neuron and the number of spikes fired by that neuron in any bin can be modelled as a Bernoulli variable. But for wider time bins, only the activity can be modelled in this way.

5.3 Moving windows for measurements

When taking measurements (e.g. moving average over the number of active neurons) or fitting distributions (eg. the beta binomial distribution) we slid a window containing a certain number of bins across the data, and made our measurements at each window position. For example, when analysing 1ms bin data, we used a window containing 100 bins, and we slid the window across the time interval for each trial moving 10 bins at a time. So that for 2560ms of data, we made 246 measurements.

For the 5ms bin width data, we used windows containing 40 bins, and slid the window 2 bins at a time when taking measurements.

For the 10ms bin width data, we used windows containing 40 bins, and slid the window 1 bin at a time when taking measurements.

By continuing to use windows containing 40 bins, we retained statistical power but sacrificed the number of measurements taken.

5.4 Fano factor

The *Fano factor* of a random variable is defined as the ratio of the variable's variance to its mean.

$$F = \frac{\sigma^2}{\mu} \tag{1}$$

We measured the Fano factor of the spike count of a given cell by measuring the mean and variance of the spike count across trials, and taking the ratio of those two quantities. When calculated in this way the Fano factor can be used as a measure of neural variability. This is similar to the calculation used in [2].

5.5 Probability Distributions suitable for modelling ensemble activty

We present here three different probability distributions that could be suitable to model the number of active neurons in an ensemble. Each distribution has the set $\{0, \ldots, n\}$ as its support, where n is the number of neurons in the ensemble. These are simple distributions with either two or three parameters each. However, we regard n as known when using these distributions for modelling, so in effect each distribution has either one or two free parameters.

5.5.1 Association

Association between random variables is similar to the correlation between random variables but is more general in concept. The correlation is a measure of association; and association doesn't have a mathematical definition like correlation does. Essentially, the association between two random variables is their tendency to take the same or similar values. Positively associated variables tend to take the same value, and negatively associated variables tend to take differnt values. In this research, we work with probability distributions of the number of successes in a set of Bernoulli trials. These Bernoulli variables may or may not be associated.

A probability distribution over the number of successes in n Bernoulli trials, where the Bernoulli variables may be associated, could constitute a good model for the number of active neurons in an ensemble of n neurons.

5.5.2 Binomial distribution

The binomial distribution is a two parameter discrete probability distribution that can be thought of as a probability distribution the number of successes from n independent Bernoulli trials, each with the same probability of success. The parameters of the binomial distribution are n, and $0 \leq p \leq 1$, the probability of success for each of these trials. A random variable with the binomial distribution can take values from $\{0, \ldots, n\}$. The probability mass function of the distribution is

$$P(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \tag{2}$$

If we have N independent samples from a binomial distribution and we know n but want to estimate p, we can maximise the log likelihood function

$$\begin{aligned} L(p) &= \log P(\{k_1, \ldots, k_N\}; n, p) && (3) \\ &= \sum_{i=1}^{N} \log \binom{n}{k_i} + k_i \log p + (n - k_i) \log(1-p) && (4) \end{aligned}$$

If we do not know n there is no closed form way way of maximising this equation. Therefore the binomial distribution is generally only used in cases where we do know n. Consequently, the distribution is practically a one parameter distribution.

As model for the activity of a neuronal ensemble, the main problem with the binomial distribution is that it treats each neuron, represented as a Bernoulli trial, as independent. It is well know that neurons are not independent, and that correlated behaviour between neurons is vital for representing sensory information. The binomial distribution falls short in this regard, but it is useful as performance benchmark when assessing the performance of other models.

5.5.3 Beta-binomial distribution

The beta distribution is the conjugate distribution of the binomial distribution. The beta-binomial distribution is the combination of the beta distribution and the binomial distribution, in that the probability of success for the binomial distribution is sampled from the beta distribution. This allows the beta-binomial distribution to capture some over dispersion relative to the binomial distribution.

The beta-binomial distribution is a three parameter distribution, n the number of Bernoulli trials, and $\alpha \in \mathbb{R}_{>0}$ and $\beta \in \mathbb{R}_{>0}$ the shape parameters of the beta distribution. The probability mass function for the beta-binomial distribution is

$$P(k; n, \alpha, \beta) = \binom{n}{k} \frac{B(k + \alpha, n - k + \beta)}{B(\alpha, \beta)} \tag{5}$$

where $B(\alpha, \beta)$ is the beta function.

This probability distribution can be reparametrised in a number of ways. One of which defines new parameters π and ρ by

$$\pi = \frac{\alpha}{\alpha + \beta} \tag{6}$$

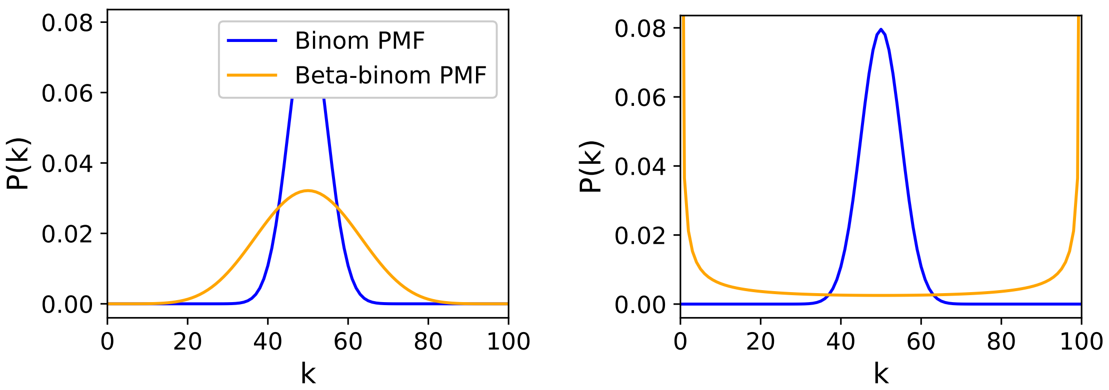
$$\rho = \frac{1}{\alpha + \beta + 1} \tag{7}$$

This reparametrisation is useful because π acts as a location parameter analogous to the p parameter of a binomial distribution. A value of $\rho > 0$ indicates over-dispersion relative to a binomial distribution.

As a model for the activity of a neuronal ensemble, the beta-binomial distribution is more suitable than a binomial distribution because the over-dispersion of the beta-binomial distribution can be used to model positive association between the neurons. An extreme example of this over-dispersion/positive association can be seen in figure 1b. In this figure, the neurons are positively associated and so tend to take the same value, consequently the probability mass of the beta-binomial distribution builds up close to $k = 0$ and $k = n$. It is worth noting that the location parameter for each distribution has the same value, $p = \pi = 0.5$.

5.5.4 Conway-Maxwell-binomial distribution

The Conway-Maxwell-binomial distribution (COMb distribution) is a three parameter generalisation of the binomial distribution that allows for over dispersion and under dispersion relative to the binomial distribution. The parameters of the distribution are n the number of Bernoulli trials, $0 \leq p \leq 1$, the location parameter, and



(a) $n = 100, p = 0.5, \alpha = \beta = 10$ (b) $n = 100, p = 0.5, \alpha = \beta = 0.3$

Figure 1: Figures showing the over-dispersion possible for a beta-binomial distribution relative to a binomial distribution. Parameters are shown in the captions.

$\nu \in \mathbb{R}$ the shape parameter.

The probability mass function of the COMb distribution is

$$P(k; n, p, \nu) = \frac{1}{S(n, p, \nu)} \binom{n}{k}^{\nu} p^k (1 - p)^{n-k} \tag{8}$$

where

$$S(n, p, \nu) = \sum_{j=0}^{n} \binom{n}{k}^{\nu} p^j (1 - p)^{n-j} \tag{9}$$

The only difference between this PMF and the PMF for the standard binomial is the introduction of ν and the consequent introduction of the normalising function $S(n, p, \nu)$.

Indeed, if $\nu = 1$ the COMb distribution is identical to the binomial distribution with the same values for n and p.

If $\nu < 1$ the COMb distribution will exhibit over-dispersion relative to the binomial distribution. If $p = 0.5$ and $\nu = 0$ the COMb distribution is the discrete uniform distribution, and if $\nu < 0$ the mass of the COMb distribution will tend to build up near $k = 0$ and $k = n$. This over-dispersion represents positive association in the Bernoulli variables.

If $\nu > 1$ the COMb distribution will exhibit under-dispersion relative to the binomial distribution. The larger the value of ν the more probability mass will build up at $n/2$ for even n, or at $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ for odd n. This under-dispersion represents negative association in the Bernoulli variables.

ν	**Relative dispersion**	**Associaton between neurons/variables**
< 1	over	positive
0	none	none
> 1	under	negative

Since the COMb distribution has the potential to capture positive and negative associatons between the neurons/Bernoulli variables, it should be an excellent candidate for modelling the number of active neurons in a neuronal ensemble.

5.6 Goodness-of-fit

5.7 Spike count correlations

6 Discussion

Point out that the Conway-Maxwell-binomial distribution could be used to measure activity and association without having to sort the voltage traces into spikes. That does defeat the purpose slightly, however.

References

[1] Patricia M. E. Altham, *Two Generalizations of the Binomial Distribution*. Journal of the Royal Statistical Society 27, 162-167, (1978)

[2] Mark M Churchland, Byron M Yu, John P Cunningham, Leo P Sugrue, Marlene R Cohen, Greg S Corrado, William T Newsome, Andrew M Clark, Paymon Hosseini, Benjamin B Scott, David C Bradley, Matthew A Smith, Adam Kohn, J Anthony Movshon, Katherine M Armstrong, Tirin Moore, Steve W Chang, Lawrence H Snyder, Stephen G Lisberger, Nicholas J Priebe, Ian M Finn, David Ferster, Stephen I Ryu, Gopal Santhanam, Maneesh Sahani, Krishna V Shenoy, *Stimulus onset quenches neural variability: A widespread cortical phenomenon.* Nature Neuroscience 13, 369-378, (2010)

[3] Fraser Daly, Robert E. Gaunt, *The Conway-Maxwell-Poisson distribution: Distributional theory and approximation.* ALEA 13, 635-658, (2016)

[4] James J. Jun, Nicholas A. Steinmetz, Joshua H. Siegle, Daniel J. Denman 5, Marius Bauza, Brian Barbarits, Albert K. Lee, Costas A. Anastassiou, Alexandru Andrei, Çağatay Aydın, Mladen Barbic, Timothy J. Blanche, Vincent Bonin, João Couto, Barundeb Dutta, Sergey L. Gratiy, Diego A. Gutnisky, Michael Häusser, Bill Karsh, Peter Ledochowitsch, Carolina Mora Lopez, Catalin Mitelut, Silke Musa, Michael Okun, Marius Pachitariu, Jan Putzeys, P. Dylan Rich, Cyrille Rossant, Wei-lung Sun, Karel Svoboda, Matteo Carandini, Kenneth D. Harris, Christof Koch, John O'Keefe, Timothy D. Harris, *Fully integrated silicon probes for high-density recording of neural activity.* Nature 551, 232-236, (2017)

[5] Joseph B. Kadane, *Sums of Possibly Associated Bernoulli Variables: The Conway–Maxwell-Binomial Distribution.* Bayesian Analysis 11, 403-420, (2016)

[6] Joseph B. Kadane, Galit Shmueli, Thomas P. Minka, Sharad Borle, Peter Boatwright, *Conjugate Analysis of the Conway-Maxwell-Poisson Distribution.* Bayesian Analysis 1, (2006)

[7] R. L. Prentice, *Binary Regression Using an Extended Beta-Binomial Distribution, With Discussion of Correlation Induced by Covariate Measurement Errors.* Journal of American Statistical Association 81, 321-327, (1986)

[8] Galit Shmueli, Thomas P. Minka, Joseph B. Kadane, Sharad Borle, Peter Boatwright, *A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution.* Applied Statistics 54, 127-142, (2005)

[9] Nick Steinmetz, Matteo Carandini, Kenneth D. Harris, *"Single Phase3" and "Dual Phase3" Neuropixels Datasets.* figshare, Dataset: https://doi.org/10.6084/m9.figshare.7666892.v2 (2019)

[10] Joseph S. Verducci, Michael E. Mack, Morris H. DeGroot, *Estimating multiple rater agreement for a rare diagnosis.* Journal of Multivariate Analysis 27, 512-535, (1988)