Supplementary material for “Measuring and interpreting neuronal correlations” by Cohen and Kohn

Spike count correlation depends on measurement window in a correlated rate model

We showed in the text that correlations arising from nearly synchronous (jittered) spikes depend on the length of the measurement window (Figure 3). Measurements of r_{SC} depend on the measurement window even for very different correlation structures. For example, consider two neurons whose spiking responses are determined by independent Poisson processes conditioned on correlated underlying firing rates. The correlation between their spike count responses n_1 and n_2 computed over time T is given by

$$r_{SC}(T) = \frac{\mathrm{cov}(n_1,n_2)}{\sqrt{\mathrm{var}(n_1)\,\mathrm{var}(n_2)}} = \frac{\langle n_1 n_2\rangle - \langle n_1\rangle\langle n_2\rangle}{\sqrt{\left(\langle n_1^{\,2}\rangle - \langle n_1\rangle^2\right)\left(\langle n_2^{\,2}\rangle - \langle n_2\rangle^2\right)}} \qquad (1)$$

The covariance and variances depend on three quantities: the joint distribution over spikes $P(n_1,n_2)$, the marginal distributions $P(n_1)$ and $P(n_2)$, and the conditional spike count distributions $P(n_1|\nu_1)$ and $P(n_2|\nu_2)$ where ν_1 and ν_2 are the correlated underlying firing rates. The joint and marginal distributions are given by

$$P(n_1,n_2) = \int d\nu_1 d\nu_2 P(n_1 \mid \nu_1)P(n_2 \mid \nu_2)P(\nu_1,\nu_2) \qquad (2)$$

$$\text{and } P(n_i) = \int d\nu_i P(n_i \mid \nu_i)P(\nu_i) \qquad (3)$$

Because the conditional spike count distributions are Poisson, the conditional spike count distributions are given by

$$P(n_i \mid \nu_i) = \frac{(\nu_i T)^{n_i}\, e^{-\nu_i T}}{n_i!}. \qquad (4)$$

Equation (3) tells us that the mean spike counts are, as expected, given by

$$\langle n_i\rangle = \sum_{n_i} n_i \int d\nu_i p(n_i \mid \nu_i)p(\nu_i) = \int d\nu_i P(\nu_i)\sum_{n_i} n_i P(n_i \mid \nu_i) = \int d\nu_i P(\nu_i)\nu_i T = \langle \nu_i\rangle T. \qquad (5)$$

Given the joint distribution in Equation (2), we can compute

$$\begin{aligned}\langle n_1,n_2\rangle &= \sum_{n_1,n_2} n_1 n_2 \int d\nu_1 d\nu_2 P(n_1 \mid \nu_1)P(n_2 \mid \nu_2)P(\nu_1,\nu_2) \\ &= \int d\nu_1 d\nu_2 P(\nu_1,\nu_2) \sum_{n_1,n_2} n_1 n_2 P(n_1 \mid \nu_1)P(n_2 \mid \nu_2) \\ &= \int d\nu_1 d\nu_2 P(\nu_1,\nu_2)\sum_{n_1} n_1 P(n_1 \mid \nu_1)\sum_{n_2} n_2 P(n_2 \mid \nu_2) \\ &= \int d\nu_1 d\nu_2 p(\nu_1,\nu_2)\nu_1\nu_2 T^2 = \langle \nu_1\nu_2\rangle T^2.\end{aligned} \qquad (6)$$

Combining Equations (5) and (6) lets us compute the numerator of r_{SC}:

$$\mathrm{cov}(n_1,n_2) = \langle n_1 n_2\rangle - \langle n_1\rangle\langle n_2\rangle = \mathrm{cov}(\nu_1,\nu_2)T^2. \qquad (7)$$

To compute the variances of n_1 and n_2, we need to know $\langle n_i^2\rangle$, which is given by

$$\langle n_i^2 \rangle = \sum_{n_i} n_i^2 \int d\nu_i P(n_i \mid \nu_i) P(\nu_i) = \int d\nu_i P(\nu_i) \sum_{n_i} n_i^2 P(n_i \mid \nu_i). \tag{8}$$

For a Poisson distribution, the average of the square of the spike count is $\nu_i^2 T^2 + \nu_i T$, so

$$\langle n_i^2 \rangle = \int d\nu_i P(\nu_i)(\nu_i^2 T^2 + \nu_i T) = \langle \nu_i^2 \rangle T^2 + \langle \nu_i \rangle T. \tag{9}$$

Then

$$\mathrm{var}(n_i) = \langle n_i^2 \rangle - \langle n_i \rangle^2 = \mathrm{var}(\nu_i) T^2 + \langle \nu_i \rangle T. \tag{10}$$

Combining Equations (7) and (10) gives

$$r_{SC}(T) = \frac{\mathrm{cov}(n_1, n_2)}{\sqrt{\mathrm{var}(n_1)\,\mathrm{var}(n_2)}} = \frac{\mathrm{cov}(\nu_1, \nu_2) T^2}{\sqrt{\left(\mathrm{var}(\nu_1) T^2 + \langle \nu_1 \rangle T\right)\left(\mathrm{var}(\nu_2) T^2 + \langle \nu_2 \rangle T\right)}} \tag{11}$$

$$= r_{SC}(\infty) \frac{T}{\sqrt{\left(T + \frac{\langle \nu_1 \rangle}{\mathrm{var}(\nu_1)}\right)\left(T + \frac{\langle \nu_2 \rangle}{\mathrm{var}(\nu_2)}\right)}},$$

where $r_{SC}(\infty)$ is the correlation between the underlying firing rates ν_1 and ν_2.

Because $\frac{\langle \nu_i \rangle}{\mathrm{var}(\nu_i)} > 0$, $\frac{T}{\sqrt{\left(T + \frac{\langle \nu_1 \rangle}{\mathrm{var}(\nu_1)}\right)\left(T + \frac{\langle \nu_2 \rangle}{\mathrm{var}(\nu_2)}\right)}} < 1$.

Therefore, $r_{SC}(T) < r_{SC}(\infty)$. As T becomes large compared to $\frac{\langle \nu_i \rangle}{\mathrm{var}(\nu_i)}$,

$\frac{T}{\sqrt{\left(T + \frac{\langle \nu_1 \rangle}{\mathrm{var}(\nu_1)}\right)\left(T + \frac{\langle \nu_2 \rangle}{\mathrm{var}(\nu_2)}\right)}} \rightarrow 1$, so $r_{SC}(T) \rightarrow r_{SC}(\infty)$.

See Appendix 2 of Bair et al. (2001) for a similar derivation.

Overly stringent spike sorting criteria reduces measured r_{SC}

We showed through simulations that overly stringent spike sorting criteria reduces measurements of correlations (main text, Figure 5) and provided an analytical description of this phenomenon. Here we derive this relationship.

Consider two neurons with spike count responses n_1 and n_2, whose spikes are discarded with probability p_1 and p_2. The spikes kept from each neuron during spike sorting are given by m_1 and m_2, where $m_i = n_i(1 - p_i)$.

The random discarding of spikes is independent for the two neurons, so $\mathrm{cov}(m_1, m_2)$ is simply a function of $\mathrm{cov}(n_1, n_2)$. Specifically,

$$\mathrm{cov}(m_1,m_2) = (1-p_1)(1-p_2)\,\mathrm{cov}(n_1,n_2). \qquad (12)$$

As above, $<m_i>$ depend on the conditional probabilities $P(m_i|n_i)$. Because m_i has a binominal distribution over n_i spike counts with probability of success (1- p_i),

$$P(m_i \mid p_i) = \sum_{n_i} P(m_i \mid n_i, p_i) P(n_i \mid p_i). \qquad (13)$$

Then

$$\begin{aligned}\langle m_i \rangle &= \sum_{m_i} m_i P(m_i \mid p_i) \qquad (14)\\ &= \sum_{n_i} P(n_i \mid p_i) \sum_{m_i} m_i P(m_i \mid n_i, p_i).\end{aligned}$$

The inner sum is equal to the average value of m_i for a particular value of n_i and p_i, which is equal to $(1-p_i)n_i$, so

$$\langle m_i \rangle = \sum_{n_i} P(n_i \mid p_i) n_i (1-p_i) = (1-p_i)\langle n_i \rangle. \qquad (15)$$

We can also use Equation (13) to compute $<m_i^2>$:

$$\begin{aligned}\langle m_i^2 \rangle &= \sum_{m_i} m_i^2 P(m_i \mid p_i) \qquad (16)\\ &= \sum_{n_i} P(n_i \mid p_i) \sum_{m_i} m_i^2 P(m_i \mid n_i, p_i).\end{aligned}$$

As above, the inner sum equals the average value of m_i^2 for a particular value of n_i and p_i, so we can use the definition of variance to obtain

$$\langle m_i^2 \rangle = \sum_{n_i} P(n_i \mid p_i) \cdot \left[\mathrm{var}(m_i) + \langle m_i \rangle^2 \right] \qquad (17)$$

Using the mean of the binomial distribution over n_i spike counts with probability of success (1- p_i) and the formula for $<m_i>$ from Equation (15), we can calculate

$$\begin{aligned}\langle m_i^2 \rangle &= \sum_{n_i} P(n_i \mid p_i) \left[n_i p_i (1-p_i) + n_i^2 (1-p_i)^2 \right] \qquad (18)\\ &= \langle n_i \rangle p_i (1-p_i) + \langle n_i^2 \rangle (1-p_i)^2\\ &= \langle n_i \rangle p_i (1-p_i) + \mathrm{var}(n_i)(1-p_i)^2 + \langle n_i \rangle^2 (1-p_i)^2\end{aligned}$$

Equations (15) and (18) let us calculate var(m_i):

$$\begin{aligned}\mathrm{var}(m_i) &= \langle m_i^2 \rangle - \langle m_i \rangle^2 \qquad (19)\\ &= \langle n_i \rangle p_i (1-p_i) + \mathrm{var}(n_i)(1-p_i)^2 + \langle n_i \rangle^2 (1-p_i)^2 - \langle n_i \rangle^2 (1-p_i)^2\\ &= \langle n_i \rangle p_i (1-p_i) + \mathrm{var}(n_i)(1-p_i)^2\end{aligned}$$

We can then use (12) and (19) to compute r_{SC}:

$$r_{SC-oversort} = \frac{\mathrm{cov}(m_1, m_2)}{\sqrt{\mathrm{var}(m_1)\,\mathrm{var}(m_2)}} \tag{20}$$

$$= \frac{\dfrac{\mathrm{cov}(n_1, n_2)}{\sqrt{\mathrm{var}(n_1)\,\mathrm{var}(n_2)}}}{\sqrt{\left(1+\dfrac{p_1\langle n_1\rangle}{(1-p_1)\,\mathrm{var}(n_1)}\right)\left(1+\dfrac{p_2\langle n_1\rangle}{(1-p_2)\,\mathrm{var}(n_2)}\right)}}$$

$$= \frac{r_{SC-original}}{\sqrt{\left(1+\dfrac{p_1\langle n_1\rangle}{(1-p_1)\,\mathrm{var}(n_1)}\right)\left(1+\dfrac{p_2\langle n_1\rangle}{(1-p_2)\,\mathrm{var}(n_2)}\right)}}$$

By definition, $0 \le p_i \le 1$, $\langle n_i \rangle \ge 0$, and $\mathrm{var}(n_i) \ge 0$. Therefore, the denominator in Equation (20) is always greater than 1, so, $r_{SC\text{-}oversort}$ will always underestimate $r_{SC\text{-}original}$.