Information differences and ΔI

Suppose you want to compare two neural codes – say spike timing and spike count. The natural thing to do is compute the information using one code, compute the information using the other, and then take the difference. What we show here is that when one of the neural codes is a sub-code of the other (as spike count is a sub-code of spike timing), then this difference is exactly equal to the ΔI, the cost function used to assess the role of correlations in Nirenberg et al. (2001) [*Nature* **411**:698-701].

As usual in information calculations in the brain, you show a set of stimuli and measure neuronal responses. The latter are denoted $\mathbf{r} \equiv (r_1, r_2, ...)$ where the r_i can be any aspect of the code – 1s and 0s in small bins to indicate the presence or absence of a spike, for example. A sub-code of $\mathbf{r}$ is any function of $\mathbf{r}$: if $\mathbf{z} = \mathbf{f}(\mathbf{r})$ then $\mathbf{z}$ is a sub-code of $\mathbf{r}$. If you observe just $\mathbf{z}$, you get no more information than if you observe $\mathbf{r}$, and you usually get less. The difference is

$$\Delta \hat{I} = I(\mathbf{r}; s) - I(\mathbf{z}; s)\,,$$

where s is the stimulus and

$$\begin{aligned} I(\mathbf{r}; s) &= -\sum_{\mathbf{r}} P(\mathbf{r}) \log P(\mathbf{r}) + \sum_s P(s) \sum_{\mathbf{r}} P(\mathbf{r}|s) \log P(\mathbf{r}|s) \\ I(\mathbf{z}; s) &= -\sum_{\mathbf{z}} P(\mathbf{z}) \log P(\mathbf{z}) + \sum_s P(s) \sum_{\mathbf{z}} P(\mathbf{z}|s) \log P(\mathbf{z}|s)\,. \end{aligned}$$

The probability distributions $P(\mathbf{z}|s)$ and $P(\mathbf{z})$ are given by the usual formulae

$$P(\mathbf{z}) = \sum_{\mathbf{r}} P(\mathbf{r}) \delta(\mathbf{z} - \mathbf{f}(\mathbf{r})) \tag{1a}$$

$$P(\mathbf{z}|s) = \sum_{\mathbf{r}} P(\mathbf{r}|s) \delta(\mathbf{z} - \mathbf{f}(\mathbf{r}))\,. \tag{1b}$$

Here δ is a Kronecker δ-like object: $\delta(\mathbf{z} - \mathbf{f}(\mathbf{r})) = 1$ if $\mathbf{z} = \mathbf{f}(\mathbf{r})$ and 0 otherwise. Had these been continuous distributions, we would have used Dirac δ-functions and the sums would have been integrals.

Let's compare $\Delta\hat{I}$ to ΔI, the latter being the number of extra yes-no questions it would take to guess the stimulus given that you observed only $\mathbf{z} = \mathbf{f}(\mathbf{r})$ rather than the full set of responses, $\mathbf{r}$. This quantity is given by [Nirenberg et al., *Nature* **411**:698-701 (2001)]

$$\Delta I = \sum_{\mathbf{r}} P(\mathbf{r}) \sum_{s} P(s|\mathbf{r}) \log \left[\frac{P(s|\mathbf{r})}{P(s|\mathbf{f}(\mathbf{r}))} \right] .$$

Using Bayes' theorem and rearranging terms slightly leads to

$$\Delta I = \sum_{s} P(s) \sum_{\mathbf{r}} P(\mathbf{r}|s) \log \left[\frac{P(\mathbf{r}|s)P(s)}{P(\mathbf{r})} \; \frac{P(\mathbf{f}(\mathbf{r}))}{P(\mathbf{f}(\mathbf{r})|s)P(s)} \right] .$$

Canceling the $P(s)$ that appears in the numerator and denominator inside the logs, and again rearranging terms, we find that

$$\Delta I = I(\mathbf{r}; s) - \left[- \sum_{\mathbf{r}} P(\mathbf{r}) \log P(\mathbf{f}(\mathbf{r})) + \sum_{s} P(s) \sum_{\mathbf{r}} P(\mathbf{r}|s) \log P(\mathbf{f}(\mathbf{r})|s) \right] . \tag{2}$$

We can rewrite the first term in brackets as

$$\sum_{\mathbf{r}} P(\mathbf{r}) \log P(\mathbf{f}(\mathbf{r})) = \sum_{\mathbf{r}} P(\mathbf{r}) \sum_{\mathbf{z}} \delta(\mathbf{z} - \mathbf{f}(\mathbf{r})) \log P(\mathbf{z}) .$$

Rearranging terms one last time, we have

$$\sum_{\mathbf{r}} P(\mathbf{r}) \log P(\mathbf{f}(\mathbf{r})) = \sum_{\mathbf{z}} \log P(\mathbf{z}) \sum_{\mathbf{r}} P(\mathbf{r}) \delta(\mathbf{z} - \mathbf{f}(\mathbf{r})) = \sum_{\mathbf{z}} P(\mathbf{z}) \log P(\mathbf{z}) \tag{3}$$

where the last equality follows from Eq. (1a). Using identical logic,

$$\sum_{\mathbf{r}} P(\mathbf{r}|s) \log P(\mathbf{f}(\mathbf{r})|s) = \sum_{\mathbf{z}} P(\mathbf{z}|s) \log P(\mathbf{z}|s) . \tag{4}$$

Finally, Inserting Eqs. (3) and (4) into Eq. (2), we find that

$$\Delta I = I(\mathbf{r}; s) - \left[- \sum_{\mathbf{z}} P(\mathbf{z}) \log P(\mathbf{z}) + \sum_{\mathbf{z}} P(\mathbf{z}|s) \log P(\mathbf{z}|s) \right] = I(\mathbf{r}; s) - I(\mathbf{z}; s) = \Delta\hat{I} .$$

Thus, $\Delta\hat{I}$ and ΔI are one and the same.