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 some one or two parameter probability distributions to spiking data collected from a mouse exposed to visual stimuli. We show that a little known probability distribution entitled the Conway-Maxwell-binomial is a suitable model for the number of active neurons in a neuronal ensemble at any given moment. This distribution fits better than binomial or beta-binomial distributions. It also captures the correlated activity in the primary visual cortex induced by stimulus onset more effectively than simply measuring the correlations, at short timescales.

1 Introduction

Recent advances in electrophysiological technology, such as 'Neuropixels' probes[7] has allowed extracellular voltage measurements to be collected from multiple brain regions simultaneously routinely, and in larger numbers than traditional methods.

These larger datasets require innovative methods to extract information from the data in a reasonable amount of time, 'reasonable' being subjective in this case.

Theoretically, all the information in an electrophysiological dataset with n neurons could be captured by calculating the probability distribution for every possible spiking pattern. This would require defining a random variable with 2^n possible values, a task that quickly becomes impossible as n increases. Attempts at approximating this random variable often involve measuring pairwise or higher order correlations [12, 5, 6]. But pairwise correlations may not be enough to characterise neural activity[14].

Measuring higher order correlations becomes computationally impractical quite quickly also (the number of 'three neuron correlations' to measure scales with $\binom{n}{3}$). In this paper, we dispense with measuring correlations directly, and attempt to characterise correlated behaviour by measuring 'association'.

In this work, we examined the ability of very simple distributions to model the number of active (spiking) neurons in a neuronal ensemble at any given timepoint. We compared a little-known distribution named the Conway-Maxwell-binomial distribution to the binomial distribution and the beta-binomial distribution. The Conway-Maxwell-binomial distribution is a probability distribution of the number of successes in a series of Bernoulli trials, but allows over- and under-dispersion relative to the standard binomial distribution. This distribution should therefore be a good candidate for our purposes. We found that Conway-Maxwell-binomial distribution was usually the best candidate of the three that we examined.

2 Results

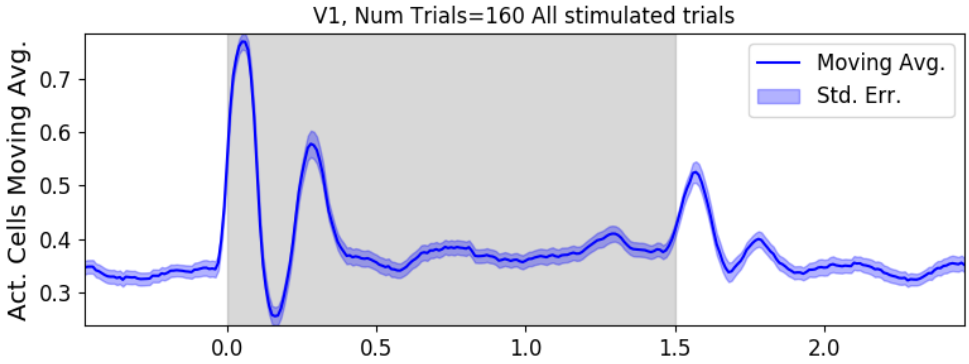
We defined a neuron as *active* in a time bin if it fires at least one spike during the time interval covered by that bin. We measured the number of active neurons in the primary visual cortex of a mouse in 1ms bins across 160 trials of a moving bar visual stimulus. We then slid a 100ms window across these 1ms bins taking mesurements, and fitting distributions along the way. We did the same for neurons in the thalamus, hippocampus, striatum, and motor cortex. We repeated the analysis for 5ms time bins with 40 bin windows, and 10ms time bins with 40 bin windows.

2.1 Increases in mean number of active neurons and variance in number of active neurons at stimulus onset in some regions

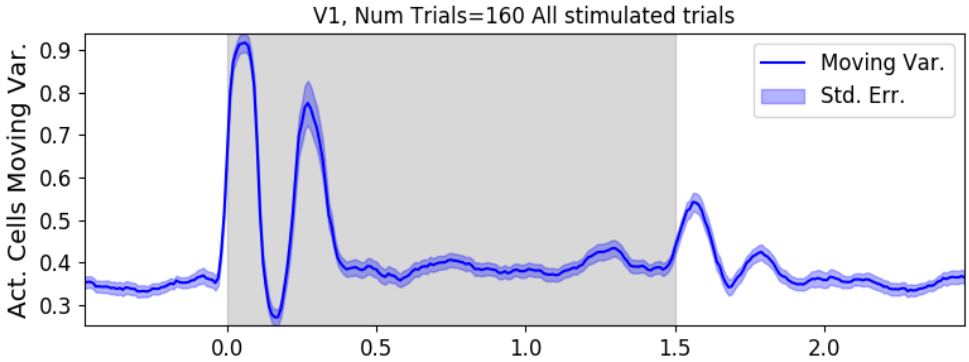
We measured the averge number of active neurons, and the variance of the number of active neurons in a 100ms sliding window starting 500ms before stimulus onset until 1000ms after stimulus onset. We found differences in the response across regions. There were no observed changes in response to the stimulus in the motor cortex or the striatum. The changes in the other regions are detailed below.

2.1.1 Primary visual cortex

We found a transient increase in both the average and variance of the number of active neurons at stimulus onset, followed by a fall to pre-stimulus levels, followed by another transient increase (see figure 1). The oscillation in both of these measurements appear to reflect the frequency of the stimulus.



(a) Moving average.

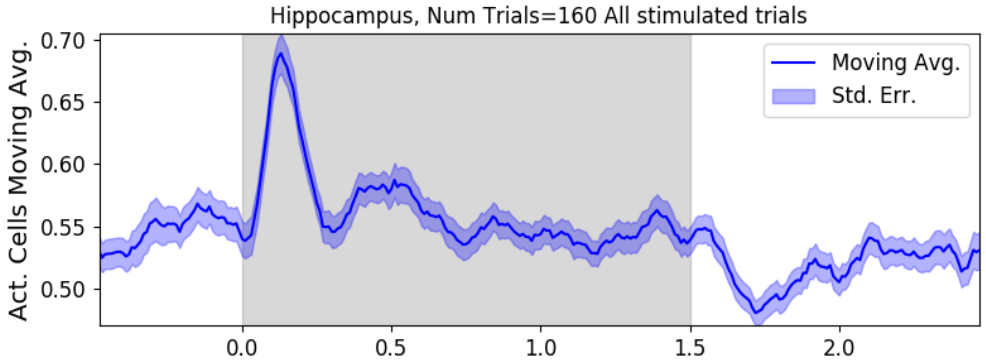


(b) Moving variance.

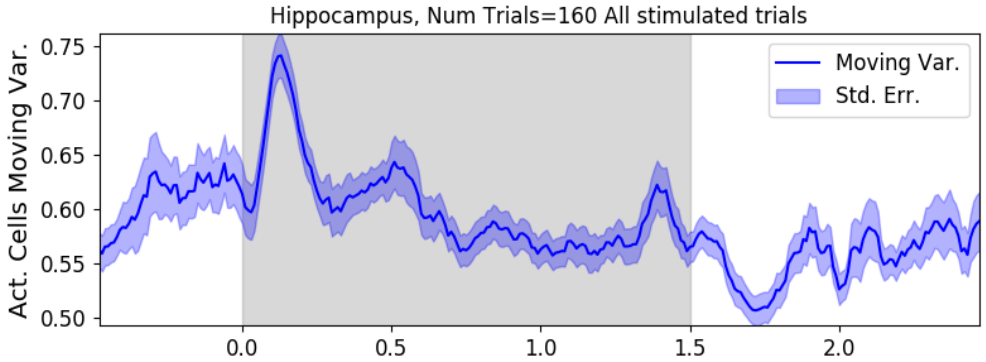
Figure 1: The (a) average and (b) variance of the number of active neurons, measured using a sliding window 100ms wide, split into 100 bins. The midpoint of the time interval for each window is used as the timepoint (x-axis point) for the measurements using that window. The grey shaded area indicates the presence of a visual stimulus. The opaque line is an average across the 160 trials that included a visual stimulus of any kind. We can see a transient increase in the average number of active neurons and the variance of this number, followed by a fluctuation and another increase.

2.1.2 Hippocampus

In the hippocampus we observed a transient increase in the mean number of active neurons and in the variance of the number of active neurons at stimulus onset (see figure 2). The increase lasted about 125ms, and the subsequent fall to baseline took the a similar amount of time.



(a) Moving average.

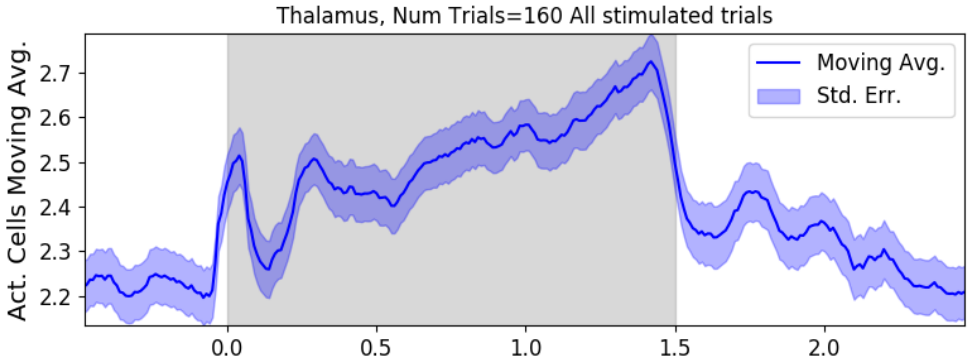


(b) Moving variance.

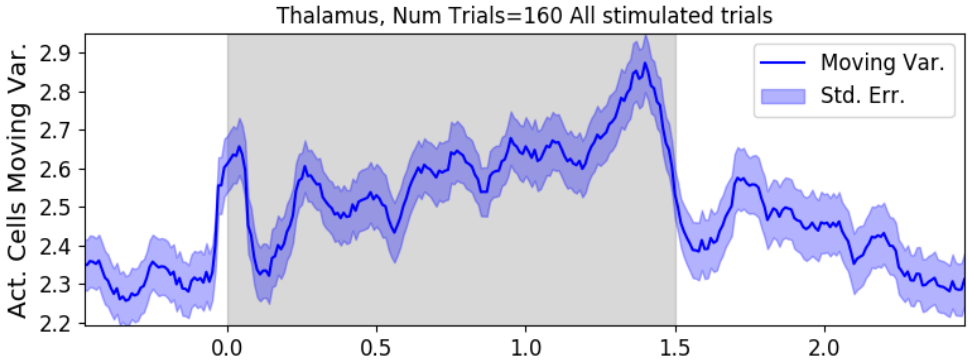
Figure 2: The (a) average and (b) variance of the number of active neurons, measured using a sliding window 100ms wide, split into 100 bins. The midpoint of the time interval for each window is used as the timepoint (x-axis point) for the measurements using that window. The grey shaded area indicates the presence of a visual stimulus. The opaque line is an average across the 160 trials that included a visual stimulus of any kind. We can see a transient increase in the average number of active neurons and the variance of this number.

2.1.3 Thalamus

In the thalamus we observed a transient increased in the both the average and variance of the number of active neurons on stimulus onset, followed by a fall to pre-stimulus levels, followed by a sustained increase until the stimulus presentation ends.



(a) Moving average.



(b) Moving variance.

Figure 3: The (a) average and (b) variance of the number of active neurons, measured using a sliding window 100ms wide, split into 100 bins. The midpoint of the time interval for each window is used as the timepoint (x-axis point) for the measurements using that window. The grey shaded area indicates the presence of a visual stimulus. The opaque line is an average across the 160 trials that included a visual stimulus of any kind.

2.2 Conway-Maxwell-binomial distribution is usually a better fit than binomial or beta-binomial

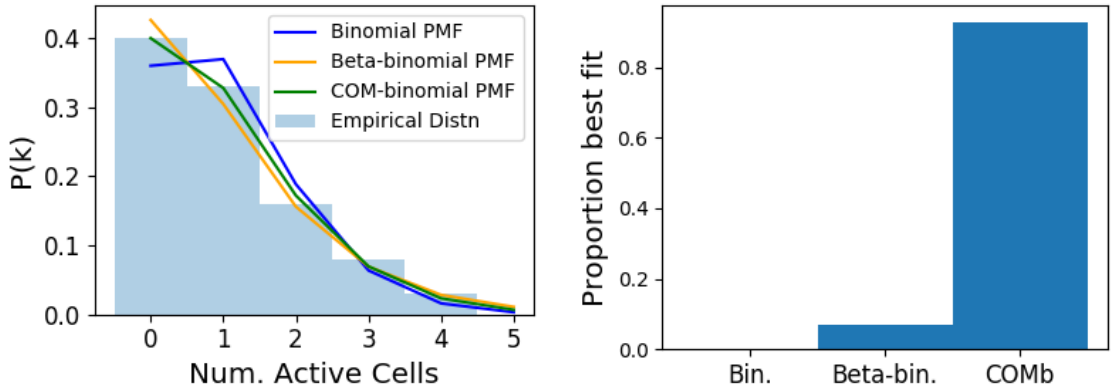
After fitting a binomial, a beta-binomial, and a Conway-Maxwell-binomial (COMb) distribution to each window for each bin width, and each region, we assessed the goodness-of-fit of each distribution by calculating the log-likelihood of each fitted distribution using the associated sample. We measured the proportion of samples for which each distribution was the best fit, for each bin width value and each region.

We found that the COMb distribution was the best fit for most of the samples regardless of bin width or region. The bin width had an effect on the number of samples for which the COMb distribution was the best fit. For a bin width of 1ms, the COMb distribution was the best fit for over 90% of samples, the beta-binomial dis-

tribution was the best fit for less than 10% of samples, and the binomial distribution was the best fit for less that 1% of samples, across regions. For 5ms bins, the COMb distribution was the best fit for $70 - 80\%$ of samples, the beta-binomial distribution was the best fit for $20 - 30\%$ of the samples, and again the binomial distribution was the best fit for less that 1% of samples, across regions. Finally, for 10ms bins, the COMb distribution was the best fit for $53 - 80\%$ of samples, the beta-binomial distribution was the best fit for $20 - 47\%$ of the samples, and the binomial distribution was the best fit for less that 0.1% of samples, across regions. These findings are summarised in table 1.

Bin Width (ms)	Binomial	Beta-binomial	COMb
1ms	$< 1\%$	$< 10\%$	$> 90\%$
5ms	$< 0.1\%$	$20 - 30\%$	$70 - 80\%$
10ms	$< 0.1\%$	$20 - 47\%$	$53 - 80\%$

Table 1: Proportion of samples for which each distribution was the best fit, grouped by bin width. The COMb distribution is the best fit most of the time.



(a) Example of fitted distributions.

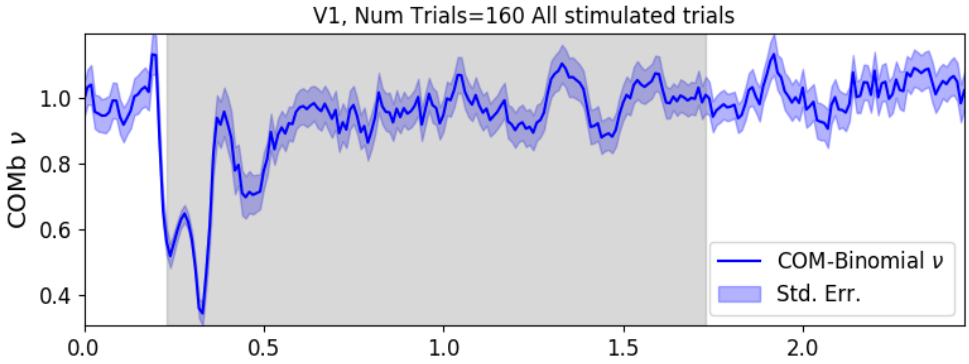
(b) Proportion of best fit

Figure 4: (a) An example of the binomial, beta-binomial, and Conway-Maxwell-binomial distributions fitted to a sample of neural activity. The Conway-Maxwell-binomial distribution is the best fit in this case. The histogram shows the empirical distribution of the sample. The probability mass function of each distribution is indicated by a differently coloured line. (b) Across all samples in all trials, the proportion of samples for which each fitted distribution was the the best fit. The Conway-Maxwell-binomial distribution was the best fit for 93% of the samples taken from V1 using a bin width of 1ms.

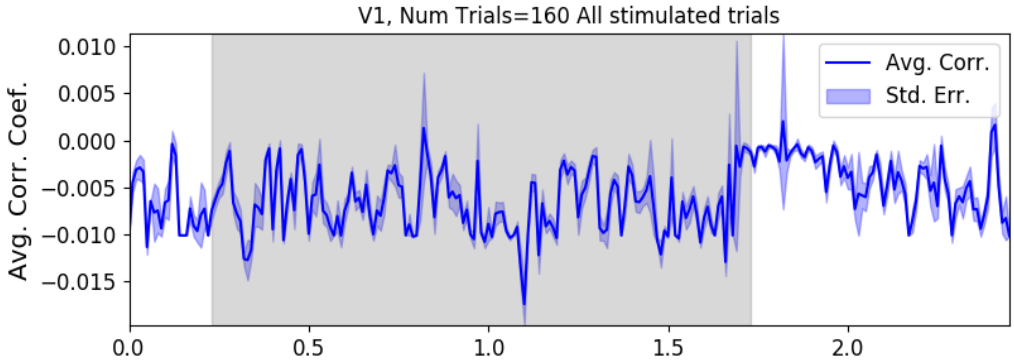
2.3 Conway-Maxwell-binomial distribution captures changes in association at stimulus onset

We fit a Conway-Maxwell-binomial (COMb) distribution to the number of active neurons in the 1ms time bins in a 100ms sliding window. We also measured the correlation coefficient between the spike counts of all possible pairs of neurons, and took the average of these coefficients. We did this for all the trials with a visual stimulus. We observed a reduction in the COMb distribution's ν parameter at stimulus onset from around 1 to between 0 and 1 (see figure 5a). A value of ν less than 1 indicates positive association between the neurons (see section 5.5.4). We might expect to see this positive association reflected in the correlation coefficients, but this is not the case. We see no change in the time series of average correlation measures at stimulus onset.

This may be due to the very short time bin we used in this case. We know that using small time bins can artificially reduce correlation measurements [3]. In this case, fitting the COMb distribution may be a useful way to measure association in a neuronal ensemble over very short timescales (< 10ms).



(a) COMb ν parameter.



(b) Average correlation coefficient.

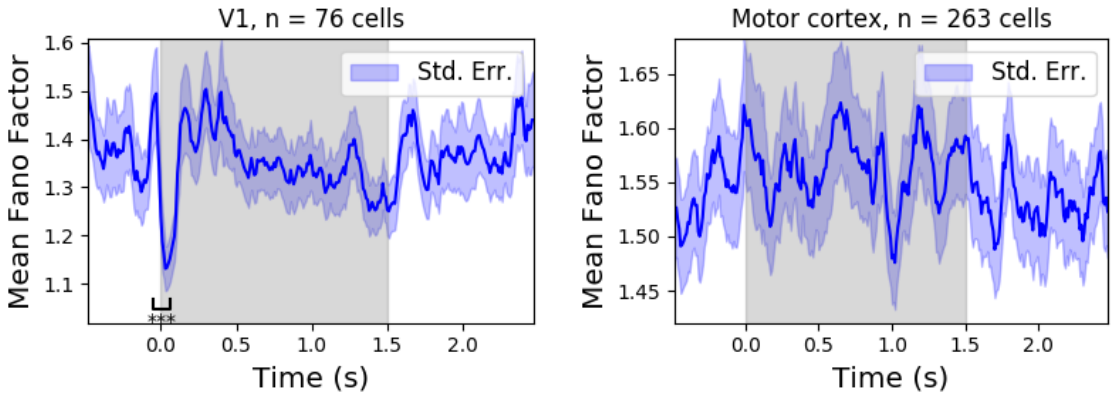
Figure 5: (a) We fit a Conway-Maxwell-binomial distribution to the number of active neurons in 1ms time bins of a 100ms sliding window. We did this for all trials with a visual stimulus and took the average across those trials. We see a transient drop in value for the distribution's ν parameter at stimulus onset. This shows an increase in association between the neurons. (b) We measured the correlation coefficient between the spike counts of all possible pairs of neurons in the same sliding window. The took the average of those coefficients. We also did this for every visually stimulated trial, and took the average across trials. The increase in association is not reflected with an increase in average correlation.

2.4 Replicating stimulus related quenching of neural variability

Churchland et al. (2010) inspected the effect of a stimulus on neural variability. One of the measures of neural variability that they employed was the Fano factor of the spike counts of individual cells (see section 5.4). They found a reduction in neural variability as measured by the Fano factor in various cortical areas in a macaque [2].

We measured the fano factor of the spike count of each cell in each brain region, during each trial. We measured the mean and standard error of these Fano factors from 500ms before stimulus onset until 1000ms after stimulus end. For the primary visual cortex, we found a transient reduction in the Fano factor immediately after stimulus onset. We used a Mann-Whitney U test to check that the Fano factors measured in a window starting at stimulus onset and ending 100ms later were significantly lower than the factors measured in a window ending at stimulus onset ($p < 0.001$, see figure 6a). We did not get this statistically significant result in any other region.

Our findings agree with those of Churchland et al for the primary visual cortex. However Churchland also found a reduction in the Fano factor in the dorsal premotor cortex (PMd) at stimulus onset. Our measurements from the mouse motor cortex show no change at stimulus onset (see figure 6b). This could indicate some difference in the functionality of the motor cortex in a macaque and the motor cortex of a mouse.



(a) Primary visual cortex.

(b) Motor cortex.

Figure 6: (a) The mean Fano factor of the spike counts of the cells in the primary visual cortex. Means were taken across cells first, then across trials. There was a significant decrease in the Fano factors immediately after stimulus onset. (b) The mean Fano factor of the spike counts of the cells in the motor cortex. No significant change in measurements at any point.

3 Discussion

4 Data

We used data collected by Nick Steinmetz and his lab 'CortexLab at UCL' [13]. The data can be found online [1] and are free to use for research purposes.

Two 'Phase3' Neuropixels [7] 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.

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

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 $(0^\circ, 22.5^\circ, \ldots, 337.5^\circ)$ 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

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.

There was an interval between each trial with a grey image in place of the moving of the moving bar stimulus. This interval varied in time. But we included some of this interval when recording the data for each trial. We started recording the number of active neurons, and the number of spikes from each neuron from 280ms before each trial until 280ms after each trial. This way, we could see the change in our measurements at the onset of a stimulus.

The actual measurements we took in each window were as follows:

Number of active neurons The number of neurons firing a spike in each bin. Most of the other measurements are aggregations of these measurements, or models fitted to these measurements.

Spike counts for each cell The number of spikes fired by each cell in each bin.

Moving average The average number of active cells in each window.

Moving variance The variance of the number of active cells in each window.

Average correlation We measured the correlation between the spike counts of each pair of cells in the ensemble, and took the average of these measurements.

Binomial p We fitted a binomial distribution to the data in each window and recorded the fitted probability of success, p in each case.

Beta-binomial α, β We fitted a beta-binomial distribution to the data in each window, and recorded the values of the fitted shape parameters, α and β, of each distribution.

Conway-Maxwell-binomial distribution p, ν We fitted a Conway-Maxwell-binomial distribution to the data in each window, and recorded the fitted values of p and ν for each distribution.

Log-likelihoods We also recorded the log-likelihood of each of the fitted distributions for each window.

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 activity

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

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

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

$$P(k; n, \alpha, \beta) = \binom{n}{k} \frac{B(k+\alpha, n-k+\beta)}{B(\alpha, \beta)} \tag{3}$$

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

$$\pi = \frac{\alpha}{\alpha + \beta} \tag{4}$$

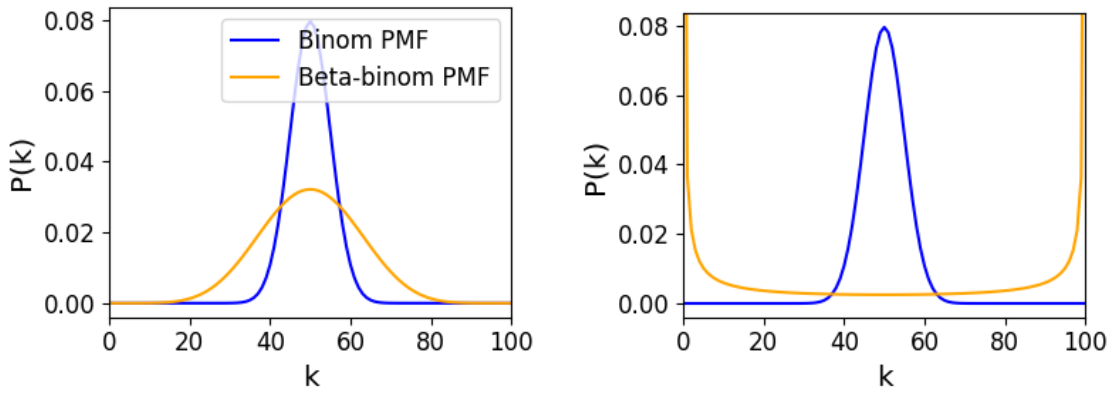
$$\rho = \frac{1}{\alpha + \beta + 1} \tag{5}$$

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 7b. 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 $\nu \in \mathbb{R}$ the shape parameter.



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

Figure 7: Figures showing the over-dispersion possible for a beta-binomial distribution relative to a binomial distribution. Parameters are shown in the captions.

The probability mass function of the COMb distribution is

$$P(k; n, p, \nu) = \frac{1}{S(n, p, \nu)} \binom{n}{k}^{\nu} p^k (1-p)^{n-k} \tag{6}$$

where

$$S(n, p, \nu) = \sum_{j=0}^{n} \binom{n}{k}^{\nu} p^j (1-p)^{n-j} \tag{7}$$

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. We can see in figure 8d that the KL-divergence $D_{KL}(P_{COMb}(n, p, \nu)||P_{Bin}(n, p)) = 0$ along the line where $\nu = 1$. The analytical expression for the divergence is

$$D_{KL}(P_{COMb}(k; n, p, \nu)||P_{Bin}(k; n, p)) = (\nu - 1) E_{P_{COMb}(k;n,p,\nu)} \left[\log \binom{n}{k} \right] \tag{8}$$

$$- \log S(n, p, \nu) \tag{9}$$

At $\nu = 1$, we have $S(n, p, 1)$ which is just the sum over the binomial PMF, so $S(n, p, 1) = 1$ and therefore $D_{KL}(P_{COMb}(n, p, \nu)||P_{Bin}(n, p)) = 0$.

If $\nu < 1$ the COMb distribution will exhibit over-dispersion relative to the bi-

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

Table 2: Relative dispersion of the COMb distribution, and association between Bernoulli variables as represented by the value of the ν parameter.

nomial 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. An example of this over-dispersion can be seen in figure 8b.

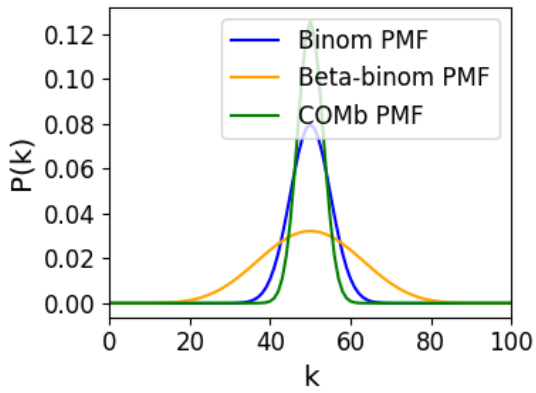
If $\nu > 1$ the COMb distribution will exhibit under-dispersion relative to the binomial distribution. The larger the value of ν the more probability mass will build up at $n/2$ for even n, or at $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ for odd n. This under-dispersion represents negative association in the Bernoulli variables. An example of this under-dispersion can be seen in figure 8a.

It should be noted that the shape parameter of the COMb distribution, p, does not correspond to the mean of the distribution, as is the case for the binomial and beta-binomial distributions. An illustration of this can be seen in figure 8c. This is because an interaction between the p and ν parameters skews the mean. There is no analytical expression for the mean of the COMb distribution.

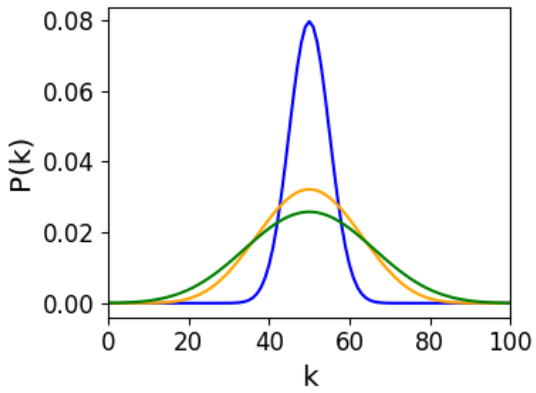
Since the COMb distribution has the potential to capture positive and negative associatons between the neurons/Bernoulli variables, it should be an excellent candidate for modelling the number of active neurons in a neuronal ensemble.

We wrote a dedicated Python package to enable easy creation and fitting of COMb distribution objects. The format of the package imitates the format of other distribution objects from the `scipy.stats` Python package. The COMb package can be found here:

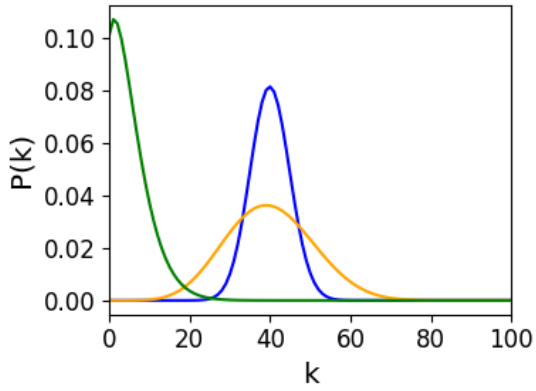
`https://github.com/thomasjdelaney/Conway_Maxwell_Binomial_Distribution`



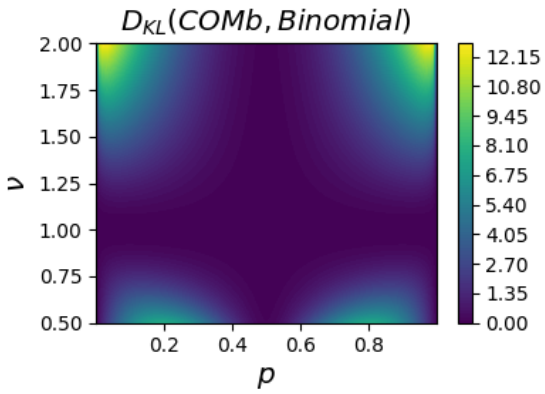
(a) $n = 100, p = 0.5, \alpha = \beta = 10, \nu = 2.5$



(b) $n = 100, p = 0.5, \alpha = \beta = 0.3, \nu = 0.1$



(c) $n = 100, p = 0.4, \alpha = 10, \beta = 15, \nu = 0.1$



(d) KL-Divergence as a function of p and ν. $n = 100$.

Figure 8: Figures showing (a) the over-dispersion and (b) under-dispersion permitted by the COMb distribution. (c) illustrates that the shape parameter of the COMb distribution, p, does not correspond to the mean of the distribution, as it does for the binomial and beta-binomial distributions. (d) shows a heatmap for the value of the Kullback-Liebler divergence between the COMb distribution and the standard binomial distribution with same value for n, as a function of p and ν. Parameters are shown in the captions.

5.6 Fitting

We fitted a binomial, beta-binomial, and Conway-Maxwell-binomial (COMb) distribution to the neural activity in each of the overlapping windows covering each trial. To fit the distributions we minimised the appropriate negative log likelihood function using the data from the window.

There is an analytical solution for the binomial distribution.

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{N} k_i \tag{10}$$

We minimised the negative log likelihood function of the beta-binomial distribution numerically.

The log likelihood function of the COMb distribution is

$$\begin{aligned} \ell(p, \nu | k_1, \ldots, k_N) =& N \left[n \log(1-p) - \log S(n, p, \nu) \right] && (11) \\ &+ \log \frac{p}{1-p} \sum_{i=1}^{N} k_i && (12) \\ &+ \nu \sum_{i=1}^{N} \log \binom{n}{k_i} && (13) \end{aligned}$$

We minimised the negation of this function using numerical methods.

More specifically, we used the `minimise` function of the `scipy.optimize` Python package.

5.7 Goodness-of-fit

After fitting, we measured the goodness-of-fit of each model/distribution with their log likelihood. We calculated this directly using the `logpmf` functions of the distribution objects in Python.

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] Marlene R Cohen, Adam Kohn, *Measuring and interpreting neuronal correlations* Nature neuroscience 14, 811-819, (2011)

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

[5] Boris Flach, *A class of random fields on complete graphs with tractable partition function.* IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 2304-2306, (2013)

[6] Elad Ganmor, Ronen Segev, Elad Schneidman, *Sparse low-order interaction network underlies a highly correlated and learnable neural population code.* PNAS 108, 9679-9684, (2011)

[7] 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)

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

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

[10] 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)

[11] 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)

[12] Elad Schneidman, Michael J. Berry II, Ronen Segev, William Bialek, *Weak pairwise correlations imply strongly correlated network states in a neural population*. Nature 440, 1007-1012 (2006)

[13] 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)

[14] Gašper Tkačik, Olivier Marre, Dario Amodei, Elad Schneidman, William Bialek, Michael J. Berry II, *Searching for Collective Behavior in a Large Network of Sensory Neurons.* PLOS computational biology 10, 1-23 (2014)

[15] 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)