A universal probabilistic spike count model reveals ongoing modulations of neural variability

Anonymous Author(s)
Affiliation
Address
email

Abstract

Neural responses are variable: even under identical experimental conditions, single neuron and population responses typically differ from trial to trial and across time. Recent work has demonstrated that this variability has predictable structure, can be modulated by sensory input and behaviour, and bears critical signatures of the underlying network dynamics and computations. However, current methods for characterising neural variability are primarily geared towards sensory coding in the laboratory: they require trials with repeatable experimental stimuli and behavioural covariates. In addition, they make strong assumptions about the parametric form of variability, rely on assumption-free but data-inefficient histogram-based approaches, or are altogether ill-suited for capturing variability modulation by covariates. Here we present a universal probabilistic spike count model that eliminates these shortcomings. Our method builds on sparse Gaussian processes and can model arbitrary spike count distributions (SCDs) with flexible dependence on observed as well as latent covariates, using scalable variational inference to jointly infer the covariate-to-SCD mappings and latent trajectories in a data efficient way. Without requiring repeatable trials, it can flexibly capture covariate-dependent joint SCDs, and provide interpretable latent causes underlying the statistical dependencies between neurons. We apply the model to recordings from a canonical non-sensory neural population: head direction cells in the mouse. We find that variability in these cells defies a simple parametric relationship with mean spike count as assumed in standard models, its modulation by external covariates can be comparably strong to that of the mean firing rate, and slow low-dimensional latent factors explain away neural correlations. Our approach paves the way to understanding the mechanisms and computations underlying neural variability under naturalistic conditions, beyond the realm of sensory coding with repeatable stimuli.

1 Introduction

Classical analyses of neural coding are based on mean spike counts or neural firing rates. Indeed, some of the most paradigmatic examples of the neural code were discovered by regressing neural firing rates to particular sensory stimuli [1, 2] or behavioural covariates [3, 4, 5, 6] to characterize their tuning properties. However, neural spiking is generally not regular. Recordings from many cortical areas show significantly different activity patterns within and across identical trials [7], despite fixing experimentally controlled variables. This irregularity is also seen in continual neural recordings without trial structure [8]. The resulting variability has classically been characterised as 'Poisson', with a Fano factor (variance to mean ratio) of one [9], but experimental data also often exhibits significantly more [10, 8, 11, 12] and sometimes less [13, 14] variability, respectively referred to as over- or underdispersion. Moreover, experimental studies have revealed that neural variability generally depends on stimulus input and behaviour [15, 16, 17, 18], and exhibits structured shared

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

variability ('noise correlations') across neurons even after conditioning on such covariates. Such correlations can have important consequences for decoding information from neural population activity [19, 20, 21] and reveal key properties of the underlying circuit dynamics [22]. Moreover, theories of neural representations of uncertainty have assigned computational significance to variability as a signature of Bayesian inference [23, 24, 25, 26]. Thus, just as classical tuning curves for firing rates have been crucial for understanding some of the fundamental properties of the neural code, a principled statistical characterisation of neural variability, and its dependence on stimulus and behavioral covariates, is a key step towards understanding the dynamics of neural circuits and the computations they subserve.

The traditional approach to characterising neural variability has been pioneered in sensory areas, and relies on repeatable trial structure with a sufficiently large number of trials using identical stimulus and behavioral correlates [27, 15, 28]. Variability in this case can be quantified by simple summary statistics of spike counts across trials of the same condition. However, this approach does not readily generalise to more naturalistic conditions and to neural populations whose covariates cannot be precisely controlled and repeated in an experiment. This more general setting requires statistical methods that take into account temporal changes in covariates for predicting neural spiking in a principled way. Generalised linear model-based approaches are a popular choice for such a method [29], but they only model the dependence of mean firing rates on covariates – with changes in variability directly coupled to changes in the mean due to a fundamental underlying assumption of Poisson spiking statistics. Methods for inferring neural tuning and latent structure [30, 31] have conventionally used similarly restrictive parametric families for spike count distributions, and thus also cannot model changes in variability that are not 'just' a consequence of changes in means. Conversely, statistical models capable of capturing arbitrary single neuron count statistics, such as histogram-based approaches or copulas [32], do not incorporate dependencies on covariates.

Here we unify these separate approaches, resulting in a single framework for jointly inferring neural tuning, single neuron count statistics, neural correlations, and latent structure. Our semi-parametric approach leads to a universal count model for a given maximum count K, in the sense that we can model arbitrary distributions over the joint count space of size $(K+1)^N$ of N neurons. The trade-off between computational overhead and model expressivity is controlled by hyperparameters, with expressivity upper bounded by the true universal model. Our approach extends the idea of a universal binary count model [33] to a finite range of integer counts, while allowing flexible dependence on observed and latent covariates to model non-stationary neural activity and correlations. The flexibility reduces biases in previous models from restrictive assumptions in any of the model components. Scalability is maintained by leveraging sparse Gaussian processes [34] with mini-batching [35, 36] to handle the size of modern neural recordings.

After introducing the universal count model, we review basic statistical quantities used in neuroscience to characterize neural data, as well as goodness-of-fit measures for evaluating model fits. As our model is able to capture arbitrary single neuron statistics, we build on the Kolmogorov-Smirnov test to extend goodness-of-fit characterization to real data. After validating our method on synthetic data that cannot be captured by currently used methods, we apply the universal model to electrophysiological recordings from two distinct brain regions in mice that show significant tuning to the head direction of the animal [37]. We find that (1) neural activity tends to be more regular than common Poisson-like models at higher firing rates, and more irregular at low rates; (2) mean and variance of counts defy a simple parametric relationship imposed by parametric count distribution families; (3) variability modulation by behaviour can be comparable or even exceed that of the mean count or firing rate; (4) a two-dimensional latent trajectory varying on timescales of around a second is sufficient to explain away neural correlations but not the non-Poisson nature of single neuron variability. Finally, we discuss related work, limitations and proposed extensions of our model.

2 Universal count model

Neural spike count activity is formally equivalent to a multivariate time series of non-negative integers. In this paper, we study a neural population of N neurons, with T time bins of count data. Due to biological constraints of neurons, the spike count has some finite upper bound K, which here is the highest observed count. A universal count model is defined here as a probabilistic model that can capture any joint distribution over population counts $\boldsymbol{y} = \{y_n\}_1^N$. In particular, our model provides a prior over count distributions, which is universal if it has support over all possible joint

distributions. We denote a discrete probability distribution over counts 0 to K by a vector $\boldsymbol{\pi}$ of length $K+1$. Additionally, we denote the count activity over neurons n and at time step t by a matrix $Y \in [0, K]^{T \times N}$ with elements y_{tn}. Similarly, we denote the observed $X \in \mathbb{R}^{T \times D_x}$ and latent $Z \in \mathbb{R}^{T \times D_z}$ covariates (the range may change with topology [38]) with elements x_{td} and z_{tq} respectively. For other quantities below, we use the capital version to denote similar multidimensional concatenations.

2.1 Generative model

The model proposed here provides a prior over the joint count distribution of a population, $p(\Pi|X)$, similar to Dirichlet priors [33] but allowing non-parametric dependence on X. For each neuron, it consists of C Gaussian process (GP) priors, a basis expansion $\boldsymbol{\phi} : \mathbb{R}^C \to \mathbb{R}^{\tilde{C}}$, and a linear-softmax mapping

$$\begin{aligned} z_{tq} &\sim p(Z), \quad h_{cn}(\cdot) \sim \mathcal{GP}(0, k_{cn}(\cdot,\cdot)) \\ f_{cnt} &= h_{cn}(\boldsymbol{x}_t, \boldsymbol{z}_t) \\ \boldsymbol{\pi}_{nt} &= \text{softmax}(W_n\, \boldsymbol{\phi}(\boldsymbol{f}_{nt}) + \boldsymbol{b}_n) \\ y_{nt} &\sim \text{Multinomial}(\boldsymbol{\pi}_{nt}) \end{aligned} \tag{1}$$

where k_{nc} is the GP kernel function. The choice of prior $p(Z)$ is described in subsection 2.3, as it allows the model to capture neural and temporal correlations. To attain scalability, we use sparse GPs [34], which add additional input points X_u as inducing points.

Depending on the choice of channels C and the basis functions in $\boldsymbol{\phi}(h)$, we obtain an approximate universal prior on joint count distributions. The true universal prior is obtained for $C = K$ and $\boldsymbol{\phi}(\boldsymbol{f}) = \boldsymbol{f}$, which is computationally expensive when $N \times C \gg 1$. The use of non-parametric GP mappings with point estimates for parameters W and $\boldsymbol{b}$ leads to an overall semi-parametric model with parameters θ (see details in Appendix E). One controls the trade-off between model expressiveness and computational overhead through C and $\boldsymbol{\phi}$. Larger expansions $\boldsymbol{\phi}$ allow one to model count distributions more expressively with small C, e.g. the linear-exponential $\boldsymbol{\phi}(\boldsymbol{f}) = (f_1, e^{f_1}, f_2, e^{f_2} \ldots)$ covers a range of distributions including the truncated Poisson with only $C = 1$ (see subsection B.3).

2.2 Stochastic variational inference and learning

The approximate posterior for our model takes the form

$$q_{\theta,\varphi}(\Pi, Z|X) = q_\theta(\Pi|X, Z)\, q_\varphi(Z) \tag{2}$$

with φ the free variational parameters for latent states. The posterior $q_\theta(\Pi|X, Z)$ is evaluated from the sparse Gaussian process posterior $q(F|X, Z)$ and analytically intractable, so in practice it is represented by Monte Carlo samples. A lower bound on the log marginal likelihood can be optimized using stochastic variational inference [39], equivalent to minimizing the variational free energy

$$\mathcal{F}_{\theta,\varphi} = -\mathbb{E}_{Z\sim q_\varphi(Z)} \mathbb{E}_{\Pi\sim q_\theta(\Pi|X_\mathcal{D},Z)} \left[\log \frac{P(Y_\mathcal{D}|\Pi)\, P_\theta(\Pi|X_\mathcal{D}, Z)\, p_\theta(Z)}{q_\theta(\Pi|X_\mathcal{D}, Z) q_\varphi(Z)} \right] \tag{3}$$

with a multinomial distribution $P(Y_\mathcal{D}|\Pi)$, which leads to tractable terms (see subsection E.1). This objective allows us to infer the approximate posterior, while its negative value provides a lower bound on the marginal log likelihood [40]. By using the framework of reparameterized Lie groups [41], we can extend Z to non-Euclidean spaces, such as the ring or the torus, in a tractable manner [38]. We used Adam [42] to perform optimization, see details of implementation and model fitting in Appendix E.

2.3 Latent variables and correlations

Without latent variables, any distribution $P(Y|X)$ drawn from the prior has conditional independence structure, i.e. spike counts are independent across neurons and time conditioned on X. To model correlations in neural activity Y, we introduce additional latent variables Z. These variables induce neural correlations as they are shared, with the resulting marginal no longer factorized over neurons

$$P(Y|X) = \int \left(\prod_{n=1}^{N} \prod_{t}^{T} P_n(y_{tn}|\boldsymbol{x}_t, \boldsymbol{z}_t) \right) p_\theta(Z)\, \mathrm{d}Z \neq \prod_{n=1}^{N} \tilde{P}_n(\boldsymbol{y}_n|X) \tag{4}$$

Correlations in the temporal dimension can be captured through Z by placing temporal priors on the latent space. In particular, we use Markovian priors with learnable parameters (see Appendix E)

$$p_\theta(Z) = p_\theta(\boldsymbol{z}_1) \prod_{t=2}^{T} p_\theta(\boldsymbol{z}_t | \boldsymbol{z}_{t-1}) \tag{5}$$

In this model, Z can then be thought of as latent trajectories describing unobserved signals represented by the neural population. The model also quantifies intrinsic or private neuronal variability that cannot be explained away by regressors or shared latent variables through $P_n(y_{tn}|\boldsymbol{x}_t, \boldsymbol{z}_t)$. Combined with latent variables, our model can describe arbitrary joint count distribution and becomes universal.

2.4 Obtaining interpretable spike count statistics from the model

Characterizing spike count distributions From the posterior $q(\Pi|X)$[1], we can compute samples of the posterior of any statistic of spike counts as a function of covariates. Single neuron statistics in particular can be characterized by tuning curves for both mean firing rates and Fano factors (FF)

$$\rho(X) = \frac{1}{\Delta} \mathbb{E}_{q(\Pi|X)} \mathbb{E}_{P(Y|\Pi)}[Y] \qquad \text{FF}(X) = \mathbb{E}_{q(\Pi|X)} \left[\frac{\text{Var}_{P(Y|\Pi)}[Y]}{\mathbb{E}_{P(Y|\Pi)}[Y]} \right] \tag{6}$$

with time bin length Δ. We define a tuning index (TI) to a set of covariates $\boldsymbol{x}_*$ with respect to a count statistic $T_y(\boldsymbol{x}_*)$ that is evaluated under the count distribution marginalized over all other covariates

$$\text{TI} = \frac{\max_{\boldsymbol{x}_*} T_y(\boldsymbol{x}_*) - \min_{\boldsymbol{x}_*} T_y(\boldsymbol{x}_*)}{\max_{\boldsymbol{x}_*} T_y(\boldsymbol{x}_*) + \min_{\boldsymbol{x}_*} T_y(\boldsymbol{x}_*)} \tag{7}$$

The marginalized tuning curves can be estimated from the mean of the count distribution over the behavioural time series, see Appendix F.

Generalized Z-scores and noise correlations The deviation of activity from the predicted statistics is commonly quantified through Z-scores [8, 43, 17], which are computed as $(y - \langle y \rangle)/\sqrt{y}$ with $\langle y \rangle$ being the mean count in the time bin. If neural activity followed a Poisson distribution, the distribution of Z asymptotically tends to a unit normal when $N \gg 1$ (Appendix C). We generalize the Z-score using the probability integral transform

$$Z = \Phi^{-1}(q) \quad \text{with} \quad q(y) = \int_0^{y+\epsilon} p(\tilde{y})\, \mathrm{d}\tilde{y} = \sum_{k=0}^{y-1} P(k) + \epsilon P(y), \quad \epsilon \sim \mathcal{U}(0,1) \tag{8}$$

which removes the bias away from Gaussianity at low counts and also generalizes to arbitrary count distributions.

With the Z-score, one can completely describe single neuron statistics with respect to the model. Correlations in the neural activity however will cause Z-scores to be correlated. We define generalized lagged correlations as

$$r_{ij}(\Delta) = \langle Z_i(t)\, Z_j(t+\Delta) \rangle_t \tag{9}$$

which describes spatio-temporal correlations not captured by the model. Noise correlations [44] refer to the case of $\Delta = 0$, when r_{ij} becomes symmetric.

2.5 Assessing model fit

Our model depends on a hyperparameter $C \leq K$ that trades off flexibility with computational burden. In practice, one likely captures the neural activity accurately with C well below K and a simple basis expansion as the linear-exponential above or quadratic-exponential $\boldsymbol{\phi}(\boldsymbol{f}) = (f_1, f_1^2, e^{f_1}, \ldots, f_1 f_2, \ldots)$. This can be quantified by the statistical measures provided below, and allows us to select appropriate hyperparameters to capture the data sufficiently well.

To assess the model fit to neural spike count data, a conventional machine learning approach is to evaluate the expected log likelihood of the posterior predictive distribution on held-out data Y, leading to the cross-validated log-likelihood

$$\text{cvLL} = \mathbb{E}_{q(Z)} \mathbb{E}_{q(\Pi|X,Z)} [\log P(Y|\Pi)] \tag{10}$$

[1]For notational convenience, X denotes both observed and latent covariates here.

where we cross-validate over the neuron dimension by using the majority of neurons to infer the latent states $q(Z)$ in the held-out segment of the data, and then evaluate Equation 10 over the remaining neurons. Without latent variables, we simply take the expectation with respect to $q(\Pi|X)$. However, the cvLL doesn't reveal how well the data is described by the model in an absolute sense. Likelihood bootstrap methods are possible [28], but become cumbersome for large datasets. To assess whether the neural data is statistically distinguishable from the single neuron statistics predicted by the model, we use the Kolmogorov-Smirnov framework [45] to quantify this discrepancy

$$T_{\text{KS}} = \max_i |F(q_i) - q_i| \tag{11}$$

with empirical distribution function $F(q)$ and q from Equation 8, for details see Appendix C. This scalar number is positive and does not indicate whether the data is less or more regular than predicted by the model. A useful measure of dispersion is the variance of Z, in particular its logarithm

$$T_{\text{DS}} = \log \langle Z^2 \rangle_T + \left(\frac{1}{T} + \frac{1}{3T^2} \right) \tag{12}$$

which provides a real number indicating over- and underdispersion for positive and negative signs, respectively. T refers to the number of time steps or Z-score values. This extends the notion of over- and underdispersion beyond Poisson reference distributions [46]. Its sampling distribution under $Z \sim \mathcal{N}(0, 1)$ is asymptotically normal, centered around 0 with a variance depending on T (Appendix D). For the case of a Poisson reference model, this statistic approximates the log Fano factor at high firing rates as shown in Appendix C.

To quantify whether the model has captured noise correlations in the data, we can then compute Z-scores[2] with respect to the posterior predictive distribution

$$Q_{\theta,\varphi}(Y|X) = \int \prod_t^T \left(\prod_n^N P_n(y_{tn}|\boldsymbol{x}_t, \boldsymbol{z}_t) \right) q_\varphi(\boldsymbol{z}_t) \, \mathrm{d}\boldsymbol{z}_t \tag{13}$$

Correlations that are caused by co-modulation of neurons by low-dimensional factors can be captured with latent states Z inferred from the same data. Intuitively, this can be seen as treating latent states Z as if it was part of observed input or behaviour. Computing Equation 26 should then show a decrease in correlations r, as the Z-scores are whitened under the posterior predictive distribution. The Fisher Z-transform of r leads to a unit Gaussian Fisher Z quantity

$$Z_{Fisher} = \frac{1}{2} \log \frac{1+r}{1-r} \tag{14}$$

which is convenient for the Kolmogorov-Smirnov test or other approaches.

3 Results

In the following results, we use $C = 3$ with an elementwise linear-exponential basis expansion as described in subsection 2.1. This empirically provided sufficient model flexibility to capture both the synthetic and real data as can be seen in goodness-of-fit metrics. We use an RBF kernel with Euclidean and cosine distances for Euclidean and angular input dimensions respectively [38].

3.1 Synthetic data

Animals maintain an internal estimate of their head direction in particular circuits of their brain [4, 47, 48]. Here, we extend simple statistical models of head direction circuits [49] for validating the ability of the universal model to capture complex count statistics, as well as neural correlations through latent structure. The task is to jointly recover the ground truth count likelihoods, their tuning to covariates, and latent trajectories if relevant from activity generated using two synthetic populations. The first population was generated with a parametric heteroscedastic Conway-Maxwell-Poisson (CMP) model [50], which has decoupled mean and variance modulation as well as simultaneously over- and underdispersed activity (Fano factors above and below 1). The second population consists of Poisson neurons tuned to head direction and an additional hidden signal, which gives rise to

[2] Note the Z in the equation refers to latent states.

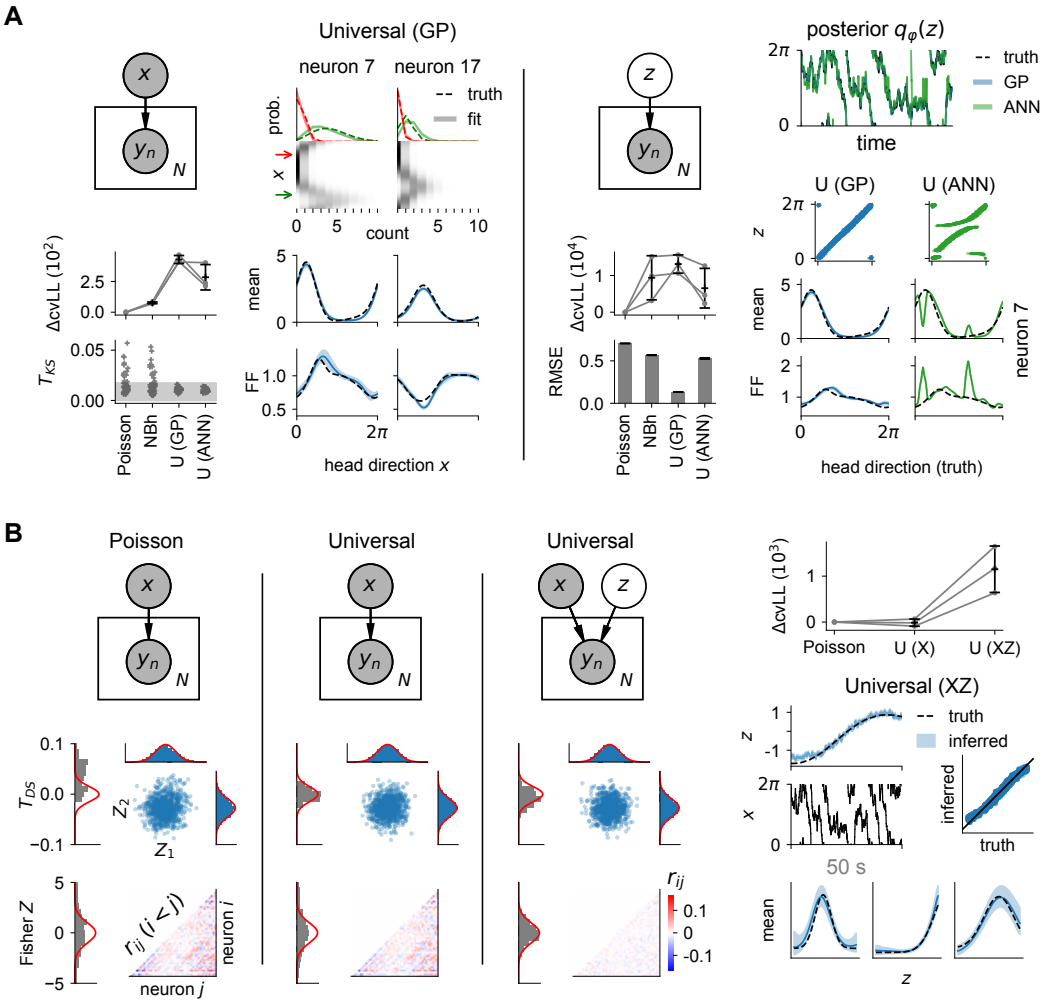

Figure 1: **Model validation with two synthetic head direction cell populations. (A)** Regression and latent variable validation experiments with synthetic data from the heteroscedastic Conway-Maxwell-Poisson population. Error bars indicate s.e.m. over cross-validation runs. The shaded region for T_{KS} indicates a 95% confidence interval. The root mean squared error (RMSE) of the inferred latent is evaluated with the geodesic distance on the ring Appendix F. **(B)** Applying regression and joint latent-regression models to the modulated Poisson population. We visualize the single neurons fits with Z-scores and T_{DS}, and noise correlations with r_{ij} and corresponding Fisher Z values (see subsection 2.5).

apparent overdispersion [28] as well as noise correlations when only regressing to observed covariates. For mathematical details of the synthetic population and count distributions, see Appendix F and Appendix B respectively.

We compare our universal model to the log Cox Gaussian process or Poisson GP model [31] and the heteroscedastic negative binomial GP (NBh) model which places GP priors on both the rate and shape parameter, a non-parametric extension of [50]. The more flexible CMPh model, analogous to NBh, has difficulty in scaling to large datasets due to the series approximation of the partition function (Appendix B). To show the power of GP based approaches, we also compare to a universal model with an artificial neural network (ANN) mapping replacing the GP. For details of the baseline models, see Appendix F. For cross-validation we split the data into 10 roughly equal non-overlapping segments, and validated on 3 chosen segments that were evenly spread out across the data. When a latent space was present, we used 90% of the neurons to infer the latent signal while validating on the

remaining neurons, and repeated this for non-overlapping subsets. We rescale the log likelihoods by the ratio of total neurons to neurons in subset and then take the average over all subsets to obtain comparable cross-validation runs to regression.

Figure 1A shows that the universal model successfully captures nontrivial count statistics of the heteroscedastic CMP population. Baseline models cannot capture cases where the Fano factor drops below 1, and indeed are outperformed. In addition, we observe that using a Bayesian GP over an ANN mapping in the model leads to a reduction in overfitting, especially in the latent setting where the ANN model fails to recover the ground truth latent signal. Figure 1B shows that the modulated Poisson population activity is seen by a Poisson regression model as overdispersed, indicated by T_{DS}. Our universal model flexibly captures the overdispersed single neuron statistics, independent from noise correlations r_{ij} that are captured when we introduce a Euclidean latent dimension. As expected, the Z-score scatter plots shows whitening under the posterior predictive distribution when the correlations are captured.

3.2 Mouse head direction cells

We apply our universal model to a recording of 33 head direction cells in the anterior nucleus of the thalamus (ANT) and the postsubiculum (PoS) of freely moving mice [37, 47]. Neural data was binned into 40 ms intervals, see details in Appendix F. Note that observed count statistics differ with bin sizes (Appendix A), which is expected as consecutive bins are not independent. Regression was performed against head direction (HD), angular head velocity (AHV), animal speed and position, and absolute time, which collectively form $X_{\mathcal{D}}$ in this model. We used 64 inducing points for regression, and added 8 for every latent dimension added (Appendix E). Cross-validation was performed similarly to the validation experiments, except that we used subsets with 85% of the neurons to infer the latent signal.

Figure 2A shows that for regression NBh performs worst, likely due to overfitting, despite containing Poisson as a special case (only approximately reachable in practice though, see Appendix B). Only the universal model captures the training data satisfactorily with respect to confidence bounds for T_{KS} and T_{DS}, although the data remained slightly underdispersed to the model with T_{DS} values slightly skewed to negative. Compared to the Poisson model, the cvLL is only slightly higher for the universal model as the data deviates from Poisson statistics in subtle ways. We see both FF above and below 1 (over- and underdispersed) across the neural firing range in Figure 2B, with quite some neurons crossing 1. Correspondingly, FF-mean correlations coefficients are often negative. Their spread away from ± 1 indicates firing rate and FF do not generally satisfy a simple relationship, especially for examples such as cell 27. Furthermore, ANT neurons seem to deviate less from Poisson statistics. From Figure 2C, we note in particular that FFs tend to decrease at the preferred head direction, but rise transiently as the head direction approaches the preferred value. We also see that tuning to speed and time primarily modulates variability rather than firing rates. All of this is impossible to pick up with baseline models, which constrain FF ≥ 1 as well as FF increasing with firing rate (Appendix B). Finally, we see more tuning of the firing rate to position in PoS cells.

When adding latent dimensions, Figure 2D shows a peak in the cvLL at two dimensions, where correspondingly the Fisher Z samples are well described by a unit normal for the first time. Kernel length scales however did not indicate redundant latent subspaces for higher dimensions as expected for automatic relevance determination, possibly due to mixing of latent dimensions. Notice the noise correlation patterns in Figure 2E tend to show positive correlations for similarly tuned neurons roughly around the diagonal of blocks, as expected from ring attractor models [22]. Intrinsic neuron variability, roughly quantified by the average FF, further decreased and thus become even more underdispersed when considering additional tuning to latents, in particular for ANT. In addition, latent signals primarily modulate firing rate as seen from TIs in Figure 2F. When looking at time scales of covariates in Figure 2G (computed as the decay time constant of the autocorrelogram (Appendix F), the latent processes seem to vary on time scales right in the gap of behavioural time scales.

Beside our main contribution of characterising the fine structure of neural variability, our results have another novel element. Using GP-based non-parametric methods, we successfully estimated the tuning of cells to as many as 8 different covariates (6 observed + 2 latent) in a statistically sound fashion (even previous GP-based approaches considered a maximum of 4 covariates). Specifically, one of our covariates was absolute experimental time to capture non-stationarities in neural tuning.

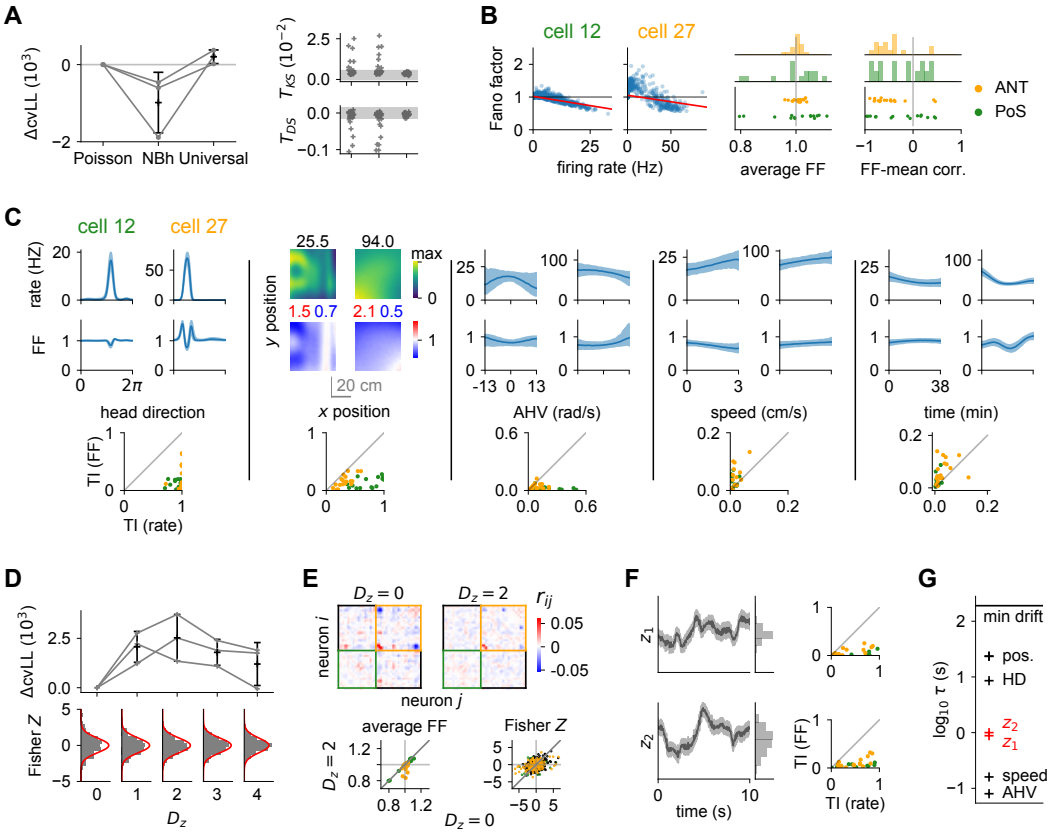

Figure 2: **Application to mouse head direction cells in the anterior nucleus of the thalamus (ANT) and the postsubiculum (PoS). (A)** Goodness-of-fit measures for Poisson, heteroscedastic negative binomial (NBh) and universal regression models. Error bars indicate s.e.m. over cross-validation runs. Shaded regions for T_{KS} and T_{DS} indicate a 95% confidence interval. **(B)** Fano factor (FF) versus the mean count of the predictive distribution across time bins (i.e. implicitly marginalising over input covariates). We also plot the average of the FFs and the Pearson r correlation between FF and mean count per cell. **(C)** Visualizing conditional tuning curves, obtained by varying the relevant covariates while keeping all others fixed (at preferred HD and centre of the arena, with zero speed, AHV and at time $t = 0$). These are generally different from marginal tuning curves used for computing TIs. **(D)** Adding latent dimensions to the universal regression model. **(E)** Comparison between the univeraal model without ($D_z = 0$) and with $Dz = 2$ latent covariates. Neurons in the noise correlation (diagonal elements of r_{ij} are always 1 and not included) plots are ordered by area first (PoS and ANT), and then by preferred head direction within area. Note that average FF is the same as in (B) but with sampled latents from $q(Z)$ as part of the input, as if they were observed. **(F)** Inferred latent timeseries for the 2D latent space with corresponding TIs. **(G)** Time scales for covariates computed from the autocorrelograms. For reference, horizontal line shows the estimated lower bound for the time scale of representational drift, computed as the minimum kernel length scale over absolute time across all neurons.

As a result, our model captured several experimental phenomena that are studied separately in the literature: drifting neural representations [51, 52, 53], anticipatory time intervals [49] and conjunctive tuning to behaviour [54]. We also applied the model in a purely latent setting similar to the example in Figure 1A, with the universal model uncovering a latent signal more closely correlated to the head direction compared to baseline models. These additional results are presented in Appendix A.

4 Discussion

Related work Neural encoding model provide a statistical description of neural count activity, and typically rely on a parametric count likelihood such as the Poisson [31], negative binomial [55, 56] or Conway-Maxwell-Poisson distribution [50]. This choice is independent of the empirical count statistics and is often mismatched to the data. Heteroscedastic count models, characterized by input-dependent noise [57], additionally regress the dispersion parameter of count distributions to covariates [58, 59]. This has shown improvements in stimulus decoding and more calibrated posterior uncertainties [50]. Copula-based models [32, 60] separate marginal distributions of single neurons from the multivariate dependency structure of the population parameterized by the copula family, and thus do not place parametric constraints on single neuron count statistics. The idea of a universal model that can capture arbitrary joint distributions has been explored for binary spike trains [33], using Bayesian non-parametric models to provide regularization and flexibility [33, 61]. However, neither approach can naturally incorporate modulation of spike count distributions by input covariates.

Our model deals with discrete spike counts ranging from 0 to K in a manner similar to categorical output variables often considered in machine learning. Similar models have been proposed mainly in the context of Gaussian process classification [62, 63], which directly pass Gaussian process function points through a softmax nonlinearity. Our approach instead passes separate Gaussian processes through a linear-softmax mapping to compute count probabilities. Introducing unobserved input variables in a Gaussian process model leads to Gaussian process latent variable models [64, 65]. Such models have recently been applied to neural data to performs dimensionality reduction [31], with extensions to non-Euclidean latent spaces and non-reversible temporal priors [38, 66].

Limitations and further work The empirical choice of hyperparameters C and basis functions ϕ is based on achieving sufficient model flexibility, as confirmed with the Kolmogorov-Smirnov approach. Recently, a multivariate extension of the Kolmogorov-Smirnov test has been proposed to directly test multivariate samples against the model [67], instead of looking at single neuron statistics. Alternatively, one could perform ARD [68, 69] by placing a Gaussian prior on W, allowing automatic selection of relevant dimensions once a basis expansion is chosen. Another avenue for future work could consider going completely non-parametric and adding a count dimension to the input space, which is evaluated at counts 0 to K for every time bin. This however increases the number of evaluation points by a factor $K + 1$. In addition, extending our model with more powerful priors for latent covariates, such as Gaussian process priors [31, 66], can improve latent variable analysis, especially at smaller time bins where the temporal prior influence becomes more important. Regularization methods may help to decorrelate inferred trajectories [70, 71].

Conclusion and impact We introduced a universal probabilistic encoding model for neural spike count data. Our model flexibly capture both single neuron count statistics and their modulation by covariates. By adding latent variables, one can additionally capture neural correlations with potentially interpretable unobserved signals underlying the neural activity. We applied our model to mouse head direction cells and found count statistics that cannot be captured with current methods. Neural activity tends to be less variable at higher firing rates, with many cells showing both over- and underdispersion. Fano factors and mean counts generally do not show a simple relation and can even be decoupled, with Fano factor modulation comparable or in some cases even exceeding that of the rate. Finally, we found that a 2D latent trajectory with a timescale of around a second explained away noise correlations in these cells.

Neural variability is usually not considered on the same footing as mean firing rates, with models assigning most computational relevance to rates [72, 73]. However, recent work on V1 has started to explore variability as playing a computationally well-defined useful role in the representation of uncertainty [24, 25, 22, 26]. The framework introduced in this paper provides a principled tool for empirically characterising neural variability and its modulations – without the biases inherent in traditional approaches, which would likely miss potentially meaningful patterns in neural activities beyond mean rates. Our model has the potential to reveal new aspects of neural coding, and may find practical applications in designing and improving algorithms for brain-machine interfaces. As progress is made in scaling and applying such technology beyond research environments [74], it becomes increasingly more important to maintain transparency, e.g. through open source code, and to raise awareness of potential ethical issues [75].

References

[1] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574, 1959.

[2] Frédéric E Theunissen, Kamal Sen, and Allison J Doupe. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *Journal of Neuroscience*, 20(6):2315–2331, 2000.

[3] John O'Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 1971.

[4] Jeffrey S Taube, Robert U Muller, and James B Ranck. Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *Journal of Neuroscience*, 10(2):420–435, 1990.

[5] Torkel Hafting, Marianne Fyhn, Sturla Molden, May-Britt Moser, and Edvard I Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005.

[6] Colin Lever, Stephen Burton, Ali Jeewajee, John O'Keefe, and Neil Burgess. Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience*, 29(31):9771–9777, 2009.

[7] Michael N Shadlen and William T Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of neuroscience*, 18(10):3870–3896, 1998.

[8] André A Fenton and Robert U Muller. Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proceedings of the National Academy of Sciences*, 95(6):3182–3187, 1998.

[9] Alexandre Pouget, Peter Dayan, and Richard Zemel. Information processing with population codes. *Nature Reviews Neuroscience*, 1(2):125–132, 2000.

[10] George J Tomko and Donald R Crapper. Neuronal variability: non-stationary responses to identical visual stimuli. *Brain research*, 79(3):405–418, 1974.

[11] A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nature reviews neuroscience*, 9(4):292–303, 2008.

[12] Johannes Nagele, Andreas VM Herz, and Martin B Stemmler. Untethered firing fields and intermittent silences: Why grid-cell discharge is so variable. *Hippocampus*, 2020.

[13] Justin Keat, Pamela Reinagel, R Clay Reid, and Markus Meister. Predicting every spike: a model for the responses of visual neurons. *Neuron*, 30(3):803–817, 2001.

[14] Gaby Maimon and John A Assad. Beyond poisson: increased spike-time regularity across primate parietal cortex. *Neuron*, 62(3):426–440, 2009.

[15] Mark M Churchland, M Yu Byron, John P Cunningham, Leo P Sugrue, Marlene R Cohen, Greg S Corrado, William T Newsome, Andrew M Clark, Paymon Hosseini, Benjamin B Scott, et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience*, 13(3):369, 2010.

[16] Adrián Ponce-Alvarez, Alexander Thiele, Thomas D Albright, Gene R Stoner, and Gustavo Deco. Stimulus-dependent variability and noise correlations in cortical mt neurons. *Proceedings of the National Academy of Sciences*, 110(32):13162–13167, 2013.

[17] Alexander S Ecker, Philipp Berens, R James Cotton, Manivannan Subramaniyan, George H Denfield, Cathryn R Cadwell, Stelios M Smirnakis, Matthias Bethge, and Andreas S Tolias. State dependence of noise correlations in macaque primary visual cortex. *Neuron*, 82(1):235–248, 2014.

[18] Neil C Rabinowitz, Robbe L Goris, Marlene Cohen, and Eero P Simoncelli. Attention stabilizes the shared gain of v4 populations. *Elife*, 4:e08998, 2015.

[19] Larry F Abbott and Peter Dayan. The effect of correlated variability on the accuracy of a population code. *Neural computation*, 11(1):91–101, 1999.

[20] Bruno B Averbeck, Peter E Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature reviews neuroscience*, 7(5):358–366, 2006.

[21] Rubén Moreno-Bote, Jeffrey Beck, Ingmar Kanitscheider, Xaq Pitkow, Peter Latham, and Alexandre Pouget. Information-limiting correlations. *Nature neuroscience*, 17(10):1410, 2014.

[22] Guillaume Hennequin, Yashar Ahmadian, Daniel B Rubin, Máté Lengyel, and Kenneth D Miller. The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron*, 98(4):846–860, 2018.

[23] Wei Ji Ma, Jeffrey M Beck, Peter E Latham, and Alexandre Pouget. Bayesian inference with probabilistic population codes. *Nature neuroscience*, 9(11):1432–1438, 2006.

[24] József Fiser, Pietro Berkes, Gergő Orbán, and Máté Lengyel. Statistically optimal perception and learning: from behavior to neural representations. *Trends in cognitive sciences*, 14(3):119–130, 2010.

[25] Gergő Orbán, Pietro Berkes, József Fiser, and Máté Lengyel. Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron*, 92(2):530–543, 2016.

[26] Rodrigo Echeveste, Laurence Aitchison, Guillaume Hennequin, and Máté Lengyel. Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. *Nature Neuroscience*, 23(9):1138–1149, 2020.

[27] David J Tolhurst, J Anthony Movshon, and Andrew F Dean. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision research*, 23(8):775–785, 1983.

[28] Robbe LT Goris, J Anthony Movshon, and Eero P Simoncelli. Partitioning neuronal variability. *Nature neuroscience*, 17(6):858, 2014.

[29] Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[30] Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen Ryu, Krishna V Shenoy, and Maneesh Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Advances in neural information processing systems*, 21:1881–1888, 2008.

[31] Anqi Wu, Nicholas A Roy, Stephen Keeley, and Jonathan W Pillow. Gaussian process based nonlinear latent structure discovery in multivariate spike train data. In *Advances in neural information processing systems*, pages 3496–3505, 2017.

[32] Pietro Berkes, Frank Wood, and Jonathan Pillow. Characterizing neural dependencies with copula models. *Advances in neural information processing systems*, 21:129–136, 2008.

[33] Il Memming Park, Evan W Archer, Kenneth Latimer, and Jonathan W Pillow. Universal models for binary spike patterns using centered dirichlet processes. In *Advances in neural information processing systems*, pages 2463–2471, 2013.

[34] Michalis Titsias. Variational learning of inducing variables in sparse gaussian processes. In *Artificial Intelligence and Statistics*, pages 567–574, 2009.

[35] James Hensman, Nicolo Fusi, and Neil D Lawrence. Gaussian processes for big data. *arXiv preprint arXiv:1309.6835*, 2013.

[36] James Hensman, Alexander Matthews, and Zoubin Ghahramani. Scalable variational gaussian process classification. 2015.

[37] Adrien Peyrache and György Buzsáki. Extracellular recordings from multi-site silicon probes in the anterior thalamus and subicular formation of freely moving mice. *CRCNS*, 2015.

[38] Kristopher Jensen, Ta-Chu Kao, Marco Tripodi, and Guillaume Hennequin. Manifold gplvms for discovering non-euclidean latent structure in neural data. *Advances in Neural Information Processing Systems*, 33, 2020.

[39] Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

[40] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations*, 2014.

[41] Luca Falorsi, Pim de Haan, Tim R Davidson, and Patrick Forré. Reparameterizing distributions on lie groups. *arXiv preprint arXiv:1903.02958*, 2019.

[42] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[43] Adam Kohn and Matthew A Smith. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *Journal of Neuroscience*, 25(14):3661–3673, 2005.

[44] I-Chun Lin, Michael Okun, Matteo Carandini, and Kenneth D Harris. The nature of shared cortical variability. *Neuron*, 87(3):644–656, 2015.

[45] Jonathan W Pillow. Time-rescaling methods for the estimation and assessment of non-poisson neural encoding models. In *Advances in neural information processing systems*, pages 1473–1481, 2009.

[46] Adam S Charles, Mijung Park, J Patrick Weller, Gregory D Horwitz, and Jonathan W Pillow. Dethroning the fano factor: a flexible, model-based approach to partitioning neural variability. *Neural computation*, 30(4):1012–1045, 2018.

[47] Adrien Peyrache, Marie M Lacroix, Peter C Petersen, and György Buzsáki. Internally organized mechanisms of the head direction sense. *Nature neuroscience*, 18(4):569–575, 2015.

[48] Rishidev Chaudhuri, Berk Gerçek, Biraj Pandey, Adrien Peyrache, and Ila Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature neuroscience*, 22(9):1512–1520, 2019.

[49] Johannes Zirkelbach, Martin Stemmler, and Andreas VM Herz. Anticipatory neural activity improves the decoding accuracy for dynamic head-direction signals. *Journal of Neuroscience*, 39(15):2847–2859, 2019.

[50] Abed Ghanbari, Christopher M Lee, Heather L Read, and Ian H Stevenson. Modeling stimulus-dependent variability improves decoding of population neural responses. *Journal of Neural Engineering*, 16(6):066018, 2019.

[51] Yaniv Ziv, Laurie D Burns, Eric D Cocker, Elizabeth O Hamel, Kunal K Ghosh, Lacey J Kitch, Abbas El Gamal, and Mark J Schnitzer. Long-term dynamics of ca1 hippocampal place codes. *Nature neuroscience*, 16(3):264, 2013.

[52] Michael E Rule, Adrianna R Loback, Dhruva V Raman, Laura N Driscoll, Christopher D Harvey, and Timothy O'Leary. Stable task information from an unstable neural population. *Elife*, 9:e51121, 2020.

[53] Daniel Deitch, Alon Rubin, and Yaniv Ziv. Representational drift in the mouse visual cortex. *bioRxiv*, 2020.

[54] Francesca Sargolini, Marianne Fyhn, Torkel Hafting, Bruce L McNaughton, Menno P Witter, May-Britt Moser, and Edvard I Moser. Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*, 312(5774):758–762, 2006.

[55] Jonathan Pillow and James Scott. Fully bayesian inference for neural models with negative-binomial spiking. *Advances in neural information processing systems*, 25:1898–1906, 2012.

[56] Yuanjun Gao, Lars Busing, Krishna V Shenoy, and John P Cunningham. High-dimensional neural spike train analysis with generalized count linear dynamical systems. In *Advances in neural information processing systems*, pages 2044–2052, 2015.

[57] Miguel Lázaro-Gredilla and Michalis K Titsias. Variational heteroscedastic gaussian process regression. In *ICML*, 2011.

[58] Seth D Guikema and Jeremy P Goffelt. A flexible count data regression model for risk analysis. *Risk Analysis: An International Journal*, 28(1):213–223, 2008.

[59] Kimberly F Sellers and Galit Shmueli. A flexible regression model for count data. *The Annals of Applied Statistics*, pages 943–961, 2010.

[60] Meng Hu, Kelsey L Clark, Xiajing Gong, Behrad Noudoost, Mingyao Li, Tirin Moore, and Hualou Liang. Copula regression analysis of simultaneously recorded frontal eye field and inferotemporal spiking activity during object-based working memory. *Journal of Neuroscience*, 35(23):8745–8757, 2015.

[61] Evan W Archer, Il Memming Park, and Jonathan W Pillow. Bayesian entropy estimation for binary spike train data using parametric prior knowledge. In *Advances in neural information processing systems*, pages 1700–1708, 2013.

[62] Kian Ming A Chai. Variational multinomial logit gaussian process. *The Journal of Machine Learning Research*, 13:1745–1808, 2012.

[63] Yarin Gal, Yutian Chen, and Zoubin Ghahramani. Latent gaussian processes for distribution estimation of multivariate categorical data. In *International Conference on Machine Learning*, pages 645–654. PMLR, 2015.

[64] Neil D Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *Nips*, volume 2, page 5. Citeseer, 2003.

[65] Michalis Titsias and Neil D Lawrence. Bayesian gaussian process latent variable model. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 844–851, 2010.

[66] Virginia Rutten, Alberto Bernacchia, Maneesh Sahani, and Guillaume Hennequin. Non-reversible gaussian processes for identifying latent dynamical structure in neural data. *Advances in Neural Information Processing Systems*, 33, 2020.

[67] Michael Naaman. On the tight constant in the multivariate dvoretzky–kiefer–wolfowitz inequality. *Statistics & Probability Letters*, 173:109088, 2021.

[68] Christopher M Bishop. Bayesian pca. *Advances in neural information processing systems*, pages 382–388, 1999.

[69] Andreas C. Damianou, Carl Henrik Ek, Michalis K. Titsias, and Neil D. Lawrence. Manifold relevance determination. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012*. icml.cc / Omnipress, 2012.

[70] Irina Higgins, Loïc Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

[71] Ricky TQ Chen, Xuechen Li, Roger Grosse, and David Duvenaud. Isolating sources of disentanglement in vaes. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 2615–2625, 2018.

[72] Guillaume Hennequin, Tim P Vogels, and Wulfram Gerstner. Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron*, 82(6):1394–1406, 2014.

[73] Klaus Wimmer, Duane Q Nykamp, Christos Constantinidis, and Albert Compte. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature neuroscience*, 17(3):431–439, 2014.

[74] Elon Musk et al. An integrated brain-machine interface platform with thousands of channels. *Journal of medical Internet research*, 21(10):e16194, 2019.

[75] Jens Clausen. Man, machine and in between. *Nature*, 457(7233):1080–1081, 2009.

[76] Riccardo Barbieri, Michael C Quirk, Loren M Frank, Matthew A Wilson, and Emery N Brown. Construction and analysis of non-poisson stimulus-response models of neural spiking activity. *Journal of neuroscience methods*, 105(1):25–37, 2001.

[77] Emery N Brown, Riccardo Barbieri, Valérie Ventura, Robert E Kass, and Loren M Frank. The time-rescaling theorem and its application to neural spike train data analysis. *Neural computation*, 14(2):325–346, 2002.

[78] William E Skaggs, Bruce L McNaughton, Matthew A Wilson, and Carol A Barnes. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, 6(2):149–172, 1996.

[79] Angus Chadwick, Mark CW van Rossum, and Matthew F Nolan. Independent theta phase coding accounts for ca1 population sequences and enables flexible remapping. *Elife*, 4:e03542, 2015.

[80] Jakob Gulddahl Rasmussen. Lecture notes: Temporal point processes and the conditional intensity function. *arXiv preprint arXiv:1806.00221*, 2018.

[81] Hongyuan Mei and Jason Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. *arXiv preprint arXiv:1612.09328*, 2016.

[82] Shuai Xiao, Junchi Yan, Xiaokang Yang, Hongyuan Zha, and Stephen Chu. Modeling the intensity function of point process via recurrent neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[83] Takahiro Omi, Naonori Ueda, and Kazuyuki Aihara. Fully neural network based model for general temporal point processes. *arXiv preprint arXiv:1905.09690*, 2019.

[84] Oleksandr Shchur, Nicholas Gao, Marin Biloš, and Stephan Günnemann. Fast and flexible temporal point processes with triangular maps. *Advances in Neural Information Processing Systems*, 33, 2020.

[85] Richard Kempter, Christian Leibold, György Buzsáki, Kamran Diba, and Robert Schmidt. Quantifying circular–linear associations: Hippocampal phase precession. *Journal of neuroscience methods*, 207(1):113–124, 2012.

[86] James T. Wilson, Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, and Marc Peter Deisenroth. Efficiently sampling functions from gaussian process posteriors. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 10292–10302. PMLR, 2020.

[87] Andrew Wilson and Hannes Nickisch. Kernel interpolation for scalable structured gaussian processes (kiss-gp). In *International Conference on Machine Learning*, pages 1775–1784. PMLR, 2015.

[88] Galit Shmueli, Thomas P Minka, Joseph B Kadane, Sharad Borle, and Peter Boatwright. A useful distribution for fitting discrete data: revival of the conway–maxwell–poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):127–142, 2005.

[89] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.

[90] Sung Soo Kim, Hervé Rouault, Shaul Druckmann, and Vivek Jayaraman. Ring attractor dynamics in the drosophila central brain. *Science*, 356(6340):849–853, 2017.

[91] Alan Benson, Nial Friel, et al. Bayesian inference, model selection and likelihood estimation using fast rejection sampling: the conway-maxwell-poisson distribution. *Bayesian Analysis*, 2021.

[92] Richard A Johnson and Thomas Wehrly. Measures and models for angular correlation and angular–linear correlation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(2):222–229, 1977.

[93] Nick I Fisher and AJ Lee. A correlation coefficient for circular data. *Biometrika*, 70(2):327–332, 1983.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See section 4.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See section 4.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] The appendices contain formulae, derivations and references for properties of statistical quantities that we defined in subsection 2.5.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] The code will be provided as a GitHub repo (zipped and URL), with instructions for running experiments in the README.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] Error bars in section 3 are taken over separate cross-validation runs.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] The relevant citations are in subsection 3.2 when presenting results for real data.
   (b) Did you mention the license of the assets? [Yes] The data was publicly available through CRCNS, link at https://crcns.org/data-sets/thalamus/th-1.
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] The data was publicly available through CRCNS.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]