

# Examining methods for estimating mutual information in spiking neural systems

Christopher J. Rozell<sup>\*</sup> and Don H. Johnson

*Department of Electrical & Computer Engineering  
Rice University, Houston, TX, 77251*

---

## Abstract

Mutual information enjoys wide use in the computational neuroscience community for analyzing spiking neural systems. Its direct calculation is difficult because estimating the joint stimulus-response distribution requires a prohibitive amount of data. Consequently, several techniques have appeared for bounding mutual information that rely on less data. We examine two upper bound techniques and find that they are either unreliable or introduce strong assumptions about the neural code. We also examine two lower bounds, showing that they can be very loose and possibly bear little relation to the mutual information's actual value.

*Key words:* mutual information, neural coding, information theory

---

## 1 Introduction

Mutual information is in many ways the cornerstone of classic information theory, playing central roles in the analysis of both digital and analog communications systems [1]. One of the primary objectives in studying neural communication is characterizing the amount of stimulus information the output spike train represents. Though it is not clear that mutual information alone answers this question [2], all agree that it is an important measure of the statistical relationship between the stimulus and the response.

Throughout this report we consider the paradigm where a continuous-valued stimulus  $S$  is encoded in a single neuron spike response  $R$  measured over the time window  $[0, T]$ . The stimulus and its probability distribution  $p(S)$  will be

---

<sup>\*</sup> Corresponding author.

*Email address:* `crozell@rice.edu` (Christopher J. Rozell).

considered in an abstract sense, where  $S$  could represent a time domain signal or a stimulus parameter. The spike response  $R$  consists of both the number of spikes in the interval  $[0, T]$  and the actual spike times. The probability distribution  $p(R)$  is therefore a mixed distribution over one discrete parameter (the count) and a variable number of continuous parameters (the spike times).

Formally, the mutual information [1, 3] between  $S$  and  $R$  is given by

$$I(S; R) = \int p(s) \int p(r|s) \log \frac{p(r|s)}{p(r)} dr ds. \quad (1)$$

It is well known that the mutual information can be written as a difference of differential entropies [3],  $I(S; R) = h(R) - h(R|S)$ , where

$$\begin{aligned} h(R) &= - \int_{-\infty}^{\infty} p(r) \log p(r) dr, \\ h(R|S) &= - \int_{-\infty}^{\infty} p(r) \int_{-\infty}^{\infty} p(r|s) \log p(r|s) ds dr. \end{aligned}$$

Mutual information can be computed for discrete random variables by changing the integrals to summations, with the differential entropies becoming discrete entropies, notated  $H(R)$  and  $H(R|S)$ . The input probability distribution  $p(S)$  is entirely under experimental control, but the input-output relationship of the system,  $p(R|S)$  needs to be measured. If the spike response contains dependencies (serial or inter-neuron), then a prohibitive number of stimulus presentations are usually required to estimate  $p(R|S)$  completely.

Neuroscientists use mutual information because they want to characterize how well a spike response conveys the information contained in external stimuli. Information theory gives meaning to mutual information in two distinct ways. In digital communications, mutual information is used to define the maximum rate that a bit sequence can be reliably communicated through a noisy channel [1]. In analog communication, mutual information is used in the context of rate-distortion theory [1, 4] to determine whether a continuous signal can be transmitted and received with an acceptable distortion level. Whichever approach is used requires mutual information to be optimized over the stimulus probability distribution  $p(S)$ , and a single measurement for some choice of  $p(S)$  may not be theoretically interesting otherwise [2]. Consequently,  $p(S)$  is often chosen to be either uniform or Gaussian in the claim of maximizing the mutual information. However, these distributions maximize mutual information in only very special cases, and the true maximizing distributions can be very abnormal in more realistic scenarios [2].

In practice, spike responses are almost always binned with binwidth  $\Delta$  and  $I(S; R)$  is estimated from this discrete approximation. The mutual information found in this way will converge to  $I(R; S)$  as  $\Delta \rightarrow 0$ , but smaller values of  $\Delta$  require more data to make reliable estimates. The “direct method” for

estimating mutual information [5] estimates  $H(R)$  by presenting stimuli drawn from  $p(S)$ , and estimates  $H(R|S)$  from repeated presentations of the same stimulus. Because the amount of data required to estimate  $p(R|S)$  is usually very large, most researchers employ techniques for bounding  $I(S; R)$ . These techniques are especially important as we consider the need to optimize  $I(S; R)$  over  $p(S)$  instead of just calculating it once. If mutual information is to be used in analyzing neural data, we must determine whether our present techniques for bounding it are adequate. We closely investigate several techniques found in the computational neuroscience literature to provide guidelines on their use. Specifically we examine cases where spiking neurons violate the simplifying assumptions to see if the bounds are always valid and useful.

## 2 Upper bound methods

There are two primary methods in the literature for calculating an upper bound on  $I(S; R)$ . For discrete (i.e., binned) responses, the simplest bound is the observation that since  $H(R|S) \geq 0$ , then we know from (1) that  $I(S; R) \leq H(R)$  [6, 7]. However,  $H(R)$  depends strongly on the choice of  $\Delta$  [8], and does *not* converge as  $\Delta \rightarrow 0$ . This is in stark contrast to mutual information, which does converge as the approximation interval gets smaller. The choice of binwidth determines this upper bound, introduces significant assumptions about the timing precision of the neural code, and does not relate objectively to the true mutual information.

The second upper bound method [5] assumes that if the stimulus presented is Gaussian, then the response is also Gaussian and the result of an additive Gaussian noise process. A Gaussian model makes little sense for a spike train that consists of discrete point events. However, because the intent is only to find an upper bound, the reality of the model does not matter if the bound is accurate and effective. With this model, the mutual information rate [1] is

$$I(S; R) = \int_0^W \log(1 + \text{SNR}(f)) df, \quad (2)$$

where  $\text{SNR}(f) = P_{SS}(f)/P_{NN}(f)$ , the ratio of the power spectral densities of the signal and the noise. In this technique, the stimulus is repeated many times and the signal estimate is formed by averaging the responses at each discrete time step (i.e., bin),  $\bar{S}_i = E[R_i]$ . The noise for each trial is estimated from the response by subtracting the mean signal,  $N_i = R_i - \bar{S}_i$ . The power spectrum of  $\bar{S}$  and  $N$  are calculated and equation (2) is an upper bound. Borst and Theunissen [5] mention that one must check the correctness of the Gaussian distribution assumption, but also say explicitly that this technique can be used on both “spiking and nonspiking neuronal responses” [5, p.953]. It is worth

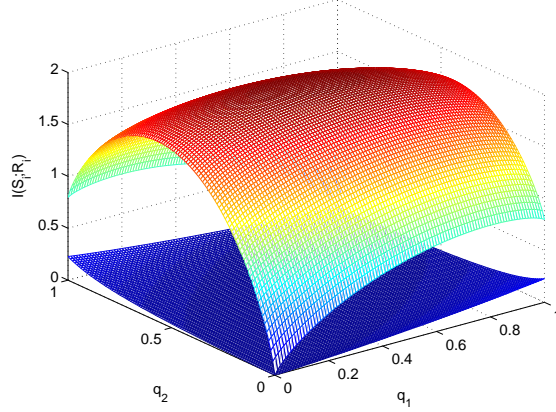
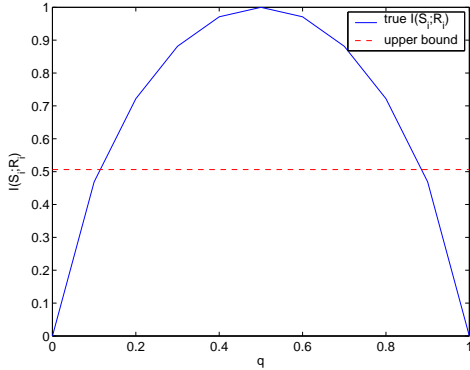


Fig. 1. A violation of the upper bound. Fig. 2. The lower bound can be trivial.

noting that this technique also assumes that the noise is independent of the stimulus. In a Poisson model of neural responses, the mean firing rate and the variance are the same (and determined by the stimulus), so the independent noise assumption is immediately violated.

When the noise source is additive and Gaussian, a Gaussian input will indeed maximize the mutual information and this would be a valid upper bound. However, this method does not produce an upper bound in the general case when the Gaussian assumptions are violated. To show the danger of applying this upper bound technique to spiking neural responses, consider a simple example where the input is a vector of  $M$  independent, identically distributed (iid) Gaussian random variables with  $S_i \sim N(0, \sigma^2)$ . The  $M$  output bins are Bernoulli random variables with  $p(R_i = 1|S_i \geq 0) = q$  and  $p(R_i = 1|S_i < 0) = (1 - q)$ . Though this is only a simple example, it could represent a spike response with probability  $q$  of firing in a bin when the input is positive and probability  $(1 - q)$  of firing when the input is negative. We can calculate  $H(R_i)$  and  $H(R_i|S_i)$  analytically to find the true mutual information rate. Using the technique described above, the mean response vector is  $\bar{S}_i = E[R_i] = \frac{1}{2}$ . The autocorrelation of  $\bar{S}$  is a triangle function, and the resulting power spectrum is  $P_{\bar{S}\bar{S}}(f) = \text{sinc}^2(\pi f M)$ . The noise vectors are  $N_i = R_i - \bar{S}_i$ , with a power spectrum of  $P_{NN}(f) = \frac{1}{4}$ . The upper bound on the mutual information rate is calculated according to equation (2), and the results (calculated numerically) are plotted in figure 1. Though this quantity is claimed as an upper bound applicable to spike responses, it simply is not valid when the (unrealistic) assumptions are violated.

### 3 Lower bound methods

There are also two primary methods for calculating a lower bound on  $I(S; R)$ . The first method [5, 6] involves estimating the stimulus  $\hat{S}$  from the spike

response  $R$ . Due to a result known as the data processing inequality [3],  $I(S; \hat{S}) \leq I(S; R)$  regardless of the estimator used. This method will always produce a lower bound, but is estimator specific. Some estimators could produce very tight bounds if they are based on an accurate model of the neural coding process. Other estimators could produce the trivial bound of zero. As [5] points out, this technique could be useful for checking the validity of a coding model *if* we knew the true value of  $I(S; R)$  or a correct upper bound.

The second method for calculating a lower bound [5, 7] again assumes that the signal  $S$  is Gaussian and the noise is additive and Gaussian. This method calculates the coherence function  $\gamma^2(f) = |P_{RS}(f)|^2 / (P_{RR}(f)P_{SS}(f))$ , using the cross-power spectrum  $P_{RS}(f)$ . In the additive Gaussian noise case,  $\text{SNR}(f) = \gamma^2(f)/(1 - \gamma^2(f))$ , and the mutual information rate is found from equation (2). This method is claimed as a lower bound to the true mutual information because it only accounts for linear correlations between  $S$  and  $R$  [7, p.967], but no proof is given that it holds for systems beyond those that simply add Gaussian noise. At the very least, the best linear fit for the correlation can be zero even though there is significant correlation present. Consider a modified version of the example from section 2, where the input is again a vector of  $M$  (iid) Gaussian random variables with  $S_i \sim N(0, \sigma^2)$ . The  $M$  output bins are Bernoulli random variables with  $p(R_i = 1 | (|S_i|) \geq \eta) = q_1$  and  $p(R_i = 1 | (|S_i|) < \eta) = q_2$ , where  $\eta$  is an arbitrary threshold. Again, this is only a simple example, but it could represent a spike response with probability  $q_1$  of firing in a bin when the input magnitude is above a threshold and probability  $q_2$  of firing otherwise.  $H(R_i)$  and  $H(R_i|S_i)$  are again readily calculable, and so the true mutual information rate  $I(S; R)$  is known. In this case, the coherence function essentially boils down to  $\gamma = \text{Cov}(R_i, S_i) / \sqrt{\text{Var}(R_i)\text{Var}(S_i)}$ , and the resulting lower bound is plotted in figure 2. When  $q_1 = q_2$ , this lower bound method produces the trivial lower bound of zero. An analysis based on this bound appears to show that the mutual information is greatest when  $|q_1 - q_2| \approx 1$  and smallest when  $q_1 = q_2$ . The true mutual information actually has the *opposite* characteristics, being largest when  $q_1 = q_2$  and smallest when  $|q_1 - q_2| \approx 1$ . If only this lower bound were used to analyze the system, the results would be very misleading.

## 4 Conclusions

Given a stimulus set  $S$  and an associated probability distribution  $p(S)$ , mutual information measures the dependence between that specific stimulus source and the spike response  $R$ . However, accurate estimation of  $p(R|S)$  requires many stimulus presentations, making mutual information difficult to calculate directly. This is especially true with responses containing complex temporal

and inter-neuron dependencies. If mutual information is directly estimated, techniques have been developed to remove estimation bias and to decompose the measurement into quantities relevant to neural coding [9].

Many of the techniques for bounding the mutual information make the simplifying assumption that spike responses are produced from an additive (independent) Gaussian noise model. We examined two upper bound techniques in this report that were either invalid or are largely determined by experimental choices. The lower bound technique based on the coherence function is not proven to be a true lower bound for spiking neural systems, but an example is shown where the lower bound it produces is minimally useful and potentially misleading. The lower bound technique based on estimating the stimulus from the spike response will always produce a valid lower bound, but is highly estimator dependent and can also produce trivial lower bounds. Because of the critical role mutual information plays in information theoretic analysis techniques such as rate-distortion theory, it is critical that we develop better tools for either its direct estimation, or for calculating bounds that are more appropriate for spike response outputs.

## References

- [1] C. Shannon, W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana, IL, 1949.
- [2] D. Johnson, Four top reasons mutual information does not quantify neural information processing, in: Proceedings of The Annual Computational Neuroscience Meeting, 2002, Chicago, IL, 2002.
- [3] T. Cover, J. Thomas, Elements of Information Theory, John Wiley & Sons, Inc., New York, NY, 1991.
- [4] T. Berger, Rate Distortion Theory, Prentice Hall, Englewood Cliffs, NJ, 1971.
- [5] A. Borst, F. Theunissen, Information theory and neural coding, *Nature Neuroscience* 2 (11) (1999) 947–957.
- [6] F. Rieke, D. Warland, R. de Ruyter van Steveninck, W. Bialek, *Spikes: Exploring the neural code*, MIT Press, Cambridge, MA, 1997.
- [7] M. Juusola, A. S. French, The efficiency of sensory information coding by mechanoreceptor neurons, *Neuron* 18 (1997) 959–968.
- [8] P. Dayan, A. Abbott, *Theoretical Neuroscience*, MIT Press, Cambridge, MA, 2001.
- [9] G. Pola, A. Thiele, K. Hoffmann, S. Panzeri, An exact method to quantify the information transmitted by different mechanisms of correlational coding, *Network: Computation in Neural Systems* 14 (2003) 35–60.