How do the eyes & brain represent the outside world? (Information theory applied to neuroscience)

[Image of visual cortex and retina diagram]

[Image of James Peak, near Denver & Boulder]

Joel Zylberberg
www.jzlab.org
The neural code is not one-to-one

“Noise” (trial-to-trial variability): many activity patterns per stimulus
What aspects of neural responses carry info to brain (& what mechanisms are responsible?)

Key: quantify information coding
Information Theory in Neuroscience

- Geometry of neural coding
- Mutual Information: \( \log(\# \text{ identifiable stimuli}) \)
- Fisher Information: precision of optimal stim. estimator
Information Theory in Neuroscience

- Geometry of neural coding
- Mutual Information: $\log(\# \text{ identifiable stimuli})$
- Fisher Information: precision of optimal stim. estimator
Geometry of highly informative neural codes

Neural Response Space

{r|s_5}  {r|s_4}  {r|s_3}  {r|s_2}  {r|s_1}
Geometry of less informative neural codes

Neural Response Space

\{r|s_5\} \{r|s_4\} \{r|s_3\} \{r|s_2\} \{r|s_1\}
# identifiable stimuli:
\[ V(\text{full set of responses})/V(\text{responses to one stimulus}) \]
Information Theory in Neuroscience

- Geometry of neural coding
- Mutual Information: $\log(\# \text{ identifiable stimuli})$
- Fisher Information: precision of optimal stim. estimator
Conditionally Gaussian Response Distributions (to start):

\[ \Sigma(s) = \text{cov}(\vec{x} | s) \]

1 std. dev. probability contours: ellipsoids. Axis lengths are eigenvalues\(^{1/2}\) of \( \Sigma(s) \)

\[ V \sim \prod_{i=1}^{k} \sigma_i = \prod_{i=1}^{k} \sqrt{\lambda_i} = \sqrt{\text{det}[\Sigma(s)]} \]
$N_{\text{identifiable}} = \sqrt{\frac{\det[\Sigma]}{\det[\Sigma(s)]}}$
Mutual Information: $\log(\text{# identifiable stimuli})$

\[ N_{\text{identifiable}} = \sqrt{\frac{\det[\Sigma]}{\det[\Sigma(s)]}} \]

\[ I(r; s) = \frac{1}{2} \log \left[ \frac{\det[\Sigma]}{\det[\Sigma(s)]} \right] \]

for Gaussian distributed responses

[Pictures like this are more useful than equations]
Mutual Information: arbitrary distributions

\[ I(r; s) = \frac{1}{2} \log \left[ \frac{\det[\Sigma]}{\det[\Sigma(s)]]} \right] \]

for Gaussian distributed responses

\[ = \log[V] - \log[V(s)] \]

For arbitrary distributions, measure “volume” differently (not just \( \det[\Sigma] \)): entropy of the distribution.

\[ H(r) = - \sum_i p(r_i) \log[p(r_i)] \]

for equi-probable responses, \( H = \log(\# \text{ possible responses}) \)

\[ h(r) = - \int p(r) \log[p(r)] \, dr \]
Mutual Information: arbitrary distributions

\[ I(r; s) = H(r) - \langle H(r|s) \rangle_{p(s)} \]

\( H(r) \): “volume” of the distribution of neural responses

\( H(r|s) \): “volume” of the neural responses to each stim

\( I(r; s) \): log of the volume ratio (log of # of stim.-conditioned distributions that fit into the marginal distribution)
Entropy is non-negative

\[ H(r) = - \sum_i p(r_i) \log[p(r_i)] \]

\[ h(r) = - \int p(r) \log[p(r)] dr \]

Mutual information is bounded by \( H(r) \)

\[ I(r; s) = H(r) - \left< H(r | s) \right>_{p(s)} \]

\[ \log(\text{# of single points in } \{r\}) \]
Many equivalent formulae for Mutual Information

\[
I(r; s) = H(r) - \langle H(r|s) \rangle_{p(s)} \\
= H(s) - \langle H(s|r) \rangle_{p(r)}
\]
Entropy and Information in Neural Spike Trains

S.P. Strong,¹ Roland Koberle,¹,² Rob R. de Ruyter van Steveninck,¹ and William Bialek¹

¹NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540
²Department of Physics, Princeton University, Princeton, New Jersey 08544
(Received 29 February 1996; revised manuscript received 11 July 1997)

The nervous system represents time dependent signals in sequences of discrete, identical action potentials or spikes; information is carried only in the spike arrival times. We show how to quantify this information, in bits, free from any assumptions about which features of the spike train or input signal
motion stim

Spikes trains of H1 neuron

r is a binary “word”: spike (1) or no spike (0) in each $\Delta t$ sized bin
Question: why take “words” & not just take single bins to compute I(r;s)?

\[ I(s; \vec{r}) = H(\vec{r}) - H(\vec{r} \mid s) \]

\[ H(\vec{r}) = - \sum_i p(\vec{r}_i) \log[p(\vec{r}_i)] \]
Inferred entropy goes down with longer counting windows (autocorrelation): extrapolate to infinite window to estimate true entropy rate

Finite data effects (worse for long T) lead to underestimates of $H$

\[ H(\tilde{r}) = \sum_{i} p(\tilde{r}_i) \log [p(\tilde{r}_i)] \]
(So far) infer $H(r)$ from data

Discuss: how does $H(r)$ relate to $I(s;r)$?

$I(s; \bar{r}) = H(\bar{r}) - H(\bar{r}|s)$

What would happen if responses were perfectly deterministic?
How close does $I(s;r)$ come to $H(r)$ (how much does noise affect info encoded by $H1$?)

$$I(s; r) = H(r) - H(r|s)$$

**Measure $H(r|s)$ ("noise entropy")**

$p(r|s_1)$  $p(r|s_3)$

take distribution of responses at different epochs (corresponds to different stimuli)
H1 transmits ~80 (=160-80) bits/s of info.

This is 50% of the theoretical maximum (i.e., noise is responsible for 50% reduction vs. a “perfect” detector).

\[ I(s; \tilde{r}) = H(\tilde{r}) - H(\tilde{r} \mid s) \]
• Mutual Information: $\log(\# \text{ identifiable stimuli})$
• Direct application to neural data doable but hard

[Pictures like this are more useful than equations]
Information Theory in Neuroscience

- Geometry of neural coding
- Mutual Information: $\log(\# \text{ identifiable stimuli})$
- Fisher Information: precision of optimal stim. estimator
Why can’t we use Shannon’s mutual information for larger neural populations?

\[ I(s; \bar{r}) = H(\bar{r}) - H(\bar{r} | s) \]
\[ H(\bar{r}) = - \sum_i p(\bar{r}_i) \log[p(\bar{r}_i)] \]

\[ \rightarrow \] estimation theory (Fisher information)

[how precisely can brain estimate \( s \) after seeing \( r \)?]
Precision = 1/MSE

\[ l^{-1/2}(s) \] gives minimum noticeable change in the stim. —> discrimination tasks from psychology
Fisher info is (formally) the curvature of the log-likelihood function

\[ I(s) = - \left\langle \frac{\partial^2}{\partial s^2} \log[p(\mathbf{r}|s)] \right\rangle_{p(\mathbf{r}|s)} \]

Still hard to calculate (i.e., need \( p(r|s) \))

Conditionally Gaussian \( p(r|s) \), with the conditional covariance \( \Sigma(r|s) \) not changing much with \( s \):

\[ I(s) = \frac{\partial \hat{f}(s)^T}{\partial s} [\Sigma(\mathbf{r}|s)]^{-1} \frac{\partial \hat{f}(s)}{\partial s} \]

\( \hat{f}(s) \): mean resp. of cells to stim. \( s \)

\( \Sigma(r|s) \): conditional covariance matrix
Start with 1 cell: Fisher info is $\sim$SNR

\[ I(s) = \frac{\left( \frac{\partial f(s)}{\partial s} \right)^2}{\sigma(r|s)^2} \]

\[ \text{how much response changes (on avg.) as stim changes (signal)} \]
\[ \text{variability in responses to a given stim (noise)} \]

In which region is $I(s)$ largest?
Multiple cells: same idea, but higher-dimensional

[consider principal components of $\Sigma(s)$]

$$I(s) = \| \bar{f}'(s) \|^2 \sum_i \frac{\cos \theta_i^2}{\lambda_i}$$

$$I(s) = \frac{\partial \bar{f}(s)^T}{\partial s} \left[ \Sigma(r|s) \right]^{-1} \frac{\partial \bar{f}(s)}{\partial s}$$

$f(s)$: mean resp. of cells to stim. $s$

$\Sigma(r|s)$: conditional covariance matrix
Basic Message: signal orthogonal to noise is benign (parallel is bad) for info coding

[Averbeck et al., 2006; Abbott & Dayan, 1999; Shamir, 2014; other work; Hu, Zylberberg, and Shea-Brown, PLoS Computational Biology 2014]
Summary

• Quantify efficacy of neural code
• Mutual Information: log(# identifiable stimuli)
• Fisher Information: precision of estimator of s given r
• Both concern the same geometry: (how much) do p(r|s) overlap for different s?
Information theory and neural coding

Alexander Borst\(^1\) and Frédéric E. Theunissen\(^2\)

\(^1\)ESPM-Division of Insect Biology and \(^2\)Dept. of Psychology, University of California, Berkeley, California 94720, USA
Correspondence should be addressed to A.B. (borst@nature.berkeley.edu)