Detecting Correlated Population Responses

Jyotirmai Uppuluri, Don H. Johnson
{*juppu, dhj*}@rice.edu
Electrical & Computer Engineering Department, MS380
Rice University
Houston, Texas 77005–1892

1 Introduction

Modern recording methods and spike-sorting algorithms now allow the simultaneous recording of many neurons' responses and their classification into single-neuron spike trains. The interrelationships among the individual responses determine the joint population code. Thus, we measure the entire population's response instead of simply inferring population behavior from single-neuron responses. These recorded responses can be interdependent in a vast number of ways, and data analysis techniques must face what statisticians call the "curse of dimensionality." To appreciate the magnitude of the problem, assume the recordings are binned in time so that no more than one spike occurs in each bin and that, from bin to bin, the responses are statistically independent. For single-neuron recordings, each bin can be described by the probability of a spike occurring within it, which amounts to determining the neuron's spiking rate for that bin. If the population response consists of N *statistically independent* neurons, each responding at their individual rate, N parameters describe the population response. However, if each neuron's response is statistically dependent on the other neurons' responses, the number of parameters needed to describe the population response's joint probability distribution is $2^N - 1$. This exponential explosion in parameters demands that proportionally more data be acquired to ascertain the parameter values that determine the population's response characteristics.

Rather than trying to estimate the most general model that encompasses the entire response, we take the tactic of information theoretic model *selection criteria*. The idea is to determine which one of increasingly complex models best fits the recordings. Of course, the simplest model assumes statistically independent responses among population members but this model may not describe the data well. A more complex model would be to assume that selected pairs are correlated and the others statistically independent of them. Which is the better description based on the data in hand? How do we select among a large set of possible models for the best description of the data?

2 Model Selection by Normalized Maximum Likelihood

The problem of selecting among models has long been a subject of information theoretic signal processing. In that context, the problem is much broader than choosing population models. For example, is a set of data best described by a Gaussian or non-Gaussian model (the Laplacian, for instance)? In this case, neither model is a special case of the other as it is in the case of population responses. Be that as it may, Rissanen's normalized maximum likelihood approach optimally selects among models in a data-dependent way [4, 5].

Mathematically, a model amounts to a parameterized probability distribution. Letting the data be generically described by the vector $\mathbf{x}$, the probability distribution for the data under a specific model is $P(\mathbf{x};\theta)$, where θ is a vector of parameters. For the case of interest here, the data vector consists of N binary-valued random variables. The n^{th} neuron's response is described by the probability distribution for N Bernoulli random variables $P_b(x_n)$.

$$P_b(x_n) = \begin{cases} p_n & x_n = 1 \\ 1 - p_n & x_n = 0 \end{cases}$$

Thus, the N values of the discharge probability p_n comprise part of the Bernoulli model's parameter vector. The general expression for the joint response probability is quite complicated [2].

$$P(\mathbf{x};\theta) = \left(\prod_{n=1}^{N} P_b(x_n)\right)\left[1 + \sum_{n_i>n_j} \rho_{i,j}\frac{(x_{n_i}-p_{n_i})(x_{n_j}-p_{n_j})}{\sigma_{n_i}\sigma_{n_j}} + \sum_{n_i>n_j>n_k} \rho_{i,j,k}\frac{(x_{n_i}-p_{n_i})(x_{n_j}-p_{n_j})(x_{n_k}-p_{n_k})}{\sigma_{n_i}\sigma_{n_j}\sigma_{n_k}} + \cdots\right]$$

The correlation-coefficient-like quantities $\rho_{i,j,\ldots}$ reflect pairwise, triple-wise, etc. correlations among population members and they, in addition to the response probabilities, comprise the parameter vector θ. All together, the number of these parameters grows exponentially with population size. Simpler models would have some of these parameter values fixed. For example, a simple model would assume statistically independent responses and set all the correlation coefficients to zero. A more elaborate model would allow pairwise correlation and only set higher correlation coefficients to zero. Whatever the model may be, let θ_m denote its parameter vector. The m^{th} model is thus described mathematically by $P(\mathbf{x};\theta_m)$ and assume we have M models.

The maximum likelihood estimates $\widehat{\theta}_m$ of a given model's parameters is that set of values that maximizes the model's probability for the observed data. Thus, $P(\mathbf{x}_1,\ldots,\mathbf{x}_R;\widehat{\theta}_m) \geq P(\mathbf{x}_1,\ldots,\mathbf{x}_R;\theta_m)$ for all possible θ_m, where $\mathbf{x}_1,\ldots,\mathbf{x}_R$ corresponds to the data collected from the population with R stimulus presentations. The normalized maximum likelihood approach to model selection amounts to choosing the model having the largest probability. The complication is that the maximum likelihood parameter estimates depend on the data, resulting in $P(\mathbf{x}_1,\ldots,\mathbf{x}_R;\widehat{\theta}_m)$ no longer being a probability distribution: it does not sum (integrate) to one. Thus, we *must* normalize the result before deciding which model is most likely. Otherwise, complex models, conceivably having as many parameters as data values, would have only one description and yield a perfectly fitting model. Only the normalization process can reject this overfitting situation.

Unfortunately, the normalization constant we need is often impossible to calculate analytically. Clarke and Barron [1] derived an amazingly simple, highly accurate approximation to this constant that applies universally. By convention, the normalized maximum likelihood technique for model selection chooses that model that *minimizes* the negative logarithm of the normalized model.

$$\min_m \left[-\ln P(\mathbf{x}_1,\ldots,\mathbf{x}_R;\widehat{\theta}_m) + \frac{\dim(\theta_m)}{2}\ln\frac{R}{2\pi} + \ln\int\sqrt{|\mathbf{F}(\theta_m)|}\,d\theta_m + o(1)\right]$$

The first term is the model with maximum likelihood estimates substituted for parameter values. The remaining terms form the approximation to the logarithm of the normalization constant. The $o(1)$ term corresponds to a quantity that decreases to zero as R increases. We found that ignoring this term does not affect the Bernoulli model's results. The integral involves $|\mathbf{F}(\theta_m)|$, the determinant of the Fisher information matrix. Interestingly, this integral can be easily shown to not depend on how the parameters are selected to describe a given model. We found, for example, that expressing a two-neuron model by the parameter vector $\theta = \{p_1, p_2, \rho_{1,2}\}$ is intractable. However, if the model is expressed by the joint probabilities of spike occurrence, $\theta = \{P(x_1=0,x_2=1), P(x_1=1,x_2=0), P(x_1=1,x_2=1)\}$, calculating the Fisher information integral is easy. In general, for independent-response models, the integral is π^N. No general expression for the totally dependent case could be obtained.

3 Results

Figure 1 shows simulation results for the probability that the normalized maximum likelihood selection process chooses incorrectly. These results disclose much about the issues surrounding model

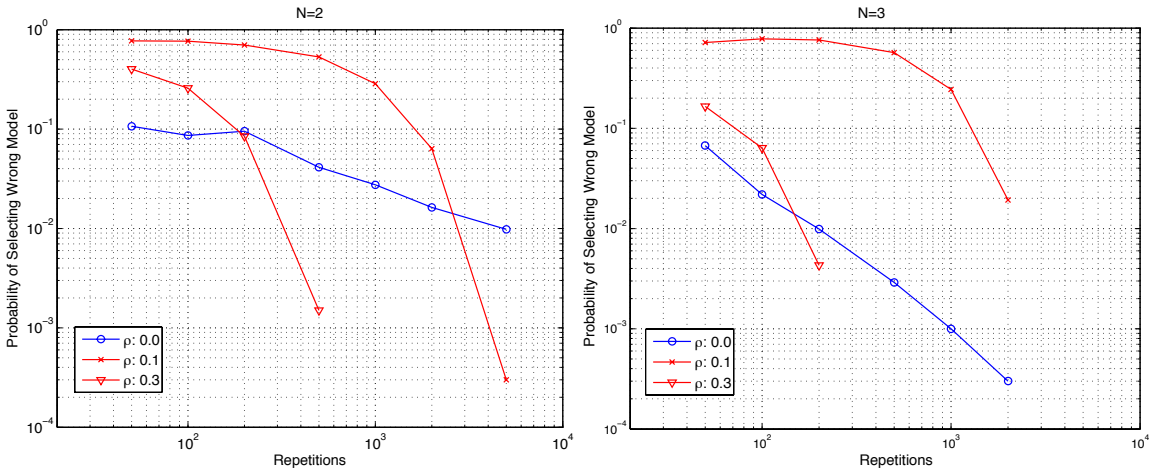

Figure 1: For both panels, the circles indicate the probability (estimated with 10,000 trials) of selecting the correlated model when in fact the uncorrelated model is correct. The other curves show the probability of choosing the uncorrelated model when the responses are indeed correlated. The values of the correlation coefficients are shown in the legend. On the left, we simulated a two-neuron population; on the right, a three neuron one. In both, individual neuron discharge probabilities are 0.1.

selection. First of all, much data is needed to select the right model with high reliability. Even when the independent-discharge model is correct, the probability of selecting a different model decays slowly, roughly proportionally to $R^{-1/2}$ for two neurons and $R^{-3/2}$ for three neurons. Correctly selecting a dependent-discharge model is even more problematic with limited data. We found that the selection error probabilities greater than 0.5 can occur when R is too small. Once a critical amount of data is available, the probability of selecting the wrong model drops precipitously. The situation is somewhat better with the larger population, but small correlations remain difficult to detect. We also found that when the incorrect independent response model best describes the data, the maximum likelihood estimates of the correlation coefficient are very accurate. Thus, relying on the accuracy of parameter estimates does not lead to accurate model selection.

In summary, we found it easy to apply the normalized likelihood model selection approach for Bernoulli population response model. One drawback is that it only selects among the models provided to it, and makes no judgement about the suitability of the best fitting one. Our approach described in a previous paper [3] can remedy the latter flaw. Even when the best model is suitable, the amount of data needed to discern subtle interneuron response dependencies can be much larger than experiments can provide. Very large population sizes can only make this situation worse, which means that determining population codes entirely from data remains a difficult, potentially insurmountable problem.

References

1. B.S. Clarke and A.R. Barron. Information-theoretic asymptotics of Bayes methods. *IEEE Trans. Info. Th.*, 36:453–471, 1990.
2. I. Goodman and D.H. Johnson. Multivariate dependence and the Sarmanov-Lancaster expansion. *J. Multivariate Analysis*, 2005. Submitted.
3. D.H. Johnson and J. Uppuluri. Finding likely models that describe population responses. In *Computational Neuroscience '04*, Baltimore, Maryland, 2004.
4. J. Rissanen. Fisher information and stochastic complexity. *IEEE Trans. Info. Th.*, 42:40–47, 1996.
5. J. Rissanen. Strong optimality of the normalized ML models as universal codes and information in data. *IEEE Trans. Info. Th.*, 47:1712–1717, 2001.