The Population Vector Could Implement Approximately Bayesian Inference

There is considerable interest in understanding how the brain might use populations of spiking neurons to encode, communicate, and combine sources of information optimally, as specified by Bayesian inference. For instance, Kording and Wolpert, in a 2004 paper, showed that performance on a sensorimotor task was consistent with optimal combination of sensory input and the statistics of the task that were learned during training, while Ma and colleagues in a 2008 paper proposed a neural modeling framework according to which Bayesian inferences could be computed. These works focused on the form in which inputs were combined to produce the posterior mean and variance. 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 have investigated performance of population vector-based inference with various tuning functions. While it is exactly Bayesian for von Mises tuning functions, it remains approximately Bayesian for many other cases. We also suggest that encoding stability within short epochs of time could lead to nearly optimal sensorimotor integration.

(*1*)

(*2*)

There is considerable interest in understanding how the brain might use populations of spiking neurons to encode and communicate probability, and to combine sources of information in an optimal, or nearly optimal way, as specified by Bayesian inference, see the review by (*3*). A population code represents information about a stimulus or behavioral feature using the simultaneous activity of a population of spiking neurons that are sensitive to that feature (*4*). Far from being deterministic, the neural response for the same action or stimulus varies from trial to trial. This suggests that the brain might encode features as probability distributions. For example, for a *center-out* reach action, the population code might represent a probability distribution with a central directional tendency μ_θ, and a measure of precision κ.

Population code Let us suppose that spikes from each neuron i, within a population of N neurons follow independent point processes $r = \{r_i\}_{i=1,\ldots,N}$ (see the inset panel of Figure 1).

$$P(r_i|\theta) = \frac{exp\{-f_i(\theta)\}f_i(\theta)^{r_i}}{r_i!} \tag{1}$$

The mean response $f_i(\theta)$ depends on $\theta - \theta_{PDi}$, where θ is the intended direction of reach, and θ_{PDi} is the preferred direction for neuron i.

Here, we define f_i as a von Mises function, which defines an exponential family on the unit circle, analogous to the normal distribution on the real line.

$$f_i(\theta) = A_i exp\{B_i \cos(\theta - \theta_{PDi})\} \tag{2}$$

Where A_i and B_i are constants representing the i_{th} neuron's amplitude and precision, respectively. High precision indicates narrow tuning for a particular preferred stimulus θ_{PDi}.

The parameters θ, θ_{PDi}, and μ_θ are directional values; for a two-dimensional workspace, they can be conveniently expressed in circular angles $[0°, 360°]$ with $0°$ being equivalent to $360°$.

Figure 1 (left) shows the response vs. reach direction for a neuron with preferred direction of

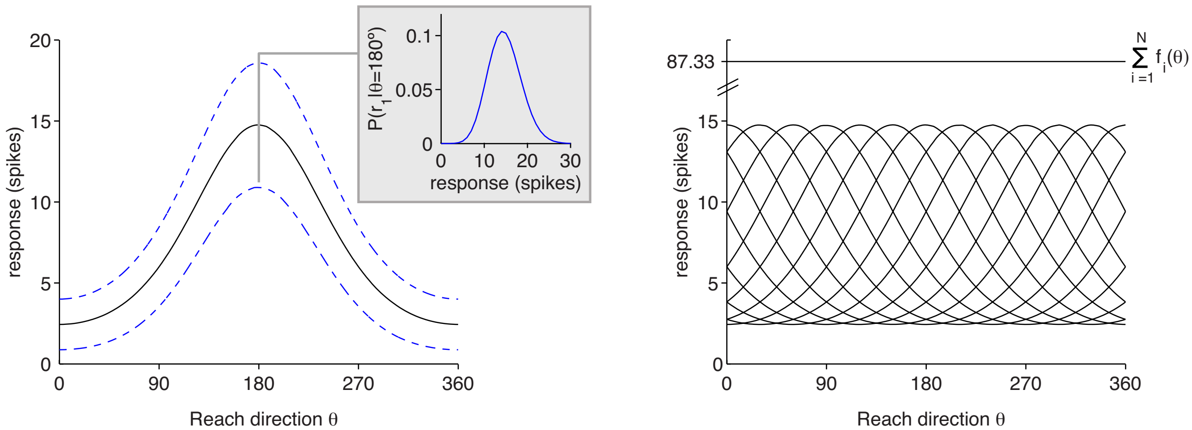

Figure 1: Encoding reach direction. The response distribution for one neuron with preferred direction of 180° is shown on the left panel (width at half amplitude= 133°). The black solid line indicates the mean, and the blue dashed lines are ± one standard deviation. The gray inset shows the poisson distribution for the neuron's response given the preferred direction stimulus. The right panel shows the mean response for a population of 12 neurons with equal precision and preferred directions spaced at 30°

$180°$. For a poisson distribution, the variance of the response is dependent on the stimulus with a Fano factor (variance to mean ratio) of 1. Tuning curves for a population of N=12 neurons with equal precision ($B_i = B$) are shown on the right panel of Figure 1, the preferred directions are spaced by 30°. Under the assumption that every neuron responds independently, the population response distribution becomes the product of the individual neuron response distributions as shown in equation 3. Experimental evidence shows that neural populations do exhibit correlations in firing rate, but for now we maintain the assumption of independence for mathematical simplicity.

$$P(r|\theta) = \prod_{i=1}^{N} P(r_i|\theta) \tag{3}$$

Now we discuss two ways of computing estimates of the stimulus from the population response: Bayesian inference and Population Vector.

Bayesian Inference Bayesian decoders use Bayes' theorem to produce a posterior probability of the stimulus given the response:

$$P(\theta|r) = \frac{P(r|\theta)P(\theta)}{P(r)} \tag{4}$$

Where $P(r|\theta)$ and $P(\theta)$ are the likelihood function and prior distribution of the stimulus respectively and $P(r)$ is a normalizing constant. Assuming a uniform prior (this assumption can be revisited later), the expression for the posterior distribution in cartesian coordinates is derived as follows:

$$P(\theta|r) \propto \prod_{i=1}^{N} P(r_i|\theta) \tag{5}$$

$$P(\theta|r) \propto \Big(\prod_{i=1}^{N} \frac{1}{r_i!}\Big)\Big(exp\{-\sum_{i=1}^{N} f_i(\theta)\}\Big)\Big(exp\{\sum_{i=1}^{N} r_i \log(A_i)\}\Big)\Big(exp\{\sum_{i=1}^{N} r_i B_i \cos(\theta - \theta_{PDi})\}\Big) \tag{6}$$

$\sum_{i=1}^{N} f_i(\theta)$ is constant over θ when the population has a uniformly dense distribution of preferred directions as shown on Figure 1B. Hence equation 6 is simply an unnormalized von Mises distribution governed by the last term on the right. Let $\bar{S}_b = N^{-1} \sum_{i=1}^{N} r_i B_i \sin(\theta_{PDi})$, and $\bar{C}_b = N^{-1} \sum_{i=1}^{N} r_i B_i \cos(\theta_{PDi})$. Then the posterior expression simply becomes:

$$P(\theta|r) = \tilde{A} exp\{\kappa \cos(\theta - \hat{\mu}_b)\}\} \tag{7}$$

Where the concentration is $\kappa^2 = \bar{S}_b^2 + \bar{C}_b^2$, the central tendency is $\hat{\mu}_b = \arctan(\bar{S}_b/\bar{C}_b)$, and the normalizing constant is defined as $\tilde{A} = [2\pi I_0(\kappa)]^{-1}$ with I_0 being the modified Bessel function of order zero. A 95% credible interval for the central tendency can be calculated directly from the von Mises probability distribution in equation 7 such that:

$$P(\hat{\mu}_b - \theta_b^* \leq \mu_b \leq \hat{\mu}_b + \theta_b^*) = 0.95 \tag{8}$$

Thus the angular size of the credible interval for the decoded stimulus is given as $L_b = 2\theta_b^*$

Population Vector Population vector is a simple way to compute an estimate of the stimulus from the population response. The estimate is an average of preferred directions weighted only by the activity of each corresponding neuron. Let $\bar{S}_{pv} = N^{-1}\sum_{i=1}^{N} r_i \sin(\theta_{PDi})$, and $\bar{C}_{pv} = N^{-1}\sum_{i=1}^{N} r_i \cos(\theta_{PDi})$. The resultant magnitude and direction are given by $\bar{R}^2 = \bar{S}_{pv}^2 + \bar{C}_{pv}^2$, and $\hat{\mu}_{pv} = \arctan(\bar{S}_{pv}/\bar{C}_{pv})$ respectively.

We can also think of the population vector as an estimate resulting from every spike carrying directional information from its emitting neuron's preferred direction. With this in mind, we consider spikes emitted by the population as samples from a circular random variable with a well defined mean direction μ_{pv}. Let $\alpha_2 = N^{-1}\sum_{j=1}^{N} r_i \cos 2(\theta_{PDi} - \hat{\mu}_{pv})$, then we can use the Central Limit Theorem to obtain an approximate 95% confidence interval for μ_{pv} as $\hat{\mu}_{pv} \pm \sin^{-1}(1.96\hat{\sigma}_{pv})$ (ref. Fisher) with $\hat{\sigma}_{pv} = \{(1-\alpha_2)/(2M\bar{R}^2)\}^{1/2}$ as the circular standard error. The angular size of the confidence interval for the decoded stimulus is given as $L_{pv} = 2\sin^{-1}(1.96\hat{\sigma}_{pv})$.

Comparing Population estimates In this section we compare the uncertainty associated with estimating the direction of reach using Bayesian Inference and Population Vector. Consider the population response to one instance in which the intended reach direction is 180°. Figure 2 (left) shows the population response plotted against the preferred direction of each neuron. The estimates of the direction of reach are the maximum likelihood $\hat{\mu}_b$ (posterior mean) for Bayesian Inference, and the activity-weighted average direction $\hat{\mu}_{pv}$ for population vector. Note that these two estimates are equal for the special case of uniform precision encoding in the population. That is if $B_i = B$ (see Figure 1) then it follows that:

$$\begin{aligned}\hat{\mu}_b &= \arctan(\bar{S}_b/\bar{C}_b) \\ &= \arctan\left(\frac{N^{-1}B\sum_{i=1}^{N} r_i \sin(\theta_{PDi})}{N^{-1}B\sum_{i=1}^{N} r_i \cos(\theta_{PDi})}\right) \\ &= \arctan(\bar{S}_{pv}/\bar{C}_{pv})\end{aligned}$$

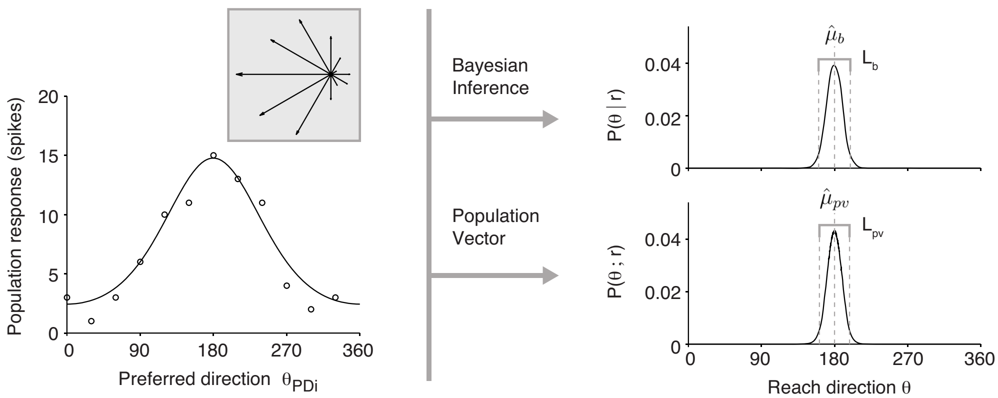

Figure 2: Computing estimates of the stimulus from the population response to an intended reach direction of 180° using Bayesian inference and Population Vector. Left: Population response plotted against the preferred direction of each neuron and shown in cartesian coordinates (inset). Right: Posterior probability distribution of the stimulus given the response using Bayesian inference (top), and probability distribution of the stimulus using Population Vector and circular Central Limit Theorem (bottom).

$$\hat{\mu}_b = \hat{\mu}_{pv} \tag{9}$$

Yet the uncertainty associated with each estimate is not necessarily equal. Figure 2 (Middle) shows the respective probability densities associated with each estimate. The credible interval of size L_b was obtained by applying Bayesian Inference under the assumption of a uniform prior distribution, and does not satisfy the coverage property. On the contrary, the confidence interval of size L_{pv} was obtained using the Central Limit Theorem, which means that it satisfies the coverage property. When repeatedly computing the ratio of credible interval size to confidence interval size we observe that the distribution is centered at 1 with a standard deviation of 0.112 (Figure 3).

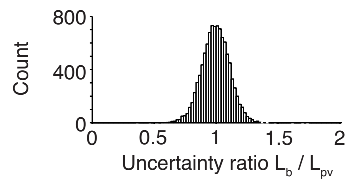

Figure 3: Distribution of the uncertainty ratio of credible to confidence interval for 10,000 repetitions for the population of neurons shown in Figure 1 and a stimulus of 180°

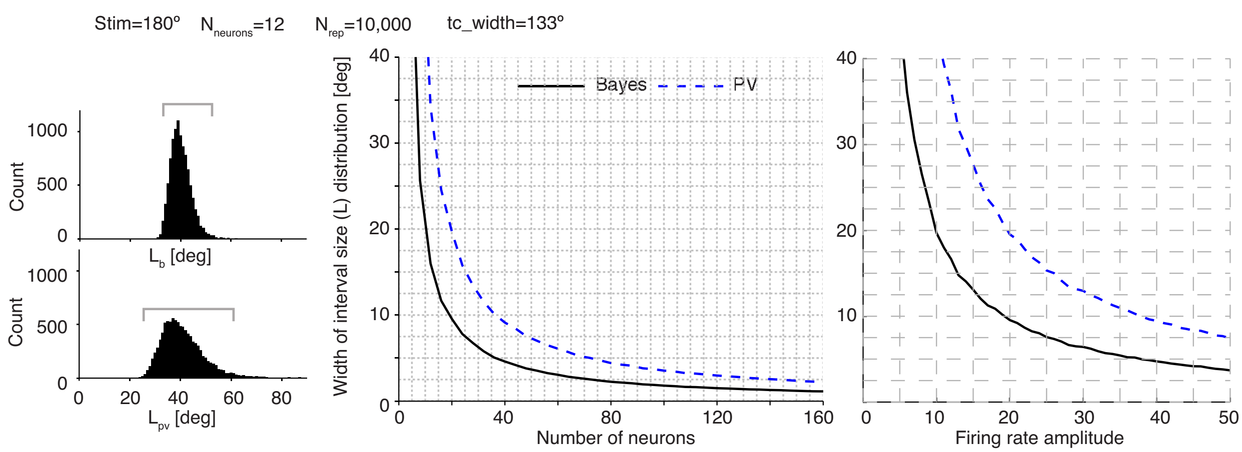

Figure 4: Comparing Performance of Bayes and Population vector based on the distribution of the credible and confidence intervals. Left: Distribution of credible (top) and confidence (bottom) intervals for 10,000 repetitions and population from Figure 1. The horizontal bars indicate the width of each distribution based on 2.5 to 97.5 percentiles. Middle: Width of interval size distribution as a function of population size (tuning curve width at half amplitude is same as Figure 1). Right: Width of interval size distribution as a function of firing rate amplitude for the population given in Figure 1

Appendix 1: Deriving the population response distribution

$$
\begin{aligned}
P(r|\theta) &= \prod_{i=1}^{N} P(r_i|\theta) \\
&= \Big(\prod_{i=1}^{N} \frac{1}{r_i!}\Big)\Big(exp\{-\sum_{i=1}^{N} f_i(\theta)\}\Big)\Big(\prod_{i=1}^{N} f_i(\theta)^{r_i}\Big) \\
&= \Big(\prod_{i=1}^{N} \frac{1}{r_i!}\Big)\Big(exp\{-\sum_{i=1}^{N} f_i(\theta)\}\Big)\Big(exp\{\sum_{i=1}^{N} r_i \log(f_i(\theta))\}\Big) \\
&= \Big(\prod_{i=1}^{N} \frac{1}{r_i!}\Big)\Big(exp\{-\sum_{i=1}^{N} f_i(\theta)\}\Big)\Big(exp\{\sum_{i=1}^{N} r_i \log(A_i) + r_i B_i \cos(\theta - \theta_{PDi})\}\Big) \\
&= \Big(\prod_{i=1}^{N} \frac{1}{r_i!}\Big)\Big(exp\{-\sum_{i=1}^{N} f_i(\theta)\}\Big)\Big(exp\{\sum_{i=1}^{N} r_i \log(A_i)\}\Big)\Big(exp\{\sum_{i=1}^{N} r_i B_i \cos(\theta - \theta_{PDi})\}\Big)
\end{aligned} \tag{10}
$$

References and Notes

1. K. P. Kording, D. M. Wolpert, *Nature* **427**, 244 (2004).

2. H. S. Seung, H. Sompolinsky, *Proceedings of the National Academy of Sciences* **90**, 10749 (1993).

3. W. J. Ma, J. M. Beck, P. E. Latham, A. Pouget, *Nat Neurosci* **9**, 1432 (2006).

4. W. Ma, A. Pouget, *Encyclopedia of neuroscience* **7**, 749 (2009).