ARTICLE **Communicated by Zhe Chen**

Population Vectors Can Provide Near Optimal Integration of Information

Josue Orellana
josue@cmu.edu
Center for the Neural Basis of Cognition, Pittsburgh, PA 15213, and Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

Jordan Rodu
jrodu@stat.cmu.edu
Center for the Neural Basis of Cognition, Pittsburgh, PA 15213, U.S.A., and Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

Robert E. Kass
kass@stat.cmu.edu
Center for the Neural Basis of Cognition, Pittsburgh, PA 15213, U.S.A., and Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A., and Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

Much attention has been paid to the question of how Bayesian integration of information could be implemented by a simple neural mechanism. We show that population vectors based on point-process inputs combine evidence in a form that closely resembles Bayesian inference, with each input spike carrying information about the tuning of the input neuron. We also show that population vectors can combine information relatively accurately in the presence of noisy synaptic encoding of tuning curves.

1 Introduction

The brain uses populations of spiking neurons to encode, communicate, and combine sources of information. There is considerable interest in whether these processes might be optimal as specified by Bayesian inference (Körding & Wolpert, 2004; Ma, Beck, Latham, & Pouget, 2006; Beck et al., 2008; Fischer & Peña, 2011). In particular, under certain circumstances (exponential families, including gaussian families), Bayes' theorem can combine information from two sources in a simple way: the posterior mean is a weighted sum of the two means representing the two sources, with the weights determined by the relative size of the two precision parameters

Neural Computation **29, 2021–2029** (2017)
doi:10.1162/NECO_a_00992

(for gaussians, the two reciprocals of the variances). Here we consider the combination of information by population vectors. While population vectors were first introduced in the context of motor control (Georgopoulos, Schwartz, & Kettner, 1986), our discussion concerns general cortical information processing.

Let $\theta \in [0, 2\pi)$ be a directional tuning parameter and let $\theta_n \in [0, 2\pi)$ be the preferred direction of neuron n, for $n = 1, \ldots, N$. We write the unit vector representation of θ as $u = u(\theta) = (\cos\theta, \sin\theta)$. Let y_n be the spike count (in a given time window) for neuron n, and let $\alpha = \sum_{n=1}^{N} y_n$ be the total number of spikes in the population. The unnormalized population vector is

$$P = \sum_{n=1}^{N} y_n(\cos\theta_n, \sin\theta_n). \tag{1.1}$$

Let $\bar{C} = \alpha^{-1}\Sigma_{n=1}^{N} y_n \cos\theta_n$ and $\bar{S} = \alpha^{-1}\Sigma_{n=1}^{N} y_n \sin\theta_n$. In directional data analysis, $(\bar{C}, \bar{S})$, which is analogous to the sample mean, is called the sample mean resultant vector, with length $\bar{R} = \|(\bar{C}, \bar{S})\|$, and the corresponding sample precision is expressed as $\bar{\kappa} = \alpha\bar{R}$. The estimate of direction provided by the population vector in angular form is

$$\begin{aligned}\bar{\theta} &= \arctan(\bar{S}/\bar{C}) \\ &= \arctan\left(\frac{\Sigma_{n=1}^{N} y_n \sin\theta_n}{\Sigma_{n=1}^{N} y_n \cos\theta_n}\right),\end{aligned}$$

and the corresponding unit vector estimate can be written as

$$\bar{u} = (\cos\bar{\theta}, \sin\bar{\theta}).$$

The points we make in this article are closely related to, but different from, observations made previously in the literature: first, population vectors combine information similarly to Bayes' theorem, using linear combinations with coefficients based on dispersions; second, population vectors can combine information relatively accurately in the presence of modest misspecifications in tuning curves.

2 Optimal Combination for von Mises Tuning Curves

We take the spike counts to be random variables Y_n (observed values being y_n) and assume the spike counts Y_n are Poisson distributed, independently across neurons, with means $\lambda_n = f_n(\theta)$ where

$$f_n(\theta) = A\exp\{B\cos(\theta - \theta_n)\}$$

has the form of a von Mises pdf (an exponential family on the unit circle; Fisher, 1996). Using unit vectors, we can write this as

$$f_n(\theta) = A\exp\{Bu^T u_n\},$$

where $u_n = (\cos\theta_n, \sin\theta_n)$. The population parameters are amplitude A and concentration B. For this model, Bayesian inference can be expressed in closed form. We review the posterior density of the stimulus given the population response: $p(\theta|y) \propto p(y|\theta)p(\theta)$, where $p(y|\theta)$ is the likelihood and $p(\theta)$ is the prior. For a uniform prior, if $\sum_{n=1}^{N} f_n(\theta)$ is constant in θ, the posterior distribution also has a von Mises form: $p(\theta|y) = \tilde{A}\exp\{\hat{\kappa}\cos(\theta - \hat{\theta})\}$. Where $\tilde{A}$ is a normalizing constant, we now focus on $\hat{\kappa}$ and $\hat{\theta}$. The unnormalized posterior estimate of direction is given by

$$Q = B\sum_{n=1}^{N} y_n(\cos\theta_n, \sin\theta_n).$$

Let $\hat{C} = B\sum_{n=1}^{N} y_n\cos(\theta_n)$ and $\hat{S} = B\sum_{n=1}^{N} y_n\sin(\theta_n)$. The precision of the posterior distribution may be summarized by $\hat{\kappa} = \|(\hat{C}, \hat{S})\|$. The Bayes estimate in angular form is

$$\begin{aligned}\hat{\theta} &= \arctan(\hat{S}/\hat{C})\\ &= \arctan\left(\frac{\Sigma_{n=1}^{N} y_n\sin\theta_n}{\Sigma_{n=1}^{N} y_n\cos\theta_n}\right),\end{aligned}$$

which can be written as a unit vector

$$\hat{u} = (\cos\hat{\theta}, \sin\hat{\theta}).$$

Note that $\hat{\theta} = \bar{\theta}$. In other words, in this special case of a population with von Mises tuning curves, all having the same precision coefficient B, with dense uniformly distributed preferred directions (specifically, $\sum_{n=1}^{N} f_n(\theta)$ is constant in θ), and Poisson spiking, the population vector estimate of direction is also the Bayes estimate.

We can rewrite Q as

$$Q = \hat{\kappa}\hat{u}.$$

The posterior density based on combining two populations (see Figures 1A and 1B) is $p(\theta|y_1, y_2) \propto p(\theta|y_1)p(\theta|y_2)p(\theta)$. For a uniform prior, the posterior is also a von Mises distribution with unnormalized Bayes estimate of direction,

$$Q_{combined} = \hat{\kappa}_1\hat{u}_1 + \hat{\kappa}_2\hat{u}_2,$$

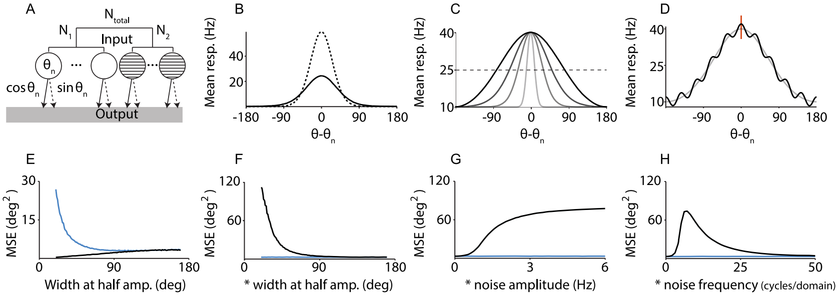

Figure 1: (A) Combining sources of information. Input: combining information within each distinct population (blank and stripes). Output: Combining information across populations to represent a tuning parameter θ. (B) Two von Mises functions corresponding to different prototypical population tuning curves (Parameters: A1 = 2, B1 = 2.5; A2 = 0.4, B2 = 5). (C) Von Mises–like tuning curves, with fixed response range between 10 Hz and 40 Hz and variable tuning width. The tuning width is measured at half amplitude. For panels in the bottom row, the *blue* trace is PV and the *black* trace is maximum likelihood (Bayes). All results are in response to a stimulus of 0 degrees. (E) Combining input sources within a population for von Mises–like tuning curves with different widths. Here, the error can be much larger for PV at the narrower widths. (F, G, H) Examining robustness under wrong tuning models. The *asterisk* indicates that a wrong model has been used for tuning curve width (F), noise amplitude (G), and noise frequency (H). In each case, PV has smaller error at most settings. The true model is shown in panel D (gray), single-peak tuning curve width of 150 degrees. The black multipeak tuning curve in panel D includes additive cosine high-frequency noise over the stimulus, frequency = 10 cycles/domain, amplitude = 2 Hz. The noise frequency was fixed at 10cycles/domain in G. The noise amplitude was fixed at 2 Hz in panel H. MSE is based on 5000 trials for all simulations.

and precision parameter $\hat{\kappa}_3 = \|Q_{combined}\|$. The unit vector Bayes estimate becomes

$$\hat{u}_3 = \hat{\kappa}_3^{-1}(\hat{\kappa}_1\hat{u}_1 + \hat{\kappa}_2\hat{u}_2). \tag{2.1}$$

Note that the posterior distribution is determined by the location parameter $\hat{u}_3$ and the precision parameter $\hat{\kappa}_3$.

3 Combining Information with Population Vectors

Writing the unnormalized population vector as $P = \alpha\bar{R}\bar{u}$, we have

$$P = \bar{\kappa}\bar{u},$$

and combining two populations, we get

$$P_{combined} = \bar{\kappa}_1 \bar{u}_1 + \bar{\kappa}_2 \bar{u}_2.$$

Let $\bar{\kappa}_3 = \|P_{combined}\|$. The unit vector estimate becomes

$$\bar{u}_3 = \bar{\kappa}_3^{-1}(\bar{\kappa}_1 \bar{u}_1 + \bar{\kappa}_2 \bar{u}_2), \tag{3.1}$$

which is very similar to equation 2.1. Specifically, the unit vector representing the combined information is a linear combination of the unit vectors representing the information from each of the two populations, with coefficients proportional to their respective precisions.

4 Efficiency and Robustness

We simulated $N = 200$ input neurons, equally spaced θ_n, under von Mises $f_n(\theta)$ with a fixed response range (10–40 Hz). We define tuning width as the width of the tuning curves at half amplitude (see Figure 1C). The mean squared error (MSE) of PV ($\bar{\theta}$) and optimal ($\hat{\theta}$) estimators are plotted as a function of tuning width in Figure 1E. In this case, PV has high statistical efficiency.

Now we examine the performance of $\bar{\theta}$ and $\hat{\theta}$ when there is a mismatch between the data-generating model for θ and the model assumed in computing the estimates. We consider both mismatches in tuning curve width (see Figure 1F) and mismatches due to high-frequency noise (see Figures 1G and 1H). For these departures from the data-generating model, PV is more robust. The simulation begins by generating trials from a true encoding model (see Figure 1D, gray single-peak tuning curve, 150 degree width). When decoding, we assume the respective wrong models (see Figures 1F to 1H) before computing the estimates of θ. For all of these cases in Figures 1F to 1H, the error is smaller for PV than for the wrong-model optimal estimate.

5 Discussion

We wrote the Bayes estimate, equation 2.1, in the von Mises case for two reasons. First, it is a weighted sum of the location parameters, with weights determined by the relative size of the two precision parameters, analogous to the gaussian case. This could be viewed as a consequence of the von Mises distribution being an exponential family on the circle. In this special case, the posterior location and precision determine the posterior distribution itself. Although in principle the posterior density can be calculated too, this requires a normalization constant, and it is plausible that if the nervous system were propagating information in terms of posterior distributions, it might use only location and precision at each step. This is part of the argument used by previous authors. Second, the Bayes estimate and the posterior precision are found from a merged population vector, meaning a population vector that merges the spike counts from the two populations, as in equation 3.1. More generally, regardless of the tuning curve, population vectors can provide Bayes-like integration of information in the sense that equation 3.1 is similar to equation 2.1, with uncertainty being represented by the magnitude of the vector (the shorter the vector, the greater the uncertainty). Exact optimality for von Mises tuning curves may be considered a special case of the more general basis function argument given by Ma et al. (2006). However, the derivation in this special case is simpler mathematically and conceptually: we consider inputs to be point processes and require only that each upstream preferred direction θ_n is encoded as a single pair of synapses in the form $(\cos\theta_n, \sin\theta_n)$, so that each spike from the upstream neuron contributes the information $(\cos\theta_n, \sin\theta_n)$ toward the sum in equation 1.1; no auxiliary network computations (such as winner-take-all, or precision found from a separate gain calculation) are needed.

While optimal estimation using richer basis representations can accommodate subtle alterations in the shapes of tuning functions, synaptic encoding of these representations would run the risk of poor performance when there is noise in the synaptic weights. It is possible that by using only a pair of synapses for transmitting information, population vectors trade some loss of statistical efficiency for a gain in robustness. Our simulations show that in certain circumstances, population vector integration of information can be less sensitive to tuning curve misspecification than Bayesian integration. This is consistent with the observation that the population vector may be considered to involve the first Fourier component of a general tuning curve (Seung & Sompolinsky, 1993), and it acts as a low-pass filter in the presence of high-frequency noise. Building on a different observation about the way population vectors can be considered Bayes-like (Shi & Griffiths, 2009), the approach outlined by Ma et al. (2006) has been adapted to account for features of the stimulus space (Ganguli & Simoncelli, 2014), and the benefits of Bayesian integration have been documented in that context. It would be interesting to consider the extent to which such representations are robust to misspecifications of tuning curve basis functions represented by synaptic weights.

Appendix A: Deriving the Likelihood of the Response

$$\begin{aligned} P(y|\theta) &= \prod_{n=1}^{N} P(y_n|\theta) \\ &= \prod_{n=1}^{N} \mathit{Poisson}(y_n; \lambda_n = f_n(\theta)) \end{aligned}$$

$$
\begin{aligned}
&= \prod_{n=1}^{N} \left(\frac{f_n(\theta)^{y_n} \exp\{-f_n(\theta)\}}{y_n!} \right) \\
&= \left(\prod_{n=1}^{N} \frac{1}{y_n!} \right) \left(\exp\left\{ -\sum_{n=1}^{N} f_n(\theta) \right\} \right) \left(\prod_{n=1}^{N} f_n(\theta)^{y_n} \right) \\
&= \left(\prod_{n=1}^{N} \frac{1}{y_n!} \right) \left(\exp\left\{ -\sum_{n=1}^{N} f_n(\theta) \right\} \right) \left(\prod_{n=1}^{N} \exp\{y_n \log(f_n(\theta))\} \right) \\
&= \left(\prod_{n=1}^{N} \frac{1}{y_n!} \right) \left(\exp\left\{ -\sum_{n=1}^{N} f_n(\theta) \right\} \right) \left(\exp\left\{ \sum_{n=1}^{N} y_n \log(f_n(\theta)) \right\} \right) \\
&= \left(\prod_{n=1}^{N} \frac{1}{y_n!} \right) \left(\exp\left\{ -\sum_{n=1}^{N} f_n(\theta) \right\} \right) \\
&\quad \times \left(\exp\left\{ \sum_{n=1}^{N} y_n \log(A \exp\{B \cos(\theta - \theta_n)\}) \right\} \right) \\
&= \left(\prod_{n=1}^{N} \frac{1}{y_n!} \right) \left(\exp\left\{ -\sum_{n=1}^{N} f_n(\theta) \right\} \right) \\
&\quad \times \left(\exp\left\{ \sum_{n=1}^{N} y_n (\log(A) + B \cos(\theta - \theta_n)) \right\} \right) \\
&= \left(\prod_{n=1}^{N} \frac{1}{y_n!} \right) \left(\exp\left\{ -\sum_{n=1}^{N} f_n(\theta) \right\} \right) \left(\exp\left\{ \sum_{n=1}^{N} y_n \log(A) \right\} \right) \\
&\quad \times \left(\exp\left\{ \sum_{n=1}^{N} y_n B \cos(\theta - \theta_n) \right\} \right) \\
&= \left(\prod_{n=1}^{N} \frac{1}{y_n!} \right) \left(\exp\left\{ -\sum_{n=1}^{N} f_n(\theta) \right\} \right) \left(A^{\sum_{n=1}^{N} y_n} \right) \\
&\quad \times \left(\exp\left\{ B \sum_{n=1}^{N} y_n \cos(\theta - \theta_n) \right\} \right) \qquad \text{(A.1)}
\end{aligned}
$$

The first three terms on the left are constant over θ. Hence, $P(y|\theta)$ is simply an unnormalized von Mises distribution governed by the last term on the right.

Let $\hat{S} = B \sum_{n=1}^{N} y_n \sin(\theta_n)$, and $\hat{C} = B \sum_{n=1}^{N} y_n \cos(\theta_n)$. Then we can write,

$$P(y|\theta) = \tilde{A}_1 \exp\{\hat{\kappa} \cos(\theta - \hat{\theta})\}, \tag{A.2}$$

where the concentration is $\hat{\kappa} = \|(\hat{C}, \hat{S})\|$, the central tendency is $\hat{\theta} = \arctan(\hat{S}/\hat{C})$, $\hat{u} = (\cos\hat{\theta}, \sin\hat{\theta})$, and $\tilde{A}_1$ is a constant obtained from the three terms on the left.

Appendix B: Deriving the Posterior for a Single Population

$p(\theta|y) \propto p(y|\theta)p(\theta)$ where $p(y|\theta)$ is the likelihood and $p(\theta)$ is the prior. For a uniform prior and following the derivation in appendix A, the posterior distribution also has a von Mises form: $p(\theta|y) = \tilde{A}_2 \exp\{\hat{\kappa} \cos(\theta - \hat{\theta})\}$, where the normalizing constant is defined as $\tilde{A}_2 = [2\pi I_0(\hat{\kappa})]^{-1}$ with I_0 being the modified Bessel function of order zero.

Appendix C: Deriving the Combined Posterior for Two Populations

The posterior density based on combining two populations is $p(\theta|y_1, y_2) \propto p(\theta|y_1)p(\theta|y_2)p(\theta)$:

$$\begin{aligned} P(\theta|y_1)P(\theta|y_2) &\propto (\exp\{\hat{\kappa}_1 \cos(\theta - \hat{\theta}_1)\})(\exp\{\hat{\kappa}_2 \cos(\theta - \hat{\theta}_2)\}) \\ &= \exp\{\hat{\kappa}_1 \cos(\theta - \hat{\theta}_1) + \hat{\kappa}_2 \cos(\theta - \hat{\theta}_2)\} \\ &= \exp\{\hat{\kappa}_3 \cos(\theta - \hat{\theta}_3)\}. \end{aligned} \tag{C.1}$$

Let $\hat{S}_3 = \sum_{n=1}^{2} \hat{\kappa}_n \sin(\hat{\theta}_n)$, and $\hat{C}_3 = \sum_{i=1}^{2} \hat{\kappa}_n \cos(\hat{\theta}_n)$. Then we can write $\hat{\kappa}_3 = \|(\hat{C}_3, \hat{S}_3)\| = \|\hat{\kappa}_1\hat{u}_1 + \hat{\kappa}_2\hat{u}_2\|$. The central tendency is $\hat{\theta}_3 = \arctan(\hat{S}_3/\hat{C}_3)$, $\hat{u}_3 = (\cos\hat{\theta}_3, \sin\hat{\theta}_3) = \hat{\kappa}_3^{-1}(\hat{\kappa}_1\hat{u}_1 + \hat{\kappa}_2\hat{u}_2)$.

For a uniform prior, the posterior is also a von Mises distribution, $p(\theta|y_1, y_2) = \tilde{A}_3 \exp\{\hat{\kappa}_3 \cos(\theta - \hat{\theta}_3)\}$, where $\tilde{A}_3$ is a function of $\hat{\kappa}_3$ and is computed in the same way as $\tilde{A}_2$.

Acknowledgments

This work was supported by grants from the National Institutes of Health: RO1 MH64537, R90DA023426, and 5R01NS050256.

References

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., . . . Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, *60*(6), 1142–1152.

Fischer, B. J., & Peña, J. L. (2011). Owl's behavior and neural representation predicted by Bayesian inference. *Nature Neuroscience*, *14*(8), 1061–1066.

Fisher, N. I. (1996). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.

Ganguli, D., & Simoncelli, E. P. (2014). Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural Computation*, *26*(10), 2103–2134.

Georgopoulos, A., Schwartz, A., & Kettner, R. (1986). Neuronal population coding of movement direction. *Science*, *233*(4771), 1416–1419.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438.

Seung, H. S., & Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences*, *90*(22), 10749–10753.

Shi, L., & Griffiths, T. L. (2009). Neural implementation of hierarchical Bayesian inference by importance sampling. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems, 22* (pp. 1669–1677). Red Hook, NY: Curran.

Received January 16, 2017; accepted April 6, 2017.