

Flexible models for spike count data with both over- and under- dispersion

Ian H. Stevenson^{1,2,3}

Received: 28 December 2015 / Revised: 14 March 2016 / Accepted: 18 March 2016 © Springer Science+Business Media New York 2016

Abstract A key observation in systems neuroscience is that neural responses vary, even in controlled settings where stimuli are held constant. Many statistical models assume that trialto-trial spike count variability is Poisson, but there is considerable evidence that neurons can be substantially more or less variable than Poisson depending on the stimuli, attentional state, and brain area. Here we examine a set of spike count models based on the Conway-Maxwell-Poisson (COM-Poisson) distribution that can flexibly account for both overand under-dispersion in spike count data. We illustrate applications of this noise model for Bayesian estimation of tuning curves and peri-stimulus time histograms. We find that COM-Poisson models with group/observation-level dispersion, where spike count variability is a function of time or stimulus, produce more accurate descriptions of spike counts compared to Poisson models as well as negative-binomial models often used as alternatives. Since dispersion is one determinant of parameter standard errors, COM-Poisson models are also likely to yield more accurate model comparison. More generally,

Action Editor: Liam Paninski

Electronic supplementary material The online version of this article (doi:10.1007/s10827-016-0603-y) contains supplementary material, which is available to authorized users.

Ian H. Stevenson ian.stevenson@uconn.edu

- ¹ Department of Psychological Sciences, University of Connecticut, Storrs, CT, USA
- ² Department of Biomedical Engineering, University of Connecticut, Storrs, CT, USA
- ³ Connecticut Institute for Brain and Cognitive Science, Storrs, CT, USA

these methods provide a useful, model-based framework for inferring both the mean and variability of neural responses.

Keywords Spike count variability · Tuning curves · Poisson · Conway-Maxwell-Poisson

1 Introduction

Variability in neural responses is ubiquitous due to cellular processes, such as channel dynamics and probabilistic vesicle release (Faisal et al. 2008), as well as, changes in network state or neuromodulation (Arieli et al. 1996; Masquelier 2013). Increasing variability, in general, reduces how accurately information about the external world can be decoded from neural responses (van Steveninck et al. 1997; Werner and Mountcastle 1963). However, variability may also serve a functional role (Ermentrout et al. 2008; Hoyer et al. 2003; Stein et al. 2005) and can act as a signature of decision making (A. K. Churchland et al. 2011), movement preparation (M. M. Churchland et al. 2006), or stimulus onset (M. M. Churchland et al. 2010). Although many model-based methods focus on describing the mean spiking response for a given stimulus or task, new techniques that also quantify how spike variability changes as a function of stimulus/movement parameters may provide another dimension for understanding the neural code. Here we examine a set of models that assume spike counts are distributed according to a Conway-Maxwell-Poisson (COM-Poisson) distribution with group/observation-level dispersion (Sellers et al. 2012). This approach can describe both overand under-dispersion and can flexibly capture patterns of variability in spike count data when fitting tuning curves and peristimulus time histograms (PSTH).

Although many tuning curve and PSTH models assume that spike counts are Poisson distributed, where the variance of the spike counts equals the mean, in practice, spike count variability can be both under-dispersed-where the count variance is less than the mean count - or over-dispersed - where variance is greater than the mean count. In early sensory systems, for instance, neurons appear to be much more precise, with less variability than would be predicted by the Poisson model (Berry et al. 1997; DeWeese et al. 2003). Variability appears to increase along sensory pathways (Kara et al. 2000; Vogel et al. 2005), and can be substantially less precise than Poisson in motor areas (Lee et al. 1998). Quantifying the variability present in spike observations (Nawrot 2010; Shinomoto et al. 2009) and interpreting the underlying causes of variability (Azouz and Gray 1999; Bair and Koch 1996; Carandini 2004; Deweese and Zador 2004; Schölvinck et al. 2015; Softky and Koch 1993; Zador 1998) are fundamental challenges to deciphering the neural code. In many cases, Poisson models where the rate is solely a function of the stimulus fail to provide an accurate description of spike count variability across trials (Amarasingham et al. 2006; Maimon and Assad 2009).

One explanation for the deviations from Poisson firing is that spike timing is driven, not just by extrinsic stimuli or behavior, but by single-neuron dynamics and other intrinsic factors. By taking these different factors into account, many models are able capture under-dispersion or over-dispersion across trials (Harris et al. 2003; Truccolo et al. 2005). Models of spike dynamics and history dependence (Berry and Meister 1998; Keat et al. 2001; Pillow et al. 2005; Reich et al. 1997; Uzzell and Chichilnisky 2004) are able to produce under-dispersion, while models using covariates that vary from trial to trial, such as interactions between neurons (Pillow et al. 2008; Stevenson et al. 2008), local field potentials (Harris et al. 2003; Kelly et al. 2010), or latent variables (Czanner et al. 2008; Eden et al. 2004; Paninski et al. 2010), are able to produce over-dispersion. However, fitting such detailed models may not always desirable. More importantly, since most rate-models only describe the mean spike count, there is typically no guarantee that the observed and model variability will match; although, see (Gao et al. 2015; Lansky and Vaillant 2000).

The COM-Poisson model has several desirable properties. Most importantly it is able to capture both over- and underdispersion (Shmueli et al. 2004). It can thus describe neural responses when spiking is less variable than Poisson as well as more variable. By modeling variability at the group- or observation-level, we show how this approach can also account for cases where the same neuron exhibits different levels of dispersion in response to different stimuli (tuning curves) or as a function of time (PSTHs). In contrast to descriptive statistics such as the Fano Factor, the COM-Poisson regression approach is model-based and allows principled model comparison. Using both simulated and experimental data we show how different noise models (e.g. Poisson, negative binomial, COM-Poisson, and generalized count) can be evaluated in terms of their ability to accurately predict spike counts. Here we find that failing to take dispersion into account results in less accurate prediction and results in biased estimates of parameter standard errors.

2 Methods

2.1 Conway-Maxwell-Poisson models

Here we model non-Poisson spike count variability using the COM-Poisson distribution. The likelihood for the COM-Poisson distribution is given by

$$p(y|\lambda,\nu) = \frac{\lambda^y}{y!^{\nu}} \frac{1}{Z(\lambda,\nu)}$$

where the probability of observing *y* counts is a function of the parameters λ and ν , normalized by a factor $Z(\lambda, \nu) = \sum_{y=0}^{\infty} \frac{\lambda^y}{y!^{\nu}}$. Note that for $\nu = 1$ the COM-Poisson likelihood has $Z(\lambda, \nu) = e^{\lambda}$, giving the Poisson likelihood. However, unlike the Poisson distribution, where the mean is always equal to the variance, the COM-Poisson distribution can exhibit both overdispersion and under-dispersion. $\nu < 1$ leads to over-dispersion with var[y] > E[y], and $\nu > 1$ leads under-dispersion with var[y] < E[y].

In addition to the Poisson distribution, the COM-Poisson also contains the Bernoulli and geometric distributions as a special cases (Sellers et al. 2012; Shmueli et al. 2004) when $\nu \rightarrow \infty$ and $\nu \rightarrow 0$, respectively.

Though there is no closed-form expression for the normalization $Z(\lambda, \nu)$, in practice we can compute it numerically with a finite sum up to some number of possibly observable spike counts (Minka et al. 2003). Since early truncation can result in misestimation of the likelihood, here we use the greater of 1000 or twice the maximum of the observed counts for the tuning curve data and the greater of 100 or twice the maximum of the observed counts for the PSTH data. More efficient methods based on asymptotic approximations or lookup tables may yield substantial speedups.

2.2 Generalized linear models with COM-Poisson observations

Although the lack of closed-form normalization makes inference more computationally difficult, the COM-Poisson is an exponential family distribution (Sellers and Shmueli 2010; Sellers et al. 2012; Shmueli et al. 2004), and we can construct a generalized linear model using the link functions

$$log(\lambda) = X\beta log(\nu) = G\gamma$$

where X and G are covariates and β and γ are parameter vectors. Here we distinguish between two cases of COM-Poisson regression: the constant dispersion parameter case—where ν is fixed—and the case with observationor group-level dispersion – where ν varies (Sellers and Shmueli 2009, 2013).

For tuning curve models, similar to (Sanger 1996), we use a Fourier basis in *X* and *G* up to some frequency *n* and *m* to generate smooth estimates of both the mean spike count and, in this case, dispersion as a function of a circular stimulus/ movement parameter θ .

 $X = \begin{bmatrix} 1 & \sin\theta & \cos\theta & \cdots & \sin n\theta & \cos n\theta \end{bmatrix}$ $G = \begin{bmatrix} 1 & \sin\theta & \cos\theta & \cdots & \sin m\theta & \cos m\theta \end{bmatrix}$

For PSTH models, instead of a Fourier basis, we use a B-spline basis (De Boor 1978) with *n* and *m* equally-spaced knots, for *X* and *G*. Sampling knot locations adaptively, for instance using Bayesian adaptive regression splines (BARS), has been shown to better capture the fast changes present in PSTHs (Dimatteo et al. 2001; Kass et al. 2003; Kaufman et al. 2005), but, for simplicity, we opt for a fixed basis here. In both the tuning curve and PSTH models we can enforce a constant dispersion parameter by setting m=0.

2.3 Maximum likelihood, MAP estimation, and Bayesian inference

Since the COM-Poisson distribution is exponential family, maximum likelihood estimates for β and γ can be found with iterative reweighted least squares, IRLS (Sellers and Shmueli 2010). Here we use a Bayesian approach—adding priors to aid convergence and avoid over-fitting. As with other generalized linear models (Gelman et al. 2008; Zhao and Iyengar 2010), under certain combinations of observations and covariates the COM-Poisson model fails to converge. In particular, convergence issues occur when the Fisher scoring matrix G^TWG or X^TWX is low rank, given the diagonal weight matrix W from IRLS. Also as with other GLMs, increasing model complexity (e.g. large n and m above) can lead to over-fitting. Adding priors allows us to flexibly overcome both these issues. Here we use normal priors with no penalty on the intercept term

```
\beta_{>1} \text{Normal}(0, \sigma_{\beta}) \\ \gamma_{>1} \text{Normal}(0, \sigma_{\gamma})
```

where $\beta_{>1}$ and $\gamma_{>1}$ denote the coefficients excepting the intercept. For the tuning curve models here we use n=2, m=1 and for the PSTH models we use n=20, m=8knots, except where noted. In both cases we use $\sigma_{\beta} = 10, \sigma_{\gamma} = 1$ (after X and G have been standardized), but as with other regularization schemes the model complexity and hyper-parameters can generally be adjusted depending on the data. Here, in illustrating the model, we have chosen these settings to give qualitatively good results, but other settings as well as priors (Kadane et al. 2006), may provide greater accuracy. To find the MAP (maximum a posteriori) estimates or posterior samples for β and γ we use LBFGS and No U-Turn Sampling (NUTS), respectively (Hoffman and Gelman 2014). Both algorithms are implemented through the Stan sampling library ("Stan: A C++ Library for Probability and Sampling, Version 2.8.0" 2015).

The major computational difference between the Poisson and COM-Poisson regression is that with the COM-Poisson model we need to numerically evaluate the normalization factor $Z(\lambda, \nu)$ every time the likelihood is evaluated. However, there is also a small difference in the structure of the likelihood calculation. If spike counts $\{y_1 \dots y_n\}$ are observed on *n* different trials of a fixed λ , with Poisson observations we only need compute the log-likelihood for the sum

$$\sum \log p(y_i | \lambda)$$

= $\sum y_i \log \lambda - \sum \lambda - \sum \log y_i! = \log p(\sum y_i | n\lambda).$

Since a sum of Poisson random variables is Poisson, $\sum y$ provides a sufficient statistic for λ . For COM-Poisson random variables (Shmueli et al. 2004), on the other hand, we have $\sum \log p(y_i | \lambda, \nu) = \sum y_i \log \lambda - \sum \nu \log y_i ! - \sum Z(\lambda, \nu) = \log \lambda \sum y_i - \nu \sum \log y_i ! - nZ(\lambda, \nu).$

In this case, $\sum y$ and $\sum \log y!$ together provide sufficient statistics for λ and ν . We find that when fitting PSTHs it is more efficient to use these sufficient statistics so that, for p unique stimuli/time-points, we only need evaluate the likelihood O(p) rather than O(np) times. As $\sum y$ become large, however, calculating $Z(\lambda, \nu)$ requires a larger support and some of this efficiency is lost.

2.4 Model comparison

To determine whether the COM-Poisson model is useful for describing spike counts in practice, we use the same methods (basis and MAP/sampling procedures) to fit Poisson, negative binomial, and generalized count observation models. The Poisson and negative binomial models have been used extensively, elsewhere (Brown

$$\log(\lambda) = X\beta$$
$$p(y|\lambda) = \frac{\lambda^{y}e^{-\lambda}}{y!}$$

And the canonical negative binomial model is given by,

$$\log(\mu) = X\beta$$

$$p(y|\mu, r) = \left(\frac{r}{r+\mu}\right)^r \frac{\Gamma(r+y)}{\Gamma(y+1)\Gamma(r)} \left(\frac{\mu}{r+\mu}\right)^y$$

where the additional parameter r (>0) allows over-dispersion and $\Gamma(\cdot)$ denotes the gamma function. One advantage of the Poisson and negative binomial models is that, unlike the COM-Poisson, they do have closed-from normalization. However, the Poisson model is always equi-dispersed while the negative binomial model is unable to describe underdispersed data, since the variance $\mu + \frac{\mu^2}{r}$ is always greater than or equal to the mean μ . Similar to the COM-Poisson model, the negative binomial model can be extended to model group-level dispersion (Taouali et al. 2016). Here we make the dispersion parameter r a function of time/stimuli by using the reparameterization $\kappa = 1/r$ and assuming $\log(\kappa) = G\gamma$.

In addition to these two canonical models we also implement the recently described generalized count distribution (del Castillo and Pérez-Casany 2005; Gao et al. 2015). Briefly, the generalized count model takes the form

$$\begin{aligned} \theta &= X\beta \\ p\Big(y\Big|\theta, g(\cdot)\Big) &= \frac{\exp(\theta y + g(y))}{y! Z(\theta, g(\cdot))} \end{aligned}$$

By varying the form of function $g(\cdot)$ this model contains the Poisson, negative binomial, and COM-Poisson distributions, among others, as special cases. In particular, after moving the intercept term out of X into g(y), Poisson regression has $g(y)=\alpha y$, negative binomial regression has $g(y)=\alpha y+\log(y+$ r-1)!, and COM-Poisson regression has $g(y)=\alpha y+\log(y+$ r-1)!, and COM-Poisson regression has $g(y)=\alpha y+(1-\nu)\log y!$. Fixing g(0)=0 ensures model identifiability, while fixing the support and adding additional smoothness constraints can allow arbitrary functions of y to be fit. Here, to examine to what extent our results are affected by the specific distributional assumptions, we use apply methods and code from (Gao et al. 2015) and fit the full function $g(\cdot)$ with a quadratic penalty on the second differences (setting the hyperparameter $\lambda_g = 50$, with no penalty on β , and LBFGS optimization).

Similar to the COM-Poisson distribution, the generalized count distribution has the advantage that it allows both overand under-dispersion under certain parameterizations of g(y) and the disadvantage that the normalization $Z(\theta, g(\cdot))$ typically needs to computed numerically. Due to its flexibility, it is, in some sense, guaranteed to out-perform the other models for a fixed stimulus given enough data.

Since the COM-Poisson model ($\nu = 1$), the negative binomial model ($r \rightarrow \infty$), and the generalized count model ($g(y) = \alpha y$) all contain the Poisson as a special case, we use the cross-validated (20-fold) log likelihood ratio relative to a homogeneous (constant λ Poisson model as a common point of comparison during our tuning curve analyses.

2.5 Simulating the relationship between variable input current and spike count variability

To illustrate the flexible origins of spike count variability (Fig. 1), we use an adaptive exponential integrate and fire model (Brette and Gerstner 2005) whose dynamics are given by

$$C\frac{dV}{dt} = -g_L(V - E_L) + g_L \Delta_T \exp\left(\frac{V - V_T}{\Delta_T}\right) - w + I(t)$$

$$\tau_w \frac{dw}{dt} = a(V - E_L) - w$$

Spikes are emitted when $V > V_{thresh}$ and the state is updated $V \rightarrow V_{reset}$, $w \rightarrow w+b$ following each spike. Here we use parameters from (Brette and Gerstner 2005) based on a regular spiking pyramidal neuron: $E_L = -70$ mV, $V_{thresh} = -50$ mV, $V_{reset} = -70$ mV, C=281pF, $g_L = 30$ nS, $\Delta_T = 2$ mV, $\tau_w = 144$ ms, a=4 nS, b=80 pA. However, similar results can be obtained with a wide range of parameter settings as well as other voltage models (Zador 1998).

2.6 Experimental Fano factors

To quantify dispersion in spike counts empirically - as opposed to using a model—we use the Fano Factor $F = \sigma^2/\mu$. Using data from a fixed stimulus/time window we estimate the sample mean and variance then quantify uncertainty about the Fano Factor using Bayesian bootstrapping (Rubin 1981). In particular, we generate bootstrap samples by drawing weights w_k for each observation from a uniform Dirichlet distribution ($\alpha = 1$ for all observations). We then compute the weighted mean $\mu' = \sum w_k v_k$ and variance $\sigma'^2 = \sum w_k (v_k - \mu')^2$. whose ratio provides a bootstrap sample for the Fano Factor itself. After many (in this case, 1000) samples we have a smooth distribution that can be used to estimate uncertainty about the Fano Factor. With a small number of count observations Bayesian boostrap typically provides a more accurate estimate of quantiles than the traditional bootstrap. See (Eden and Kramer 2010) for an alternative approach.

Fano Factors reported in Fig. 2c were collected, *ad-hoc* from ten studies. Since the goal of this collection is simply to illustrate the range of realistic spike count variability, data



Fig. 1 Trial-to-trial variability depends on input. a) Spike responses of an adaptive exponential integrate and fire neuron to different types of repeated input currents: DC noise (*left*), fixed DC (*middle*), and frozen noise (*right*). Although the parameters of the neuron model are identical between the three different settings, trial-to-trial variability can range from over-dispersed (*left*) to highly under-dispersed (*right*) depending on the input, and a single noise model for spike counts may not be able to accurately describe all situations. Trials are sorted by total number of

spikes, and the initial voltage is random. **b**) The same model neuron responding to current whose trial-to-trial variability is a function of time. The average current (*top*) is denoted by the light trace with standard deviation denoted by the *gray band*. A single current trace for a typical trial is also shown. Note that the Fano Factor varies over time (50 ms sliding window) and can be both less than 1 (under-dispersion) and greater than 1 (over-dispersion)

come from a wide range of species, brain areas, and tasks with substantial heterogeneity among firing rates and windows used to compute *F*. Values for the mean Fano Factor, standard deviation, and number of neurons were collected from text and digitized figures (see Supplementary Material for list of sources).

2.7 Tuning curve data

Tuning curve data was obtained from the neural signal archive (neuralsignal.org) nsa2004.6b and nsa2004.6c and consists of 56 recordings from 45 cells in areas MT/V5 of anesthetized (sufentanil), paralyzed (vecuronium) adult male macaque



Fig. 2 The Conway-Maxwell-Poisson Distribution can describe both under- and over-dispersion. **a**) COM-Poisson likelihood for various values of ν Note that $\nu = 1$ corresponds to a Poisson model, while $\nu < 1$ generates over-dispersion, and $\nu > 1$ generates under-dispersion. The *solid lines* denote mean, dashed lines denote quartiles, and the likelihoods are scaled and smoothed for clarity. **b**) The mean-variance relationships for the COM-Poisson distributions in **a** (*left*) and for two alternative models of count data: the negative binomial distribution which can describe over-dispersion and the binomial model which can describe

under-dispersion (right). The *solid black line* denotes equal mean and variance as in the Poisson distribution. c) Experimentally observed Fano Factors are often close to 1 but can exhibit over- as well as underdispersion depending on the species, brain area, time scale, and stimulus. The *position of each circle* denotes the mean Fano Factor reported in one study, *error bars* denote standard deviation, and the circle area indicates the number of neurons in the study (20 and 200 shown for reference). The *violin plot at right* shows the combined distribution from all studies along with the median and quartiles (*gray lines*)

monkeys (macaca fascicularis, N=5 and macaca nemestrina, N=1). For details of the experimental methods see (Kohn and Movshon 2003). Briefly, visual stimuli consisted of drifting sine-wave gratings and moving dots (at 100 % contrast/coherence). Both grating and random dot stimuli were presented with 16 movement directions plus 1 blank stimulus presented in random order with each trial lasting 1000–1280 ms and presented continuously, without breaks. The total number of trials varied between 51 and 204. Here we use spike counts 50–550 ms following stimulus onset to estimate direction tuning and ignore responses to the blank stimulus.

2.8 PSTH data

PSTH data was obtained from the neural signal archive (neuralsignal.org) nsa2009.1, which consists of 52 recordings from 26 cells in areas MT/V5 of an awake, fixating adult male macaque monkey (macaca mulatta, N=1) performing a direction discrimination task. For details of the experimental methods see (Britten et al. 1992). Here trials are neural responses to repeated presentations of particular incoherent, random dot patterns, either with frozen noise or with a variable seed. Trials lasted 2000–2048 ms, and recording lengths varied from 14 to 210 trials. See previous analysis in (Bair and Koch 1996).

3 Results

Although many models assume that spike counts are Poisson distributed, the fact that spiking is close to Poisson is somewhat remarkable when you consider how spikes are generated. Previous work on voltage-models and intracellular experiments has illustrated that the reliability of spiking over time and across trials heavily depends on the input that a neuron receives (Zador 1998). In a simulated neuron model (see Methods), neurons can readily generate trial-to-trial variability ranging from over-dispersed to highly under-dispersed (Fig. 1). If the synaptic input is reliable then spike count variability can be very low (Mainen and Sejnowski 1995), but even when the input to a single neuron is highly controlled, variations in the initial state of the neuron can produce irregular firing patterns (Fig. 1a). In vivo, differences in intrinsic covariates, such as network state (Arieli et al. 1996), will lead to changes in the exact input that a neuron receives. Small variations in synaptic input and membrane potential appear to be amplified by the non-linear thresholding effect of spike generation (Carandini 2004), and spiking responses can easily be both under- and over-dispersed (Shadlen and Newsome 1998). In general, synaptic input varies both over time and from trial to trial (Fig. 1b), producing corresponding variations in the mean firing rate and the spike count variability.

Rather than modeling the synaptic input and membrane potential explicitly, here we aim to describe the distribution of spike counts statistically. Whereas the Poisson model has a single mean parameter λ , the COM-Poisson model has two parameters λ and ν that, together, determine both the spike count mean and variance. $\nu < 1$ results in over-dispersion where the variance of spike counts is greater than the mean, while $\nu > 1$ results in under-dispersion where the variance is less than the mean (Fig. 2a). Several models have been used to describe either under-dispersion (*e.g.* binomial) or overdispersion (*e.g.* negative binomial) alone (Fig. 2b). However, the ability to model both regimes may be a useful property of COM-Poisson models, since, empirically, neurons tend to show both types of behavior (Fig. 2c).

3.1 COM-Poisson models of tuning curves

To illustrate how inferring dispersion explicitly can provide another dimension to neural coding, we fit tuning curve models to simulated, non-Poisson spike counts. Figure 3 shows spike counts simulated from three COM-Poisson tuning curve models along with Bayesian model fits (samples from the posterior) for Poisson and COM-Poisson regression. In each case, the Poisson model does recover the tuning curve for mean spike counts. However, since the Poisson model does not have an explicit representation of variability, the predicted Fano Factor, the ratio of the spike count variance to the spike count mean, is always 1. In contrast, the COM-Poisson model allows explicit inference of the Fano Factor, and accurately recovers the true Fano Factor for under-dispersed (Fig. 3a) and over-dispersed (Fig. 3b) spiking, as well as for a neuron that has both under- and over-dispersion depending on the stimulus (Fig. 3c). For the examples show here, we use a Fourier basis and stimulate group-level dispersion where the relationship between the mean and variance is not fixed, but depends on the stimulus. COM-Poisson models with a constant dispersion parameter, on the other hand, are only be able to describe data that is either under- or over-dispersed, not both.

In experimental data, we find that tuning curve estimation with COM-Poisson noise models is more accurate than the Poisson models, as well as, negative binomial models frequently used to describe over-dispersed responses (Fig. 4). In data from MT/V5 recorded during presentation of moving sine-wave grating stimuli, we find that, for some neurons, the dispersion can be both under- and over-dispersed depending on the stimulus (Fig. 4a). We also find that the pattern of dispersion can be negatively correlated with the mean response (Fig. 4b). Note that, in general, a constant dispersion parameter means that the mean-variance relationship (Fig. 2b) is fixed. Curvature in the mean-variance relationship allows the Fano Factor to vary as a function of the stimulus but in a highly constrained way. A single, fixed mean-variance



Fig. 3 Comparison of Bayesian tuning curve estimates for simulated tuning curves with **a**) under-dispersion, **b**) over-dispersion, and **c**) both under- and over-dispersion. Spike counts (16 trials in each of 16 directions) were generated by sampling from a Conway-Maxwell-Poisson model then fit using Bayesian tuning curve estimation. Posterior samples are shown for the Poisson model (*blue*) and the Conway-Maxwell-Poisson model (*red*). Dots in the top plots denote spike counts on single trials, jittered for clarity. The *middle plots* show the Fano factor as a function of stimulus direction. *Error bars* denote

relationship appears to be insufficient to describe patterns of spike count variability found experimentally. The examples here highlight the cases where a COM-Poisson model with group-level dispersion, by allowing both over- and underdispersion and flexible mean-variance relationships, may provide a more accurate account of the underlying structure of the spike counts.

Across the n=56 neurons in these data we find that the COM-Poisson model with a constant dispersion parameter is more accurate (Fig. 5a) than both the Poisson model ($26\pm6\%$, two-tailed, paired t-test on the cross-validated log likelihood ratio, p < 0.001) and negative binomial model (1.2±0.002 %, p=0.002). Additionally, the COM-Poisson model with group-level dispersion out-performs the model with constant dispersion parameter (0.4 \pm 0.001 %, p<0.001). In this dataset, the largest improvement, compared to the Poisson model, appears to be due to allowing over-dispersion, but allowing under-dispersion and flexible mean-variance relationships also improve model performance slightly. Allowing deviations from the Poisson model also improves reconstruction of Fano Factors (Fig. 5b). In this case, allowing over-dispersion improves the Fano Factor RMSE relative to the Poisson model by 2 % with the negative binomial model, 34 % with the COM-Poisson, 40 % with the COM-Poisson model with group-level dispersion. Here, since the majority of neurons in this dataset are over-dispersed with convex meanvariance relationships, the negative binomial model performs quite well. However, as expected, the COM-

empirical Fano Factor estimates with inter-quartile range estimates using Bayesian bootstrapping. The *solid black curves* denote the ground truth values; however, the predicted Fano Factor for the Poisson model (*blue*) is always 1. The *bottom plots* show the posterior predictive distribution for the tuning curve along with the mean (*solid*) \pm 1 standard deviation (*dashed*). Note that the Conway-Maxwell-Poisson model can accurately describe both under- and over-dispersion, and, by allowing ν to vary as a function of stimulus direction, both types of dispersion can be modeled in the same neuron

Poisson model with group-level dispersion does tend to perform better for those cases where spike counts are less-dispersed (Fig. 5c, r=-0.16, p=0.22) and for neurons where the Fano Factor and mean response as a function of the stimulus are not strongly correlated (Fig. 5d, r=-0.41, p=0.002).

For the data here we also compare the accuracy of the COM-Poisson models to a recently described generalized count model that also allows both over- and underdispersion (Gao et al. 2015). The generalized count model contains the Poisson, negative binomial, and COM-Poisson models as special cases, and, might generally be expected to out-perform the other models with fixed meanvariance relationships (see Methods). However, in this case, we find that both the negative binomial $(10\pm8\%)$, p=0.1) and COM-Poisson models (10±7 %, p=0.5) outperform the generalized count model. One explanation for these results may be that, with limited data (3–12 trials per stimulus), the constraints on the mean-variance relationships imposed by the negative binomial and COM-Poisson models are beneficial. Allowing too much flexibility in the mean-variance relationship fit by the generalized count model is likely to lead to over-fitting, even with the smoothing constraints implemented here.

3.2 Dispersion affects standard errors

In the examples above we show how Poisson models tend to provide accurate estimates of the mean, even



Fig. 4 Model fits for four different Bayesian tuning curve models to data from two typical MT neurons in response to moving sine-wave gratings. **a**, **b**) Spike rasters in response to the moving sine-wave grating stimuli (50-700ms after stimulus onset). *Colors* denote different grating directions. C,D) Tuning curve and Fano factor for the different Bayesian models. Individual colored lines denote samples from the

when the data is non-Poisson. Poisson regression is consistent when the data is not equi-dispersed. However, one aspect of Poisson models that is often over-looked in neuroscience is that, when data is non-Poisson, standard errors and confidence intervals under the Poisson model are incorrect (Cameron and Trivedi 2001). As with any model, it is important to consider what happens when the assumptions are not met.

For canonical Poisson regression

 $y_i \sim Poisson(\exp(x_i^T \beta)).$

And the standard errors of the maximum likelihood estimate $\hat{\beta}$ follow from the covariance matrix

$$\operatorname{var}\left(\hat{\beta}\middle|y\right) = \left(\sum_{i} x_{i} x_{i}^{T} \exp\left(x_{i}^{T} \hat{\beta}\right)\right)^{-1}.$$

posterior distribution for the Poisson, Negative Binomial, Conway-Maxwell Poisson with constant dispersion, and COM-Poisson with group-level dispersion, respectively. For the Fano factor, dots and *error bars* denote the median and inter-quartile range estimated from the data using Bayesian bootstraping

However, for data that is not equi-dispersed a correction is needed: most commonly,

$$\operatorname{var}(\hat{\beta}|y) = \alpha \left(\sum_{i} x_{i} x_{i}^{T} \exp\left(x_{i}^{T} \hat{\beta}\right)\right)^{-1}$$

Under this correction, the dispersion is assumed fixed, with the dispersion parameter α estimated by

$$\hat{\alpha} = \frac{\sum_{i} \left(y_i - \hat{y}_i \right)^2 / \hat{y}_i}{(n-k)}$$

where *n* is the number of observations and *k* is the number of covariates. This adjustment results in what is sometimes known as a quasi-Poisson or quasi-Maximum Likelihood (QML) model where the mean count (determined by β) is estimated by traditional Poisson maximum likelihood estimation, but over- or under-dispersion can be accounted for during prediction and inference by estimating a dispersion parameter α post-hoc (Gourieroux et al. 1984). Without this QML



С D 0.06 0.06 L_{Group COM}-pL_{NB} (Bits/Spike) 0.04 0.04 0.02 0.02 0 0 -0.02 -0.02 ά -0.5 0.5 ò à ò Fano Factor Fano-Mean Correlation

Fig. 5 Model performance on the *n*=56 neurons in nsa2004.6. **a**) Crossvalidated (20-fold) log likelihood ratio between a homogeneous Poisson model and tuning curve models with six different noise assumptions: Poisson, generalized count (GC), negative binomial (NB), negative binomial with group-level dispersion, Conway-Maxwell Poisson (COM-P), and Conway-Maxwell Poisson with group-level dispersion. *Error bars* denote SEM. **b**) Square-root of the mean squared error (RMSE) for reconstructing the Fano Factor under each of the six models - fit to all data (not cross-validated). *Bars* denote median. *Error bars* denote inter-quartile range. **c**) Log likelihood ratio between the COM-Poisson model with group-level dispersion and the negative binomial model as a function of Fano Factor (averaged across stimulus conditions). The negative binomial model only captures over-dispersion,

and the COM-Poisson model tends to be better for smaller Fano Factors. **d**) Log likelihood ratio between the COM-Poisson model with grouplevel dispersion and the negative binomial model as a function of the correlation between the Fano Factor and the mean rate for each stimulus condition. Under the negative binomial model, since the meanvariance relationship is convex, negative correlations between the Fano Factor and mean spike count are not possible. The COM-Poisson model with group-level dispersion, on the other hand, allows more flexible mean-variance relationships and tends to outperform the Negative Binomial model when the Fano Factor and mean spike count are uncorrelated or negatively correlated. *Black line* denotes linear fit to the data. *Dashed curves* denote 95 % confidence intervals

correction Poisson regression will tend to be overly-confident when data is over-dispersed ($\alpha > 1$) and under-confident when data is under-dispersed ($\alpha < 1$).

Consider estimating a single parameter $\hat{\lambda}$ to describe n count observations with unknown mean count λ . Since the Poisson model assumes that the count variance is equal to the mean, the standard error for the Poisson model is $\sqrt{\hat{\lambda}/n}$ and for the QML model $\sqrt{\hat{\alpha}\hat{\lambda}/n}$ where, in this case, k=1 and $\hat{\alpha} = \hat{F}$, the Fano Factor from the sample variance and mean. Without the correction, the Poisson model will thus misestimate the standard error by a factor of $\sqrt{\hat{F}}$, even as the number of observations increases. That is, with F = 4 the standard error under the Poisson model is half what it should be, while with F = 0.25 the standard error is twice what it should be.

Many experiments in neuroscience rely on comparing the parameters of Poisson regression models, *e.g.* tuning width before and after adaptation or preferred direction over time. Since misspecification of the noise model can result in overor under-confidence, model-based results that fail to correct for non-Poisson spiking may need a second look, even when standard errors or confidence intervals are determined by Bayesian methods or bootstrapping (Cronin et al. 2010; Stevenson et al. 2011). Note for instance, that in the examples



Fig. 6 An illustration of how the choice of noise model affects the estimation of standard errors. In this case, we consider estimating a single mean spike count under different Fano Factors. Sampling from COM-Poisson distributions with fixed mean and different variances, we find that both the negative binomial and COM-Poisson model are able to accurately track the true standard error (*gray curve*, QML) for overdispersed data, but the negative binomial model is unable to track standard errors for under-dispersed data. For comparison, when the population mean and variance are known, the Poisson model produces a fixed estimate of standard error (*black line*) that does not take dispersion into account. This problem can be remedied by using the Quasi-Maximum Likelihood (QML) approach to calculate standard errors that take dispersion into account post-hoc (*gray curve*)

above, the samples from the posterior are less variable for the COM-Poisson model than for the Poisson model when the spike counts are under-dispersed (Fig. 3a) and *vice versa* when spike counts are over-dispersed (Fig. 3b). In general, by allowing the spike count variance to explicitly differ from the mean, the COM-Poisson model produces confidence/ credibility intervals that take dispersion into account.

To illustrate this point empirically, we sample COM-Poisson random variables (n=100) with a fixed mean E[y](5) and different variances var[y] (Fig. 6). If the moments were known exactly, using the equations above, the Poisson model should have constant standard errors given by $\sqrt{E[y]/n}$ even when the variance changes, while the QML model yields corrected standard errors $\sqrt{var[y]/n}$. The standard errors under the negative binomial and COM-Poisson models can be calculated by evaluating the Hessian at the maximum likelihood estimate or, as in the case shown here, by calculating the standard deviation of samples from the posterior. When estimating a single unknown mean from COM-Poisson samples, we find that the COM-Poisson credibility intervals match the intervals from QML (Fig. 6). The negative binomial model also takes dispersion into account, to some extent. In this case, the credibility intervals match for over-dispersed data, but not for under-dispersed. In the more general case, however, we have shown that dispersion may not even be fixed, suggesting that group/observation-level dispersion estimates may be necessary even with QML.

3.3 COM-Poisson models of PSTHs

Finally, to illustrate the flexibility of COM-Poisson regression for spike count data, we apply these models to estimating peristimulus time histograms (PSTHs). Just as modeling dispersion as a function of stimulus parameters can provide more accurate descriptions of tuning curves, modeling dispersion as a function of time may provide more accurate descriptions of PSTHs. Here, rather than using a Fourier basis, we use Bsplines with equally spaced knots (see Methods). As with tuning curves, modeling group-level dispersion allows us to fit a wider range of both simulated and experimental data.

In simulation, modeling dispersion explicitly and at the group-level allows us to fit cases where there is both overand under-dispersion in the same neuron (Fig. 7a), as well as cases where the dispersion changes without changes in the



Fig. 7 COM-Poisson PSTH fits for two simulated neurons. Here we use the COM-Poisson model with group-level dispersion to track both the PSTH mean and variance over time. a) Illustrates a case where the rate decreases and the Fano Factor increases to generate both over- and underdispersion in the same neuron. b) Illustrates a case where the mean is constant, but the variance switches from equi-dispersed to underdispersed. The *red and blue curves* indicate Bayes estimators for the

COM-Poisson model with variable (group-level) and fixed dispersion parameter, respectively (n=4, m=8). Dashed lines denote 95 % credibility intervals. Solid black lines denote ground-truth, simulated values. Gray bars in the upper panels denote empirical mean spike counts along with standard deviation. Gray dots in the lower panels denote empirical Fano factors



Fig. 8 COM-Poisson PSTH fits for two MT neurons responding to incoherent moving dot stimuli generated from either the same seed (*left*) or a variable seed (*right*) on repeated trials. a) Neuron nsa2009.1.ec082, and b) neuron nsa2009.1.ec094 show similar reduction in Fano Factor following stimulus onset. *Red and blue curves* indicate Bayes estimators for the COM-Poisson model with variable and fixed dispersion parameter, respectively. *Gray bars and dots* denote the empirical mean

mean (Fig. 7b). These two cases illustrate how dispersion can act as an alternative dimension for the neural code.

Similarly, in experimental data the COM-Poisson model allows accurate tracking of both the mean spike count and its variance (Fig. 8). Here we find that, in general, models with group-level dispersion are necessary to account for changes in dispersion over time. Models with constant dispersion parameters are able to track changes in Fano Factor to some extent (Fig. 8a), but the data are generally better fit when the meanvariance relationship is flexible.



Fig. 9 Model performance for n = 46 PSTH responses in nsa2009.1. a) Cross-validated (4-fold) log likelihood ratio between a homogeneous Poisson model and PSTH models with six different noise assumptions: Poisson, generalized count (GC), negative binomial (NB), negative binomial with group-level dispersion, Conway-Maxwell Poisson (COM-P), and Conway-Maxwell Poisson with group-level dispersion. *Error bars* denote SEM. b) Square-root of the mean squared error (RMSE) for reconstructing the Fano Factor under each of the six models - fit to all data (cross-validated). *Bars* denote median. *Error bars* denote inter-quartile range



(1ms bins) and Fano Factor (20ms bins), and *dashed lines* denote 95 % credibility intervals for the variable dispersion model. Note that both the fixed and variable dispersion parameter models accurately track the mean. However, the fixed dispersion parameter model fails to track the Fano Factor in many cases, suggesting that the relationship between mean and variance may not be fixed, or at least deviates from the relationship under the COM-Poisson distribution

Across the n = 46 neurons (with firing rates >0.5 Hz) in the PSTH dataset we compare the same six models (Fig. 9) as with the tuning curves. Here, with 20ms bins, the responses are largely under-dispersed, and we find that the COM-Poisson models with constant dispersion parameters out-perform both the Poisson (60±33 %, one-tailed, paired t-test on the crossvalidated log likelihood ratio, p=0.05) and negative binomial models (59 \pm 34 %, p=0.07). The generalized count models, since they are able to capture under-dispersion, also outperform the Poisson (57 \pm 32 %, p=0.01) and negative binomial models ($55\pm34\%$, p=0.01), as well as the COM-Poisson model with constant dispersion parameter $(3\pm3\%, p=0.3)$. However, the COM-Poisson model with group-level dispersion slightly out-performs the generalized count model (8±3 %, p=0.38). In this case, allowing under- and over-dispersion improves the Fano Factor reconstruction (RMSE) compared to the Poisson model by 19 % for the generalized count model, 0.1 % for the negative binomial, 35 % for the COM-Poisson, and 37 % for the COM-Poisson with group-level dispersion.

4 Discussion

Inhomogeneous Poisson models that aim to describe spike rate as a function of stimuli are often unable to effectively capture the full range of spike count variability observed in experimental data. Observed spike counts can be both overand under-dispersed and can have mean-variance relationships that differ from common canonical count models. Here we illustrate how Conway-Maxwell-Poisson models with group-level dispersion provide an accurate description of spike count statistics in these non-Poisson regimes. In tuning curve and PSTH data from MT/V5 we find that COM-Poisson models out-perform Poisson and negative binomial models and naturally correct estimates of standard error for dispersion in the observations. In general, for both tuning curves and PSTHs, the COM-Poisson approach can accurately and flexibly track both the spike count mean and variance.

Model-based methods have several important advantages over traditional, descriptive measures of spike count variability, such as the sample Fano Factor. Most importantly, model-based methods are generally more efficient than descriptive approaches (Cronin et al. 2010; Kass et al. 2003); that is, they can produce better estimates with smaller amounts of data. In the cases shown here, for both tuning curves and PSTHs, using COM-Poisson regression with a smooth basis allows us to leverage the fact that spike count distributions tend to vary smoothly over stimuli and time, where similar stimuli can be expected to result in similar spike count distributions. Another challenge of describing variability with sample Fano Factors is that the results depend heavily on bin-size (Teich 1989). Estimates from the COM-Poisson model still depend on binsize, but we can reduce our uncertainty about the dispersion in a given bin by incorporating estimates from nearby bins.

An additional advantage of model-based methods is that they allow principled model comparison (Cronin et al. 2010; Kottas et al. 2012). Previous studies have used Bernoulli models to describe under-dispersion (DeWeese et al. 2003) and negative binomial (Scott and Pillow 2012) or mixture models (Goris et al. 2014; Moshitch and Nelken 2014; Shidara et al. 2005; Wiener and Richmond 2003) to describe over-dispersion. Here we demonstrate how the COM-Poisson model can be directly compared to several alternative noise models: Poisson, negative binomial, and generalized count models. By allowing both under- and over-dispersion the COM-Poisson models generally provides a more accurate description of the broader range of variability present in neural responses. Although the COM-Poisson model is a special case of the generalized count model, it is among the simplest models that allows both under- and over-dispersion. Here we find that allowing additional flexibility in the meanvariance relationship by using the full generalized count model only provides a slightly more accurate fits. By modeling grouplevel dispersion we can add an additional level of flexibility where the mean-variance relationship need not be fixed across stimuli/time. Just as we can compare different noise models, including group-level dispersion allows us to compare patterns of dispersion. For instance, by comparing COM-Poisson models with m = 0 and m>0, we can examine whether a specific pattern of variability across stimuli is consistent with the COM-Poisson mean-variance relationship or whether additional parameters are necessary to capture these trends.

In some sense, the COM-Poisson models presented here only provide a summary of spike count statistics without an explanation for spike count variability. As mentioned above, many extensions of the inhomogeneous Poisson model are able to account for non-Poisson variability by including intrinsic covariates-variables that differ from trial-to-trial even with a fixed stimulus (Masquelier 2013). While the COM-Poisson approach simply describes variability, single-trial models can begin to explain the sources of trial-to-trial variability. Similarly, rather than modeling spike counts it is often useful to model detailed spike timing. As binsizes become smaller, one challenge for count models is that all observations begin to look like Bernoulli random variables. In this limit, the difference between the Poisson and COM-Poisson distributions is negligible and it is often more sensible to use point-process formulations, either Poisson or non-Poisson (Barbieri et al. 2001; Kass and Ventura 2001). Recently, point process models with power-law meanvariance relationships have been developed that explicitly allow both over- and under-dispersed spike counts (Koyama 2015), and there has also been some preliminary work developing a COM-Poisson process (Zhu et al. 2015). Although the COM-Poisson approach presented here is a flexible way to summarize the spike count statistics of tuning curves and PSTHs, it may not always be the desired level of description.

Spike count variability appears to be an increasingly informative dimension for understanding the neural code (Hussar and Pasternak 2010; Mandelblat-Cerf et al. 2009; Scaglione et al. 2011). Here we have shown that in the absence of other explanatory variables, COM-Poisson models with group-level dispersion can provide a flexible, efficient description of spike count variability. These models may, thus, be useful tools for linking behavioral variables, such as decision making or stimulus/movement onset, to spike count variability, even with limited data. Although here we have focused on characterizing single neuron variability, these and similar methods may ultimately provide insight into population decoding (Averbeck et al. 2006) and how the variability of multiple neurons may be related (Cohen and Kohn 2011).

Acknowledgments Thanks to Mike DeWeese, Heather Read, and Monty Escabi for helpful comments and discussions. IHS was supported by an NSF Computing Innovation Fellowship (NSF-0937060 CIF-D-018).

Compliance with ethical standards

Conflict of interest The author declares that he has no conflict of interest.

References

Amarasingham, A., Chen, T.-L., Geman, S., Harrison, M. T., & Sheinberg, D. L. (2006). Spike count reliability and the Poisson hypothesis. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 26*(3), 801–809. doi:10.1523/ JNEUROSCI.2948-05.2006.

- Arieli, A., Sterkin, A., Grinvald, A., & Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*, 273(5283), 1868–1871. doi:10.1126/ science.273.5283.1868.
- Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews. Neuroscience*, 7(5), 358–366.
- Azouz, R., & Gray, C. M. (1999). Cellular mechanisms contributing to response variability of cortical neurons in vivo. *Journal of Neuroscience*, 19(6), 2209.
- Bair, W., & Koch, C. (1996). Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Computation*, 8(6), 1185–1202. doi:10.1162/neco.1996.8.6.1185.
- Barbieri, R., Quirk, M. C., Frank, L. M., Wilson, M. A., & Brown, E. N. (2001). Construction and analysis of non-Poisson stimulus-response models of neural spiking activity. *Journal of Neuroscience Methods*, 105(1), 25–37. doi:10.1016/S0165-0270(00)00344-7.
- Berry, M. J., & Meister, M. (1998). Refractoriness and neural precision. The Journal of Neuroscience : The official Journal of the Society for Neuroscience, 18(6), 2200–2211.
- Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences*, 94(10), 5411–5416. doi:10.1073/pnas.94.10. 5411.
- Brette, R., & Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, 94(5), 3637–3642. doi:10.1152/jn.00686.2005.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, *12*(12), 4745.
- Brown, E., Barbieri, R., Eden, U., & Frank, L. (2003). Likelihood methods for neural data analysis. In J. Feng (Ed.), *Computational Neuroscience: A comprehensive approach* (pp. 253–286). London: Chapman and Hall.
- Cameron, A. C., & Trivedi, P. K. (2001). Essentials of count data regression. In A companion to theoretical econometrics (Vol. 331). Blackwell Publishing Ltd.
- Carandini, M. (2004). Amplification of trial-to-trial response variability by neurons in visual cortex. *PLoS Biology*, 2(9), E264. doi:10.1371/ journal.pbio.0020264.
- Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G., & Shenoy, K. V. (2006). Neural variability in premotor cortex provides a signature of motor preparation. *Journal of Neuroscience*, 26(14), 3697.
- Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience*, 13(3), 369–378. doi:10.1038/nn.2501.
- Churchland, A. K., Kiani, R., Chaudhuri, R., Wang, X. J., Pouget, A., & Shadlen, M. N. (2011). Variance as a signature of neural computations during decision making. *Neuron*, 69(4), 818–831. doi:10. 1016/j.neuron.2010.12.037.
- Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7), 811–819. doi:10.1038/nn. 2842.
- Cronin, B., Stevenson, I. H., Sur, M., & Kording, K. P. (2010). Hierarchical Bayesian modeling and Markov Chain Monte Carlo sampling for tuning-curve analysis. *Journal of Neurophysiology*, 103(1), 591.
- Czanner, G., Eden, U. T., Wirth, S., Yanike, M., Suzuki, W. A., & Brown, E. N. (2008). Analysis of between-trial and within-trial neural spiking dynamics. *Journal of Neurophysiology*, 99(5), 2672–2693. doi: 10.1152/jn.00343.2007.
- De Boor, C. (1978). A practical guide to splines. Applied mathematical sciences 27. Verlag: Springer.

- del Castillo, J., & Pérez-Casany, M. (2005). Overdispersed and underdispersed Poisson generalizations. *Journal of Statistical Planning and Inference*, 134(2), 486–500. doi:10.1016/j.jspi.2004. 04.019.
- Deweese, M. R., & Zador, A. M. (2004). Shared and private variability in the auditory cortex. *Journal of Neurophysiology*, 92(3), 1840–1855. doi:10.1152/jn.00197.2004.
- DeWeese, M. R., Wehr, M., & Zador, A. M. (2003). Binary spiking in auditory cortex. *The Journal of Neuroscience : The Official Journal* of the Society for Neuroscience, 23(21), 7940–7949.
- Dimatteo, I., Genovese, C. R., & Kass, R. E. (2001). Bayesian curvefitting with free-knot splines. *Biometrika*, 88(4), 1055–1071. doi:10. 1093/biomet/88.4.1055.
- Eden, U. T., & Kramer, M. a. (2010). Drawing inferences from Fano factor calculations. *Journal of Neuroscience Methods*, 190(1), 149–152. doi:10.1016/j.jneumeth.2010.04.012.
- Eden, U. T., Frank, L. M., Barbieri, R., Solo, V., & Brown, E. N. (2004). Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Computation*, 16(5), 971–998. doi:10.1162/ 089976604773135069.
- Ermentrout, G. B., Galán, R. F., & Urban, N. N. (2008). Reliability, synchrony and noise. *Trends in Neurosciences*, 31(8), 428–434.
- Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews. Neuroscience*, 9(4), 292–303. doi:10.1038/nm2258.
- Gao, Y., Buesing, L., Shenoy, K. V, & Cunningham, J. P. (2015). Highdimensional neural spike train analysis with generalized count linear dynamical systems. In *NIPS*.
- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4), 1360–1383. http://projecteuclid.org/euclid.aoas/1231424214. Accessed 30 July 2015.
- Goris, R. L. T., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*, 17(6), 858–65. doi:10. 1038/nn.3711.
- Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: applications to Poisson models. *Econometrica*, 52(3), 701–720.
- Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G., & Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature*, 424(6948), 552–556.
- Hoffman, M., & Gelman, A. (2014). The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15, 30.
- Hoyer, P. O., Hyvarinen, A., & Hyvärinen, A. (2003). Interpreting neural response variability as Monte Carlo sampling of the posterior (Vol. 15, pp. 277–284,). MIT Press.
- Hussar, C., & Pasternak, T. (2010). Trial-to-trial variability of the prefrontal neurons reveals the nature of their engagement in a motion discrimination task. *Proceedings of the National Academy of Sciences* of the United States of America, 107(50), 21842–7. doi:10.1073/ pnas.1009956107.
- Kadane, J. B., Shmueli, G., Minka, T. P., Borle, S., & Boatwright, P. (2006). Conjugate analysis of the Conway-Maxwell-Poisson distribution. *Bayesian Analysis*, 1(2), 363–374. http://projecteuclid.org/ euclid.ba/1340371067. Accessed 11 December 2015.
- Kara, P., Reinagel, P., & Reid, R. C. (2000). Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron*, 27(3), 635–646. doi:10.1016/S0896-6273(00)00072-6.
- Kass, R. E., & Ventura, V. (2001). A spike-train probability model. *Neural Computation*, 13(8), 1713–1720.
- Kass, R. E., Ventura, V., & Cai, C. (2003). Statistical smoothing of neuronal data. *Network (Bristol, England)*, 14(1), 5–15. http://www.ncbi.nlm.nih.gov/pubmed/12613549. Accessed 29 October 2015.

- Kaufman, C. G., Ventura, V., & Kass, R. E. (2005). Spline-based nonparametric regression for periodic functions and its application to directional tuning of neurons, 24(14), 2255–2265.
- Keat, J., Reinagel, P., Reid, R. C., & Meister, M. (2001). Predicting every spike a model for the responses of visual neurons. *Neuron*, 30(3), 803–817.
- Kelly, R. C., Smith, M. A., Kass, R. E., & Lee, T. S. (2010). Local field potentials indicate network state and account for neuronal response variability. *Journal of Computational Neuroscience*, 29(3), 567– 579. doi:10.1007/s10827-009-0208-9.
- Kohn, A., & Movshon, J. A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. *Neuron*, 39(4), 681–691. doi:10. 1016/S0896-6273(03)00438-0.
- Kottas, A., Behseta, S., Moorman, D. E., Poynor, V., & Olson, C. R. (2012). Bayesian nonparametric analysis of neuronal intensity rates. *Journal of Neuroscience Methods*, 203(1), 241–53. doi:10.1016/j. jneumeth.2011.09.017.
- Koyama, S. (2015). On the spike train variability characterized by variance-to-mean power relationship. *Neural Computation*, 27(7), 1530–48. doi:10.1162/NECO a 00748.
- Lansky, P., & Vaillant, J. (2000). Stochastic model of the overdispersion in place cell discharge. *Biosystems*, 58(1), 27–32.
- Lee, D., Port, N. L., Kruse, W., & Georgopoulos, A. P. (1998). Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *Journal of Neuroscience*, 18(3), 1161– 1170. http://www.jneurosci.org/content/18/3/1161.abstract?ijkey= bd8ccb3d3a84873b46d8a3414a579c19586b02c6&keytype2=tf_ ipsecsha. Accessed 11 November 2015.
- Maimon, G., & Assad, J. a. (2009). Beyond Poisson: increased spike-time regularity across primate parietal cortex. *Neuron*, 62(3), 426–440. doi:10.1016/j.neuron.2009.03.021.
- Mainen, Z. F., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, 268(5216), 1503–1506.
- Mandelblat-Cerf, Y., Paz, R., & Vaadia, E. (2009). Trial-to-trial variability of single cells in motor cortices is dynamically modified during visuomotor adaptation. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 29*(48), 15053–62. doi:10. 1523/JNEUROSCI.3011-09.2009.
- Masquelier, T. (2013). Neural variability, or lack thereof. Frontiers in Computational Neuroscience, 7(February), 7. doi:10.3389/fncom. 2013.00007.
- Minka, T. T. P., Shmueli, G., Kadane, J. B. J., Borle, S., & Boatwright, P. (2003). Computing with the COM-Poisson distribution., *PA: Department of*, (776). http://lib.stat.cmu.edu/cmu-stats/tr/tr776/ tr776.pdf
- Moshitch, D., & Nelken, I. (2014). Using Tweedie distributions for fitting spike count data. *Journal of Neuroscience Methods*, 225, 13–28. doi:10.1016/j.jneumeth.2014.01.004.
- Nawrot, M. P. (2010). Analysis and interpretation of interval and count variability in neural spike trains. In *Analysis of parallel spike trains* (pp. 37–58). Springer.
- Paninski, L., Ahmadian, Y., Ferreira, D. G., Koyama, S., Rahnama Rad, K., Vidne, M., et al. (2010). A new look at state-space models for neural data. *Journal of Computational Neuroscience*, 29(1), 107– 126. doi:10.1007/s10827-009-0179-x.
- Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P., & Chichilnisky, E. J. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *Journal of Neuroscience*, 25(47), 11003–11013.
- Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., & Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207), 995–999.
- Reich, D. S., Victor, J. D., Knight, B. W., Ozaki, T., & Kaplan, E. (1997). Response variability and timing precision of neuronal spike trains in vivo. *Journal of Neurophysiology*, 77(5), 2836–41. http://jn.

physiology.org/content/77/5/2836.abstract. Accessed 12 November 2015.

- Rubin, D. B. (1981). The Bayesian bootstrap. *The Annals of Statistics*, 9(1), 130–134. http://projecteuclid.org/euclid.aos/1176345338. Accessed 30 October 2015.
- Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. *Journal of Neurophysiology*, 76(4), 2790–2793.
- Scaglione, A., Moxon, K. A., Aguilar, J., & Foffani, G. (2011). Trial-to-trial variability in the responses of neurons carries information about stimulus location in the rat whisker thalamus. Proceedings of the National Academy of Sciences of the United States of America, 108(36), 14956–61. doi:10.1073/ pnas.1103168108.
- Schölvinck, M. L., Saleem, A. B., Benucci, A., Harris, K. D., & Carandini, M. (2015). Cortical state determines global variability and correlations in visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 35(1), 170–8. doi:10.1523/JNEUROSCI.4994-13.2015.
- Scott, J., & Pillow, J. W. (2012). Fully Bayesian inference for neural models with negative-binomial spiking. In Advances in Neural Information Processing Systems (pp. 1898–1906). http://papers. nips.cc/paper/4567-fully-bayesian-inference-for-neural-modelswith-negative-binomial-spiking. Accessed 27 July 2015.
- Sellers, K. F., & Shmueli, G. (2009). A regression model for count data with observation-level dispersion. In 24th International Workshop on Statistical Modelling (IWSM).
- Sellers, K. F., & Shmueli, G. (2010). A flexible regression model for count data. *The Annals of Applied Statistics*, 943–961.
- Sellers, K. F., & Shmueli, G. (2013). Data dispersion: Now you see it... now you don't. *Communications in Statistics-Theory and Methods*, 42(17), 3134–3147.
- Sellers, K. F., Borle, S., & Shmueli, G. (2012). The COM-Poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry*, 28(2), 104–116.
- Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10), 3870–3896.
- Shidara, M., Mizuhiki, T., & Richmond, B. J. (2005). Neuronal firing in anterior cingulate neurons changes modes across trials in single states of multitrial reward schedules. *Experimental Brain Research*, 163(2), 242–5. doi:10.1007/s00221-005-2232-y.
- Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., et al. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Computational Biology*, 5(7), e1000433. doi:10.1371/journal.pcbi.1000433.
- Shmueli, G., Minka, T., Kadane, J., Borle, S., & Boatwright, P. (2004). A useful distribution for fitting discrete data:revival of the conway-Maxwell Poisson distribution. *Applied Statistic*, 54(1), 127–142.
- Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 13(1), 334–350.
- *Stan: A C++ Library for probability and sampling, version 2.8.0.* (2015). Retrieved from http://mc-stan.org/
- Stein, R. B., Gossen, E. R., & Jones, K. E. (2005). Neuronal variability: noise or part of the signal? *Nature Reviews. Neuroscience*, 6(5), 389–397. doi:10.1038/nrn1668.
- Stevenson, I. H., Rebesco, J. M., Miller, L. E., & Körding, K. P. (2008). Inferring functional connections between neurons. *Current Opinion* in Neurobiology, 18(6), 582–588.
- Stevenson, I. H., Cherian, A., London, B. M., Sachs, N. A., Lindberg, E., Reimer, J., et al. (2011). Statistical assessment of the stability of neural movement representations. *Journal of Neurophysiology*, *106*(2), 764–774. doi:10.1152/jn.00626.2010.

- Taouali, W., Benvenuti, G., Wallisch, P., Chavane, F., & Perrinet, L. U. (2016). Testing the odds of inherent vs. observed overdispersion in neural spike counts. *Journal of Neurophysiology*, 115(1), 434–44. doi:10.1152/jn.00194.2015.
- Teich, M. C. (1989). Fractal character of the auditory neural spike train. *IEEE Transactions on Bio-Medical Engineering*, 36(1), 150–60. doi:10.1109/10.16460.
- Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, 93(2), 1074–1089.
- Uzzell, V. J., & Chichilnisky, E. J. (2004). Precision of spike trains in primate retinal ganglion cells. *Journal of Neurophysiology*, 92(2), 780–789. doi:10.1152/jn.01171.2003.
- van Steveninck, R. R. R., Lewen, G. D., Strong, S. P., Koberle, R., & Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science*, 275(5307), 1805–1808.
- Vogel, A., Hennig, R. M., & Ronacher, B. (2005). Increase of neuronal response variability at higher processing levels as revealed by

simultaneous recordings. *Journal of Neurophysiology*, 93(6), 3548–59. doi:10.1152/jn.01288.2004.

- Werner, G., & Mountcastle, V. B. (1963). The variability of central neural activity in a sensory system, and its implications for the central reflection of sensory events. *Journal of Neurophysiology*, 26(6), 958–977.
- Wiener, M. C., & Richmond, B. J. (2003). Decoding spike trains instant by instant using order statistics and the mixture-of-poissons model. *Journal of Neuroscience*, 23(6), 2394–2406. http://www.jneurosci. org/content/23/6/2394.full. Accessed 14 December 2015.
- Zador, A. (1998). Impact of synaptic unreliability on the information transmitted by spiking neurons. *Journal of Neurophysiology*, 79(3), 1219–1229.
- Zhao, M., & Iyengar, S. (2010). Nonconvergence in logistic and poisson models for neural spiking. *Neural Computation*, 22(5), 1231–1244.
- Zhu, L., Morris, D. S., Sellers, K. F., & Shmueli, G. (2015). Bridging the gap: a generalized stochastic process for count data. *Under Review*.