

# Statistical Assessment of Time-Varying Dependence Between Two Neurons

Valérie Ventura, Can Cai, and Robert E. Kass

June 24, 2004

Department of Statistics

and

Center for the Neural Basis of Cognition  
Carnegie Mellon University

*Running Head:* Smoothed time-varying synchrony

## ABSTRACT

The joint peristimulus time histogram (JPSTH) provides a visual representation of the dynamics of correlated activity for a pair of neurons. There are many ways to adjust the JPSTH for the time-varying firing-rate modulation of each neuron, and then to define a suitable measure of time-varying correlated activity. Our approach is to introduce a statistical model for the time-varying joint spiking activity so that the joint firing rate can be estimated more efficiently. We have applied an adaptive smoothing method, which has been shown to be effective in capturing sudden changes in firing rate, to the ratio of joint firing probability to the probability of firing predicted by independence. A bootstrap procedure, applicable to both Poisson and non-Poisson data, was used to define a statistical significance test of whether a large ratio could be due to chance alone. A numerical simulation showed that the bootstrap-based significance test has very nearly the correct rejection probability, and can have markedly better power to detect departures from independence than does an approach based on testing contiguous bins in the JPSTH. In a companion paper (Cai *et al.* 2004b) we show how this formulation can accommodate latency and time-varying excitability effects, which can confound spike timing effects.

**Keywords:** Bootstrap, Correlation, Cross-correlogram, Excursion significance test, Firing rate, Gamma process, IMI process, Non-Poisson Spiking, Peri-Stimulus Time Histogram, Point Process, Smoothing, Statistical model, Spike Train Analysis

# 1 Introduction

Advances in multineuronal recording methods have increased interest in studying the simultaneous activity of two or more neurons. The timing of potentially correlated activity, relative to presentation of a stimulus or occurrence of some behavior, is often studied. An immediate concern is that fluctuations in firing rate will create fluctuations in near-coincident spiking behavior, even if the neurons act independently. Thus, any attempt to quantify time-varying correlated activity must take account of time-varying firing rates. In addition, it is possible that the neurons respond with differing time lags relative to the stimulus or behavior. Therefore, in examining spike trains from a pair of simultaneously-recorded neurons, the activity of one neuron at peri-stimulus times  $t$  must be examined together with the activity of the second at times  $t + \delta$ , for many lags  $\delta$ , as in Aertsen *et al.* (1989). From these considerations two statistical problems emerge. The first is to assess deviations from the joint spiking behavior that would be expected if the two neurons were statistically independent. The second is to acknowledge the large number of assessments that must be produced, across both peri-stimulus times  $t$  and lag times  $\delta$ , which may lead to spuriously significant results. (For example, one thousand significance tests at the .05 level would be expected to yield 50 “significant” assessments by chance alone.) On the one hand, potentially large deviations from expected coincident spike rates that occur at many neighboring times  $t$  would be much more convincing than those that might occur at isolated times. This suggests smoothing the assessments across time. On the other hand, smoothing alone does not eliminate the opportunities for statistical false alarms, so it is highly desirable to have a global evaluation of statistical significance.

A starting point is to collect the joint spike counts for the pair of neurons, across times  $t$  and  $t + \delta$ , at an appropriate time resolution (such as 1 millisecond). These may be displayed with the joint peri-stimulus time histogram (JPSTH). In order to take account of firing rate variation in the neurons, Aertsen *et al.* (1989) proposed a normalized version of the JPSTH and showed how the normalized departures from expected joint spike counts could

be assessed under the hypothesis that the two neurons fired independently. Aertsen *et al.* also recognized the importance of smoothing (see also Kass *et al.* 2003, and the appendix) but they did not provide a global evaluation of statistical significance. In this paper we use a smoothed version of the JPSTH to define an alternative, more powerful, test, and we show how the bootstrap may be used to evaluate global statistical significance (Davison and Hinkley 1997; Efron and Tibshirani 1993). See Ventura (2004) for an overview of bootstrap testing and model selection in the analysis of spike train data.

A short summary of our procedure is as follows. If two neurons were to fire independently, in the sense of probability theory, then the joint spiking probability would equal the product of the two neurons' individual spiking probabilities. Using the notations  $P^1(t)$  for the probability that neuron 1 spikes at time  $t$ ,  $P^2(t+\delta)$  for the probability that neuron 2 spikes at time  $t + \delta$  and  $P^{12}(t, t + \delta)$  for the probability that both neuron 1 spikes at time  $t$  and neuron 2 spikes at time  $t + \delta$ , if the neurons were independent then the ratio

$$\zeta_\delta(t) = \frac{P^{12}(t, t + \delta)}{P^1(t) \cdot P^2(t + \delta)} \quad (1)$$

would equal 1 for all  $t$  and all  $\delta$ . The quantity  $(\zeta_\delta(t) - 1)$  may be interpreted as the excess proportion, above what is predicted by independence, in the probability that neuron 1 will fire at time  $t$  and neuron 2 will fire at time  $t + \delta$ . As we explain in detail in Sections 2–4, our procedure begins with a smooth estimate  $\hat{\zeta}_\delta(t)$  of the function  $\zeta_\delta(t)$ . Throughout this paper we assume time  $t$  takes on discrete values defined by the recording resolution (such as 1 millisecond). A point process representation in continuous time would be possible, but is not necessary for our purposes.

Our test of excess synchronous activity is based on the magnitude of the excursion, across time, of  $\hat{\zeta}_\delta(t)$  outside certain bounds, centered at 1. When  $\hat{\zeta}_\delta(t)$  is either high above its expected value of 1 for a brief period of time, or moderately above for a substantial period of time, the magnitude of the excursion becomes large, providing evidence against independence. We have used the bootstrap both to define excursion boundaries and to compute statistical significance.

Our bootstrap excursion test was developed with the goal of being useful for small or moderate numbers of trials. In related work, Pipa and Grün (2003) show how to use both a permutation test and a bootstrap test to assess the significance of synchrony using the unitary event coincidence count as the test statistic. Their approach, however, considers only the total coincidence count across the trial interval and does not attempt to assess time-varying excess firing. The method here provides temporal information in the spirit of the normalized JPSTH. An important additional motivation for this work is that the excursion test may be extended to adjust for excess trial-to-trial variability, as described in a companion article (Cai *et al.* 2004b). In Section 3 we present results from simulation studies, under the assumption of Poisson spiking, that show that our test has the correct type I error. In Section 4 we show how the procedure may be extended to non-Poisson spiking, which is important in many settings, and we report additional simulation studies of the non-Poisson case that show that non-Poisson spiking behavior can often be ignored without damaging the properties of the test.

## 2 Bootstrap Significance Test

Figure 1 illustrates the statistical procedure presented in this paper. Panels A–C display the simulated data, while panel D displays  $\zeta_\delta(t)$  and a smooth estimate  $\hat{\zeta}_\delta(t)$ . Under the null hypothesis of independence  $\hat{\zeta}_\delta(t)$  will, because of random fluctuations, differ from  $\zeta_\delta(t) = 1$ . Real departures of  $\zeta_\delta(t)$  from 1, the kind we wish to detect, will be statistically substantial and will be sustained over some interval of time. We therefore measure the deviation of  $\hat{\zeta}_\delta(t)$  from 1 by assessing the magnitude of its excursion beyond 95% probability boundaries for  $\hat{\zeta}_\delta$  based on the assumption of independence. This is illustrated in Figure 1E. Note that some excursions beyond the boundaries remain likely to occur due to chance alone, under the null hypothesis. We evaluate the probability of large excursions, and thereby obtain a  $p$ -value for the statistical significance test.

The computation begins with a set of bootstrap samples, which are first used to compute 95% probability boundaries for  $\zeta_\delta(t) = 1$  under the assumption of independence, and are subsequently re-used to compute the probability of large excursions outside those boundaries. The upper boundary  $h_U(t)$ , at each time  $t$ , is defined to be the value such that the probability of  $\hat{\zeta}_\delta(t) > h_U(t)$ , under independence, is 2.5%; similarly, the lower boundary  $h_L(t)$  is defined to be the value such that the probability of  $h_L(t) < \hat{\zeta}_\delta(t)$ , under independence, is 2.5%. We refer to these boundaries as pointwise null bands. This is explained in Section 2.1. Section 2.2 defines a test statistic based on the null bands, and then specifies how the bootstrap samples drawn to compute the bands are re-used to compute the  $p$ -value of the test statistic. Because this procedure is computationally intensive, in the Appendix we also discuss the use of Normal approximations in creating the null bands, based on a smaller bootstrap simulation.

In specifying the bootstrap significance test we will assume that smooth estimates of the firing rate functions  $P^1(t)$  and  $P^2(t + \delta)$  (smoothed versions of the PSTHs) are available from a preliminary analysis, which we write as  $\hat{P}^1(t)$  and  $\hat{P}^2(t + \delta)$ . A smoothing method may similarly be applied to the joint spike counts at times  $t$  for neuron 1 and  $t + \delta$  for neuron 2 in order to obtain an estimate  $\hat{P}^{12}(t, t + \delta)$  of  $P^{12}(t, t + \delta)$ . An estimate  $\hat{\zeta}_\delta(t)$  of  $\zeta_\delta(t)$  in (1) may thereby be obtained as

$$\hat{\zeta}_\delta(t) = \frac{\hat{P}^{12}(t, t + \delta)}{\hat{P}^1(t) \cdot \hat{P}^2(t + \delta)}.$$

Details about smoothing are provided in the Appendix.

In the work reported here we have applied a recently-developed smoothing method, called BARS, that has been shown to perform very well in similar smoothing applications (DiMatteo *et al.* 2001; Kass and Wallstrom 2002; Zhang *et al.* 2003). However, in many applications Gaussian filtering, which is available in statistical software packages, will be adequate. The advantage of introducing  $\zeta_\delta(t)$  is that we can estimate it relatively efficiently, which in turn will provide increased power to detect synchrony, as shown later in Figure 3. More details can be found in the Appendix; see also Kass *et al.* (2003).

## 2.1 Pointwise Null Bands for $\zeta_\delta(t)$

Assuming the data set consists of  $R$  trials in which spiking events for a pair of neurons are observed, the following steps may be used to obtain 95% pointwise null bands for  $\zeta_\delta(t) = 1$  under the assumption of independence:

1. Simulate  $R$  trials of spike trains for the two neurons independently using the estimated firing rate functions  $\hat{P}_1(t)$  and  $\hat{P}_2(t)$  for the two neurons; this is a bootstrap sample;
2. Obtain a smooth estimate  $\hat{\zeta}_\delta(t)$  based on this bootstrap sample, for each  $\delta$  of interest;
3. Repeat steps 1–2  $N$  times to get  $N$  estimates  $\hat{\zeta}_\delta(t)$ ;
4. For each time  $t$ , define  $h_L(t)$  and  $h_U(t)$  to be the .025 and .975 quantiles of the  $N$  values  $\hat{\zeta}_\delta(t)$  (so that, for example, 2.5% of the sampled  $\hat{\zeta}_\delta(t)$  values lie below  $h_L(t)$ ).

Note that Step 1 is an implementation of a *parametric* bootstrap. An alternative nonparametric bootstrap procedure is also straightforward: form  $R$  joint spike trains by sampling at random with replacement the observed spike trains of neuron 1, and separately of neuron 2. This nonparametric option is appropriate if there are a large number of trials and there is no excess trial-to-trial variability (discussed in Cai *et al.* 2004b). Other non-parametric bootstrap simulations can be found in Ventura (2004).

To determine the bands accurately we have used bootstrap sample sizes of  $N = 1000$ . To obtain  $(1 - Q)\%$  bands for any  $Q$ , instead of 95% bands, we take  $h_L(t)$  and  $h_U(t)$  to be the  $\frac{Q}{2}$  and  $1 - \frac{Q}{2}$  quantiles. To reduce the computation effort, a normal approximation can be used with  $N \doteq 50$ , as described in the Appendix.

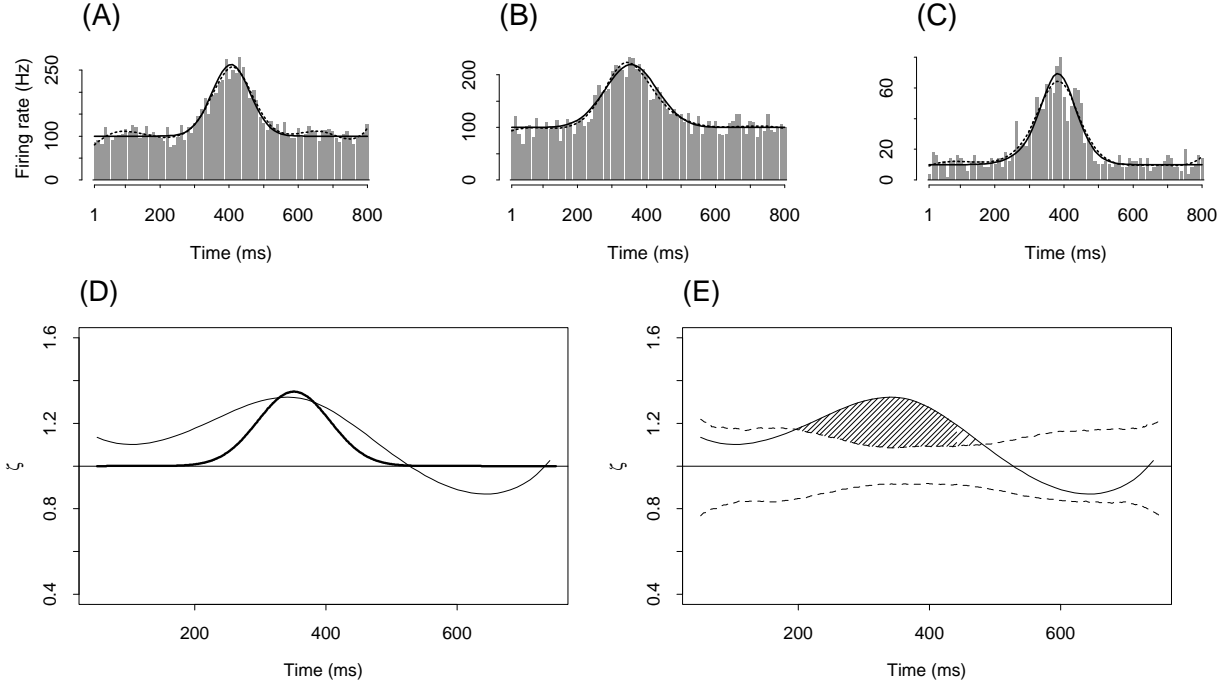


Figure 1: (A) and (B): PSTHs for the two simulated neurons, based on 200 trials, together with BARS estimated firing rates (dashed curves), and true firing rates  $P^1(t)$  and  $P^2(t)$  overlaid in bold. (C): Main diagonal of the JPSTH, together with BARS estimated (dashed), and true joint firing rate  $P^{12}(t, t)$  (bold). (D): True function  $\zeta_0(t)$  defined in Equation (1) (bold), and estimated  $\hat{\zeta}_\delta(t)$  (solid). (E): Estimate  $\hat{\zeta}_\delta(t)$  (solid) with 95% bootstrap confidence bands  $h_{L=.025}$  and  $h_{U=.975}$  (dashed). The excursion  $\hat{G}_{obs}$  of  $\hat{\zeta}(t)$  beyond the bands is the area of the shaded region.



## 2.2 Significance Test for Assessing Time-Varying Synchrony

We define  $G_{obs}$  to be the largest area of any contiguous portion of  $\hat{\zeta}(t)$  that exceeds the bands, where “obs” stands for “observed”, shown as the shaded area in Figure 1E. A mathematical definition of  $G_{obs}$  is given in the Appendix. To calculate its bootstrap  $p$ -value, let  $\hat{\zeta}_\delta^{(n)}(t), n = 1, \dots, N$  stand for the estimate of  $\zeta_\delta(t)$  obtained from the  $n$ th bootstrap sample. For each bootstrap sample we compute  $G_{boot}^{(n)}$ , the largest contiguous portion of  $\hat{\zeta}_\delta^{(n)}(t)$  that exceeds the bands. Then the bootstrap  $p$ -value for  $G_{obs}$  is

$$p = \frac{\text{number of bootstrap samples for which } G_{boot}^{(n)} > G_{obs}}{N + 1},$$

as is standard for bootstrap tests. As an illustration, we obtained  $p < .0001$  for the shaded area in Figure 1E, so we would conclude correctly that there exists synchrony between this pair of neurons at time lag  $\delta = 0$ .

## 3 Properties of the Excursion Test

Formally, a statistical test of hypotheses yields either the correct decision, or one of two types of errors. A type I error occurs when the null hypothesis is rejected erroneously, as we would here if we detect synchronous activity when in fact the two neurons are independent. Conversely, a type II error occurs if we fail to detect synchrony when in fact the two neurons are not independent. In applying a test, the truth is unknown. A good test will have small probabilities of type I and type II errors. However, the probabilities of making the two types of errors vary in opposite directions, and it is impossible to eliminate errors. Additionally, it is often difficult to design a test so that it has specified type I and type II errors. The accepted way to proceed is to build a test that has prespecified probability  $\alpha$  of a type I error, say 0.05, and also has a small probability of a type II error. That is, in statistical jargon, we try to build a powerful test.

In this section we evaluate the operating characteristics of the bootstrap significance test, and compare it to a simpler test based on the normalized JPSTH of Aertsen *et al.* (1989), which we describe below. We consider first, in Section 3.1, whether the bootstrap test has the correct probability of a type I error. Then, in Section 3.2 we consider the power of the test (the probability of rejecting the null when it is truly false) in comparison to that based on the normalized JPSTH.

As a comparison, we used a significance test based on the normalized JPSTH of Aertsen *et al.* (1989), the  $(t, t + \delta)$  pixel of which is

$$\rho_\delta(t) = \frac{P^{12}(t, t + \delta) - P^1(t)P^2(t + \delta)}{\sqrt{P^1(t)(1 - P^1(t))P^2(t + \delta)(1 - P^2(t + \delta))}}, \quad (2)$$

with estimate  $\hat{\rho}_\delta(t)$  taken to be the value of  $\rho_\delta(t)$  when the spiking probabilities  $P^1(t)$ ,  $P^2(t)$  and joint spiking probability  $P^{12}(t, t + \delta)$  are replaced with their observed-data counterparts, i.e., the spike counts divided by the number of trials (and the pixel width). If the two neurons are independent, then  $\hat{\rho}_\delta(t)$  has approximately a standard normal distribution for all  $\delta$  and all  $t$ . Large values of  $|\hat{\rho}_\delta(t)|$  are evidence that the two neurons are correlated at lag  $\delta$ , but because large magnitudes for isolated values of  $t$  would likely be interpreted as chance fluctuations, we define the test to reject the null hypothesis whenever  $|\hat{\rho}_\delta(t)| > z_{\alpha/2}$  for two contiguous values of  $t$ , where  $z_{\alpha/2}$  is the  $\alpha/2$ th quantile of the standard normal distribution. We also consider the corresponding three contiguous values test.

### 3.1 Significance Level – Probability of a False Negative

We examine only the case  $\delta = 0$ , corresponding to synchrony, because we expect performance for other values of  $\delta$  to be similar: the procedures and data are essentially identical (except that for large  $|\delta|$  the data become sparse, as we would be examining data that go into the upper left and lower right corners of the JPSTH, and there the statistical

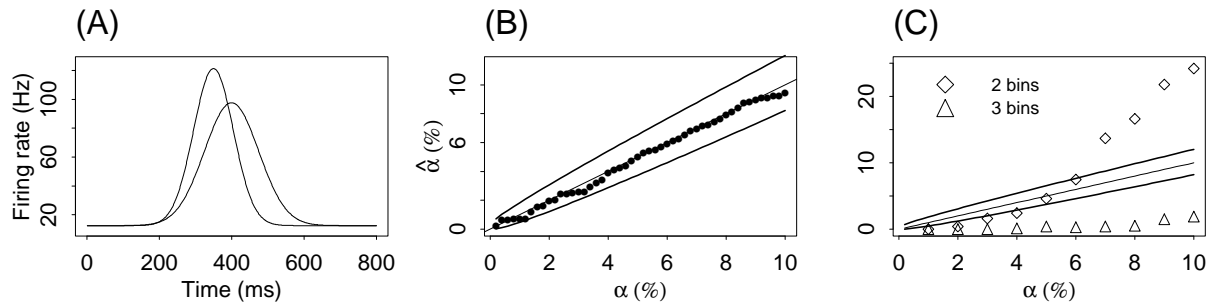


Figure 2: Observed type I error  $\hat{\alpha}$  versus its nominal value  $\alpha$  for (B) the bootstrap excursion test, and (C) for the alternative tests based on contiguous bins of the JPSTH. (A) Firing rate functions used for the two simulated neurons. (B, C) Having set the nominal type I error to  $\alpha$  (shown on the x-axis), the observed type I error is the proportion among 1000 simulated data samples for which the  $p$ -value satisfies  $p < \alpha$ . The solid line corresponds to perfect agreement  $\alpha = \hat{\alpha}$ , and the bold bands are 95% confidence intervals for  $\hat{\alpha}$ . The plot displays extremely good agreement between  $\hat{\alpha}$  and  $\alpha$  for our bootstrap test in (B), but not for the JPSTH based tests in (C).

power drops off sharply). We simulated 1000 pairs of independent neurons, i.e.  $H_0$  true:  $\zeta_0(t) = 1$  for all  $t$ . The firing rates of the neuron pairs are shown in Figure 2A. For each simulated neuron pair, we calculated the  $p$ -value specified in Section 2.2. If we reject the null hypothesis based on  $\alpha = 0.05$ , then the percentage of the  $p$ -values less than 0.05 is the empirical type I error  $\hat{\alpha}$  of the test. Figure 2B displays a plot of  $\hat{\alpha}$  versus  $\alpha$ , for values of  $\alpha$  ranging up to 10%. We find, in this case, that the empirical type I error of our excursion test matches the nominal  $\alpha$  level closely. This is not the case for the JPSTH based tests, as shown in Figure 2C. These tests do not have the correct properties. In fact, the empirical type I error of the JPSTH based tests is an unknown function not only of the prespecified  $\alpha$  and number of contiguous bins considered, but also of the width of the bins.

## 3.2 Power – Probability of a False Positive

To examine the power of significance tests for time-varying dependence we must define time-varying alternatives to the null independence model. In our framework (and again confining attention to  $\delta = 0$ ) this means we must define some set of functions  $\zeta_0(t)$  that are not identically equal to 1. We continue to use two neurons with the firing rates displayed in Figure 2A, and set  $\zeta_0(t) = 1 + 4\beta f(t)$ , where  $f(t)$  is a bell shaped function we took to be the Normal density function with mean 350 msec and standard deviation 55, and  $\beta$  is a gain coefficient taking increasing values 1, 2, 3, 4, 6, 8, and 12, which correspond to a maximum percentage of excess joint firing rate of 2.9, 5.8, 8.7, 11.6, 17.4, 23.2, 34.8, respectively. The  $\zeta_0(t)$  functions we used here are proportional to the  $\zeta_0(t)$  plotted in Figure 1D.

In order for two alternative statistical tests, here the bootstrap-based excursion test versus the two contiguous-bin JPSTH test, to be comparable with respect to power they must be defined to have the same type I error. Because the JPSTH based test does not have the specified type I error, we modified it to reject the null hypothesis whenever  $|\hat{\rho}_0(t)| > \rho^{**}$  for two contiguous values of  $t$ , where  $\rho^{**}$  is a threshold value so that the type I error of this procedure is the prespecified  $\alpha = .05$ . The value  $\rho^{**}$  is not the same as  $z_{\alpha/2}$ , the .05-level threshold for a single  $t$ : it is computed by comparing many alternative thresholds under the null hypothesis and choosing the value that rejects (falsely) 5% of the time.

Figure 3 displays the results of the power calculation. The bootstrap test is able to achieve very good power when the maximum percentage of excess joint spiking activity is around 20%. The JPSTH-based method using (2) is very much less powerful. In fact, when the maximum excess firing rate is 17.5%, four times as many trials would be needed using the JPSTH-based method than using the bootstrap significance test.

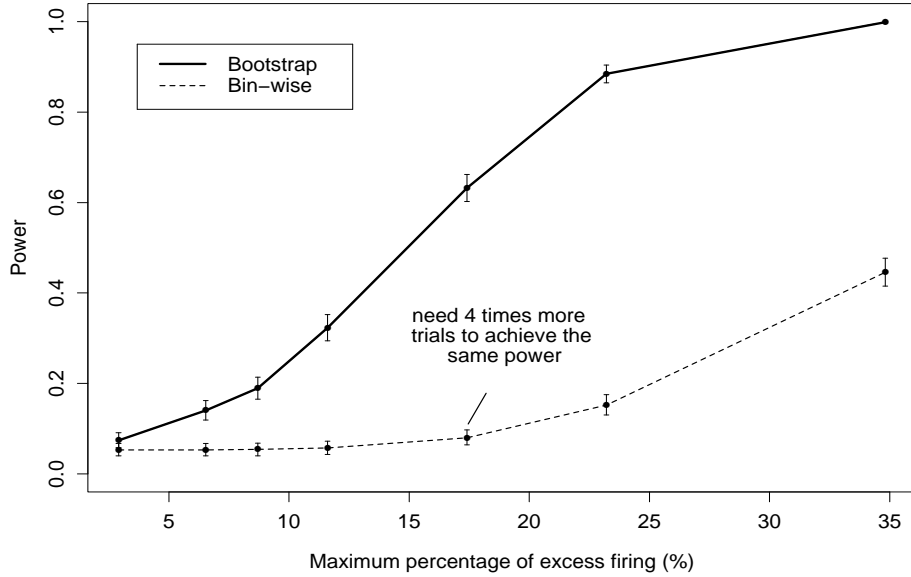


Figure 3: *Power of the bootstrap-based excursion test compared with the test using two contiguous bins of the JPSTH. Power (probability of rejecting  $H_0$ ) is plotted against maximum percentage excess firing ( $\max \zeta_0(t) - 1$ ). Both tests have the same probability of rejecting  $H_0$  when  $H_0$  holds (type I error)  $\alpha = .05$ , indicated by the coincidence of the two power graphs when the percentage excess firing is zero.*

## 4 Non-Poisson Variability

The foregoing material assumed Poisson variation either implicitly, by ignoring spiking history (timing of sequences of spikes) within trials, or explicitly, by generating spike trains from Poisson distributions, both within the bootstrap procedure and in our simulation studies. Two remaining issues are (1) the extent to which the Poisson-based bootstrap  $p$ -value becomes inaccurate in the presence of non-Poisson variability in the observed spike trains, and (2) the manner in which the excursion test be modified for non-Poisson variability, and its resulting statistical properties. We consider these issues in this section.

### 4.1 Robustness of the Excursion Test applied to non-Poisson Spike Trains

To examine the effect of departures from Poisson spiking on the excursion test for synchrony we performed further simulation studies. These consisted of applying the test to non-Poisson data and recording the empirical significance level  $\hat{\alpha}$  to check whether or not  $\hat{\alpha} \doteq \alpha$ .

We considered two types of non-Poisson spike trains. The first type were Poisson, pruned back to enforce a hard refractory period. The second type were gamma spike trains of order  $q$  ranging from 0.05 (much more variable than Poisson processes) to 16 (much less variable). An easy way to understand a gamma process with  $q$  an integer is to simulate one. A Poisson spike train is simulated as follows: divide the time in small intervals centered on times  $t_i$ , and generate a spike in each interval with probability  $P(t_i)$ , where  $P(t)$  denotes the spiking probability at time  $t$ . A gamma process of order  $q$  is defined as the waiting time until the  $q$ th event of a Poisson process; therefore, to generate such a process with firing rate  $P(t)$ , we generate a Poisson process with rate  $q \cdot P(t)$ , but retain only every  $q$ th spikes. For more general  $q$ 's (integer or not, above or below 1), we generate

gamma processes using the time rescaling theorem, as described in Brown *et al.* (2002).

To further illustrate the properties of gamma processes, note that for a Poisson process with constant rate  $\lambda$ , the inter-spike intervals (ISI) have an exponential distribution with mean  $\lambda^{-1}$  and variance  $\lambda^{-2}$ . For a gamma( $q$ ) process with rate  $\lambda$ , the distribution of the ISI is  $q^{-1}$  times a gamma distribution with mean  $q\lambda^{-1}$  and variance  $q\lambda^{-2}$ . Hence the ISI have mean  $\lambda^{-1}$ , as do the ISI of the Poisson process, but variance  $\lambda^{-2}/q$  that is smaller (larger) if  $q$  is larger (smaller) than 1. Therefore the spikes of a gamma process with  $q > 1$  ( $q < 1$ ) occur with more (less) regularity than the spikes of a Poisson process with the same firing rate. Figure 4 illustrates this. It shows raster plots, each based on 10 spike trains generated from gamma processes with rate  $\lambda = 200$  Hz, and  $q = 0.25, 1$  (Poisson), and 4, respectively. The right panels of Figure 4 show the distributions of the number of spikes in a 10 msec window; all have mean 2, as expected, since the average ISI for all spike trains is 5 msec. The larger (smaller)  $q$  is, the smaller (larger) the ISI variability is about the mean. Also, as  $q$  decreases, the probability of observing 10 spikes in a 10 msec window increases.

Figure 5A shows the empirical significance level  $\hat{\alpha}$  of the test applied to a pair of neurons that are Poisson with equal firing rates ranging from  $\lambda = 25$  to 125 Hz, and pruned back to have the same hard refractory period  $d_1 = d_2 = d = 0, 1, 5, \text{ or } 10$  msec. For each combination of  $d$  and  $\lambda$  we simulated 1000 independent neuron pairs, applied the joint spiking model under the Poisson assumption, and calculated the  $p$ -value of the excursion test  $\zeta_\delta(t)$  for  $\delta = 0$ . When  $d = 0$ , so that the two neurons are Poisson,  $\hat{\alpha}$  should not, and indeed does not, differ significantly from  $\alpha$ .

The type I error also remains correct provided the refractory period  $d$  is small compared to the average ISI, which is  $1000\lambda^{-1}$  msec. As a rule of thumb based on Figure 5 and additional simulations, we found that, when both neurons have the same rate and the same refractory period, there should be little concern about applying the Poisson based excursion test to non-Poisson spike trains provided  $d < 0.1(1000\lambda^{-1})$  msec. We next

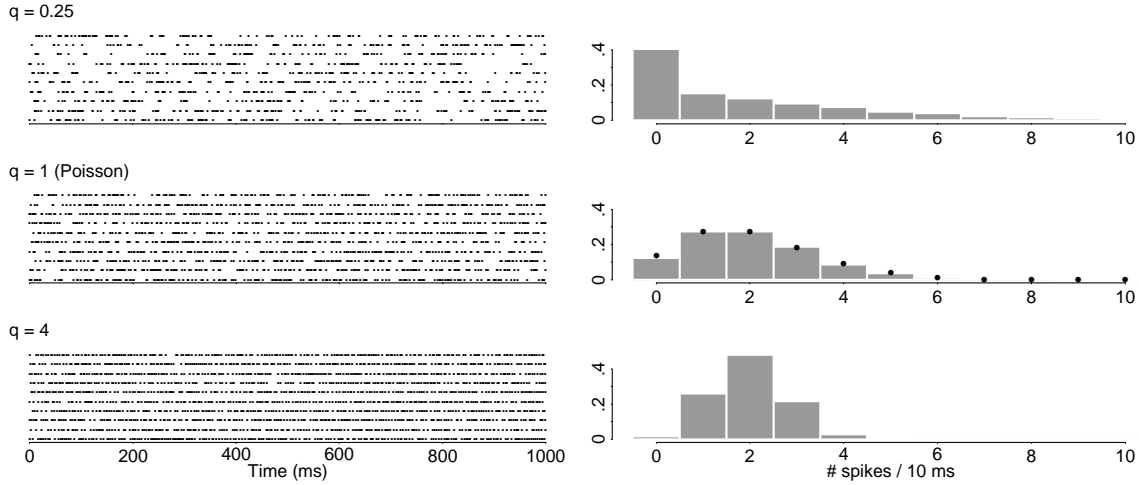


Figure 4: *Properties of gamma processes. Left: raster plots of  $\text{gamma}(q)$  spike trains with firing rate  $\lambda = 200$  Hz. Right: simulated distributions of the number of spikes in 10 msec windows. The black dots in the middle panel are the theoretical distribution of the counts for a Poisson process; simulated and true distributions match.*

investigate the robustness of the Poisson based test when the two neurons do not have the same rates or refractory periods.

Figure 5B shows  $\hat{\alpha}$  when neuron 1 has rate 200 Hz, and extreme ISI to refractory period ratio with  $d_1 = 5$  or 10 msec, so that  $d_1 \gg 0.1(1000\lambda^{-1})$ . We chose the fairly extreme rate of 200 Hz to demonstrate more clearly the limitations of our test. Neuron 2 has refractory period  $d_2$  ranging from 0 to 10 msec, and rate 200, 125 or 50 Hz. Note that when the rate of neuron 2 is as low as 50 Hz,  $0.1(1000\lambda^{-1}) = 2$ , so that by the standards of Figure 5A, the ISI to refractory period of neuron 2 is fairly extreme still for  $d_2 > 2$  msec, yet with no damaging effects on  $\hat{\alpha}$ . This suggests that the departure of  $\hat{\alpha}$  from  $\alpha$  is severe only if both neurons, not just one, have extreme ISI to refractory periods ratios, and that the departure is all the more severe if the refractory periods and the rates of the two neurons are similar.

Figure 5C shows  $\hat{\alpha}$  for a variety of gamma point processes. Neuron 1 is  $\text{gamma}(q_1)$  with



rate 200Hz, while neuron 2 is gamma( $q_2$ ) with rate 130Hz. We observe that the Poisson procedure applied to gamma spike trains produces approximately the correct empirical significance level  $\hat{\alpha}$ , unless  $q_1$  and  $q_2$  are both either very large, or very small.

## 4.2 Non-Poisson Models

Neuronal spiking behavior that follows a Poisson process is determined by the spiking probability  $P(t)$  which, importantly, depends only on time  $t$ . For a general (not-necessarily Poisson) point process the firing rate is governed by the conditional spiking probability  $P(t | \mathcal{H})$ , where  $\mathcal{H}$  is the spiking history (for a given trial) up to time  $t$ . Various non-Poisson alternatives may be used, including the inhomogeneous Markov interval (IMI) models discussed by Kass and Ventura (2001), Gamma process models or other inhomogeneous versions of renewal processes (Barbieri, *et al.*, 2001; Brown *et al.*, 2002). The joint spiking model under the alternative non-Poisson assumption may be fitted just as under the Poisson assumption except the form for the firing rate probability functions,  $P^1(t)$  and  $P^2(t)$ , of each neuron must be changed. We make the simplifying assumption that while the overall joint spiking depends on the spiking history for each neuron through their conditional individual firing rates, their excess joint spiking above that predicted by independence  $\zeta_\delta(t)$  is not itself mediated by recent spiking activity. Some mathematical details are given in the Appendix. The smoothed estimate of  $\zeta_\delta(t)$  is obtained, as before, by smoothing the joint spike counts to obtain a smoothed version of the numerator of (1) and then dividing by the similarly estimated product for the denominator. In our work we have again used spline-based likelihood methods to estimate the conditional intensity functions, following Kass and Ventura (2001).

We repeated the simulation study of the previous subsection, but we used an IMI rather than a Poisson process to define the bootstrap excursion test. Figure 5D displays  $\hat{\alpha}$  in the case where both neurons were Poisson with firing rate 200 Hz and hard refractory

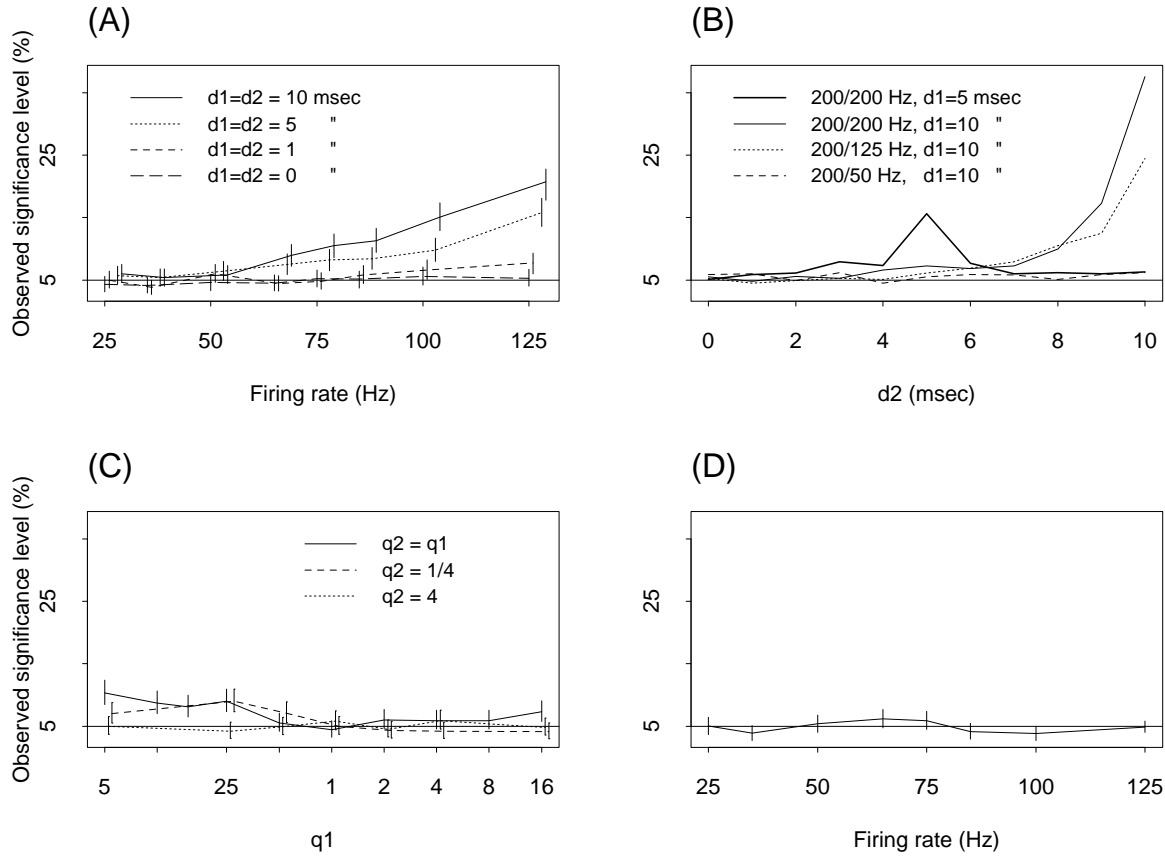


Figure 5: Type I error of excursion test under the Poisson assumption when (A,B) the data are non-Poisson due to a hard refractory period, and when (C) the data are gamma. (A) Both neurons have the same firing rate (x-axis) and the same refractory period  $d_1 = d_2$ . The horizontal line is at  $\alpha = 5\%$ , the type I error the test is supposed to achieve. Vertical bars show the simulation error in the form of 95% confidence intervals; they were omitted in (B) for clarity. (The vertical bars have been separated horizontally to make them more distinct.) (B) Neuron 1 has firing rate 200Hz and refractory period  $d_1 = 5$  or 10 msec; neuron 2 has firing rate 200, 130 or 50 Hz, and refractory period  $d_2$  (x-axis). (C) Neurons 1 is  $\text{gamma}(q_1)$  with rate 200Hz; neurons 2 is  $\text{gamma}(q_2)$  with rate 130Hz. The x-axis is on the log scale for clarity. (D) Type I error of excursion test for the Poisson spike trains with hard refractory period of 10 ms in (A), when model (5) is used in the bootstrap simulation.

period 10 msec. We can see that the type I errors are now indistinguishably close to the desired level  $\alpha = .05$ . The outcome of the IMI based bootstrap test applied to the other truncated Poisson and gamma spike trains of the previous subsection were similar.

## 5 Discussion

The function  $\zeta_\delta(t)$  uses probability to characterize theoretically the evolving dependence in the firing of two neurons. A smooth estimate  $\hat{\zeta}_\delta(t)$ , like a smoothed diagonal of the JPSTH, describes evolving dependence in the data. Because the data are noisy, a statistical procedure must be used to judge whether an observed deviation of  $\hat{\zeta}_\delta(t)$  away from the value of 1, which would hold under independence, is due to chance fluctuations. The magnitude of the excursion of  $\hat{\zeta}_\delta(t)$  beyond the 95% null bounds matches the intuition that increases in joint spiking activity *over continuous intervals of time* above that expected under independence should provide evidence of dependence. Creation of an approximately correct  $p$ -value for any excursion test of this type could be difficult. With the bootstrap, however, it is quite straightforward. A substantial literature on the bootstrap indicates that it has good statistical properties (e.g., Davison and Hinkley, 1997, Sections 2.6 and 5.4, and references therein), and the results presented here indicate that the bootstrap excursion test performs well in the sense of yielding accurate  $p$ -values and having good power for moderate sample sizes.

The particular choice of  $\zeta_\delta(t)$  is not essential to the bootstrap approach applied here, and alternative measures of departure from independence could be used instead. As Ito and Tsuji (2000) observed, there is no uniquely compelling normalization of the JPSTH and each normalization effectively models the occurrence of excess joint spiking activity beyond what would be predicted by independence. Thus, when dependence evolves over time different measures could lead to distinct pictures of the phenomenon. For example, in the presence of time-varying marginal firing rates, excess joint spiking activity that

is constant in time when measured in the ratio form of  $\zeta_\delta(t)$  would appear non-constant when measured in an additive form.

Non-Poisson neurons can produce a greater or smaller number of joint spikes than that predicted under Poisson firing. We discussed and reported on applications of the bootstrap excursion test to non-Poisson spike trains. For mild departures from non-Poisson spiking, and low firing rates, the effects on performance of the Poisson-based test are not large. However, good data-analytical practice would involve checking for non-Poisson behavior by fitting non-Poisson models and, if indicated, applying the non-Poisson versions of the bootstrap excursion test discussed in Section 5.

As Bar-Gad *et al.* (2001) have documented convincingly, omission of spikes due to erroneous spike sorting can have a substantial effect on assessments of correlated activity. We have assumed throughout that the recorded spike trains are accurate representations of neuronal action potential sequences. Under these circumstances, the benefit of the statistical approach adopted here is that it used the information in the data efficiently. In addition, the framework allows us to extend the significance test to include effects such as excess trial-to-trial variation, which is discussed in the companion paper (Cai *et al.* 2004b).

## Appendix

At several points in this paper we simulate joint spike trains. It is easy to do under the assumption that the two neurons are independent. Here, we describe how we simulated spike trains for two correlated neurons with marginal rates  $P^1(t)$ ,  $P^2(t)$ , and joint rate  $P^{12}(t) = P^1(t)P^2(t)\zeta_0(t)$ . This algorithm is valid for Poisson and non-Poisson spike trains. It extends immediately to correlations at other lags. It does not extend to correlations that spread across several lags.

1. Simulate a spike train for neuron A with firing rate  $P^1(t)$ .
2. Simulate the spike train of neuron B conditional on the spike train of neuron A, that is
  - (a) If neuron A had a spike at time  $t$ , generate a spike for neuron B at time  $t$  from a Bernoulli distribution with conditional probability

$$\begin{aligned}
 P^2(t \mid \text{neuron A spiked at } t) &= \frac{P^{12}(t, t)}{P^1(t)} \\
 &= \frac{P^1(t)P^2(t)\zeta_0(t)}{P^1(t)} \\
 &= P^2(t)\zeta_0(t).
 \end{aligned}$$

- (b) If, on the other hand, neuron A did not fire at time  $t$ , generate a spike for neuron B at time  $t$  from a Bernoulli distribution with probability

$$P^2(t \mid \text{neuron A did not spike at } t) = \frac{P^2(t) - P^2(t)P^1(t)\zeta_0(t)}{1 - P^1(t)}.$$

## Section 2 – The Statistical Efficiency of Smoothing

The function  $\zeta_\delta(t)$  was introduced in Equation (1) as a measure of time-varying departure from independence. The advantage of introducing  $\zeta_\delta(t)$  is that we can estimate it relatively efficiently, which in turn will provide increased power to detect synchrony, as was shown in Figure 3.

Figure 1 displays the result of a simulation of two correlated Poisson neurons in which the correlated activity occurs at lag  $\delta = 0$ , i.e., synchronously. Figure 1 also shows the estimate  $\hat{\zeta}_0(t)$  of  $\zeta_0(t)$  obtained by smoothing the sequences  $Y^1(t)/R$  to estimate  $P^1(t)$ ,  $Y^2(t)/R$  to estimate  $P^2(t)$ , and  $Y_0^{12}(t)/R$  to estimate  $P^{12}(t, t + 0)$ , where  $Y^i(t)$  is the number of times out of  $R$  trials neuron  $i$ ,  $i = 1, 2$  fires at time  $t$ , and  $Y_0^{12}(t)$  is the number of times neuron 1 fires at time  $t$  and neuron 2 fires at time  $t = t + \delta$ , with  $\delta = 0$ . To do the smoothing we prefer a spline-based method called BARS (DiMatteo *et al.* 2001) because

it produces relatively good statistical estimates in many contexts, and works particularly well when a firing-rate function varies rapidly in some part of the time domain. However, this particular smoothing method is not essential to the methodology presented here: any other smoothing method could be used, such as Gaussian filtering (see Kass *et al.* 2003).

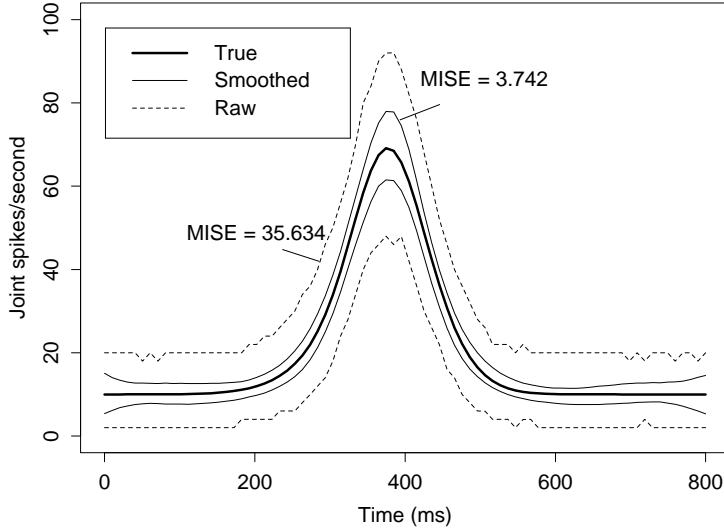


Figure 6: 95% simulation bands for the main diagonal of the JPSTH, plotted as the solid curve, based on 1000 simulated neuron pairs, each having 50 trials. Bands are shown for the raw JPSTH diagonals (without smoothing) and for the smoothed version, using BARS. The bands for BARS are much closer to the true diagonal, indicating much higher accuracy after smoothing. The mean integrated squared errors (MISE) confirm that.

It is not possible to illustrate the efficiency gain of smoothing based on  $\zeta_\delta(t)$ , because an unsmoothed estimate of  $\zeta_\delta(t)$  is not defined at times  $t$  when either  $Y^1(t) = 0$  or  $Y^2(t+\delta) = 0$ . Instead, we illustrate the benefits of smoothing based on the main diagonal of the JPSTH,  $P^{12}(t, t)$  with unsmooth estimate  $Y_0^{12}(t)/R$ , and smooth estimate  $\hat{P}^{12}(t, t)$ , which we take to be a BARS smooth of  $Y_0^{12}(t)/R$ .

Figure 6 shows the true diagonal of the JPSTH we used in Figure 1, along with 95% simulation bands obtained from 1000 simulations from our model for the raw diagonal of

the JPSTH,  $Y_0^{12}(t)/R$ , and for the BARS smooth  $\hat{P}^{12}(t, t)$ . This clearly shows that the variability of the smooth estimate is much smaller. Another way of measuring the quality of an estimate  $\hat{f}(t)$  of  $f(t)$  is to calculate (or approximate) its mean integrated squared error (MISE), that is

$$MISE = \int (\hat{f}(t) - f(t))^2 dt.$$

The MISE values for the smooth and unsmooth estimates of  $P^{12}(t, t)$  are indicated on Figure 6, and suggest that for the unsmoothed diagonal of the JPSTH to have the same accuracy as  $\hat{P}^{12}(t, t)$ , one would need to collect approximately 10 times as much data. See also Kass *et al.* (2003) for additional examples of the advantages of smoothing.

### Section 2.1 – Normal Approximations for the bootstrap null bands of $\zeta_\delta(t)$

As a rough guideline we recommended  $N = 1000$  bootstrap samples to construct 95% null bands for  $\zeta_\delta(t)$ . To reduce computing time, if desirable, a Normal approximation may be used: for large number of trials  $R$  one may assume  $\hat{\zeta}_\delta(t) \sim \text{Normal}(1, \sigma_z^2(t))$  where  $\sigma_z^2(t)$  is the variance of  $\hat{\zeta}_\delta(t)$ , which can be estimated for each  $t$  by the sample variance of the  $\hat{\zeta}_\delta(t)$  computed from a small bootstrap sample (e.g.,  $N = 50$ ).

In practice, to make Normal approximations more accurate a change of variable (or “transformation”) is usually helpful. Our simulations suggest that the square root transformation improves Normality, that is, we may use

$$\sqrt{\hat{\zeta}_\delta(t)} \sim \text{Normal}(1, \sigma^2(t))$$

where, again,  $\sigma^2(t)$  is estimated from a small bootstrap simulation, say with  $R \approx 50$ .

### Section 2.2 – Mathematical definition of the test statistic

Focusing on a particular diagonal specified by  $\delta$ , the null hypothesis of no synchrony between the two neurons at lag  $\delta$  is that  $\zeta_\delta(t) = 1$  for all  $t$ . (The complete null hypothesis of independence is that  $\zeta_\delta(t) = 1$  for all  $t$  and all  $\delta$ ). A test statistic is a formal ordering of the deviations from the null hypothesis, that is, a way of saying which deviations are

greater than others. In our context, deviations of interest will be those that affect many contiguous values of time, so we define our test statistic to be the magnitude of the largest excursion of the estimate of  $\zeta_\delta(t)$ , which we denoted by  $\hat{\zeta}_\delta(t)$ , either above the null band  $h_U(t)$  or below the null band  $h_L(t)$ . More specifically, our test statistic is the largest area between the fitted curve and the null band. This is pictured as the shaded area in Figure 1 for our simulated example. Mathematically, the observed value of the test statistic is

$$G_{obs} = \max \left( \max_{t_s, t_e} \left( \int_{t_s}^{t_e} (\hat{\zeta}_\delta(t) - h_U(t)) dt, \max_{t_s, t_e} \int_{t_s}^{t_e} (h_L(t) - \hat{\zeta}_\delta(t)) dt \right) \right) \quad (3)$$

where  $t_s$  and  $t_e$  are the starting and ending times of the periods during which  $\hat{\zeta}_\delta(t)$  is outside the null bands, and the maximum is taken across all  $t_s$  and  $t_e$ . Then  $G_{obs}$  is the largest area of  $\hat{\zeta}(t)$  exceeding the band.

## Section 4.2 – Non-Poisson Models

When the observed spike trains are severely non-Poisson, it will be safer to apply the bootstrap excursion test based on a model more appropriate than Poisson. Indeed, for non-Poisson spike trains, the spiking probabilities of the two neurons depend on the past, so that Equation (1) must be changed to allow for that, and is thus replaced by

$$P^{12}(u, v) = P^1(u | \mathcal{H}^1) \cdot P^2(v | \mathcal{H}^2) \cdot \zeta(u, v), \quad (4)$$

where  $\mathcal{H}^i$ ,  $i = 1, 2$ , are the respective spiking histories for the two neurons up to time  $t$ . In the special case where we use IMI models to fit the observed spike trains, each neuron's conditional intensity probability function may be written in the form

$$P(t | \mathcal{H}) = P(t, s_*(t))$$

where  $s_*(t)$  is the most recent spike time prior to  $t$  so that Equation (4) becomes

$$P^{12}(u, v) = P^1(u, s_*^1(u)) \cdot P^2(v, s_*^2(v)) \cdot \zeta(u, v). \quad (5)$$

The estimate of  $\zeta(u, v)$  is then obtained as the ratio of smooth estimates of  $P^{12}(u, v)$ , and  $P^1(u, s_*^1(u)) \cdot P^2(v, s_*^2(v))$ , the latter obtained as in Kass and Ventura (2001).



## Grants

Cai and Kass: MH64537

Ventura: MH64537 and N01-NS-2-2346

## References

- Aertsen A, Gerstein G, Habib M, and Palm G.** Dynamics of neuronal firing correlation–Modulation in effective connectivity. *J Of Neurophysiol* 61: 900–917, 1989.
- Barbieri R, Quirk MC, Frank LM, Wilson MA, and Brown EN.** Non-Poisson stimulus-response models of neural spike train activity. *J Neurosci Methods* 105: 25–37, 2001.
- Bar-Gad I, Ritov Y., Vaadia E., and Bergman H.** Failure in identification of multiple neuron activity causes artificial correlations. *J Neurosci Methods* 107: 1–13, 2001.
- Brown EN, Barbieri R, Ventura V, Kass RE, and Frank LM.** A note on the time-rescaling theorem and its implications for neural data analysis. *Neural Comput* 14: 325–46, 2002.
- Cai C, Kass RE, and Ventura V.** Trial-to-trial variability and its effect on time-varying dependence between two neurons, 2004. *Submitted*
- Davison AC and Hinkley DV.** *Bootstrap Methods and Their Applications*. Cambridge University Press, 1997.
- DiMatteo I, Genovese CR, and Kass RE.** Bayesian curve fitting with free-knot splines. *Biometrika* 88: 1055–1073, 2001.
- Efron B and Tibshirani R.** *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.
- Ito H and Tsuji S.** Model dependence in quantification of spike interdependence by joint peri-stimulus time histogram. *Neural Comput* 12: 195–217, 2000.
- Kass RE and Ventura V.** A spike-train probability model. *Neural Comput* 13: 1713–1720, 2001.
- Kass RE, Ventura V, and Cai C.** Statistical smoothing of neuronal data. *Network: Comput Neural Syst* 14: 5–15, 2003.
- Kass RE and Wallstrom G.** Spline adaptation in extended linear models. Invited comment. *Statistical Science*, 17: 24–29, 2002.

**Pipa G and Grün S.** Non-parametric significance estimation of joint-spike events by shuffling and resampling. *Neurocomput*, 52–54: 31–37, 2003.

**Ventura V.** Testing for, and estimating latency effects for Poisson and non-Poisson spike trains. *Neural Comput* 16: 11, 2323-2350, 2004.

# List of Figures

- 1 (A) and (B): PSTHs for the two simulated neurons, based on 200 trials, together with BARS estimated firing rates (dashed curves), and true firing rates  $P^1(t)$  and  $P^2(t)$  overlaid in bold. (C): Main diagonal of the JPSTH, together with BARS estimated (dashed), and true joint firing rate  $P^{12}(t, t)$  (bold). (D): True function  $\zeta_0(t)$  defined in Equation (1) (bold), and estimated  $\hat{\zeta}_\delta(t)$  (solid). (E): Estimate  $\hat{\zeta}_\delta(t)$  (solid) with 95% bootstrap confidence bands  $h_{L=.025}$  and  $h_{U=.975}$  (dashed). The excursion  $\hat{G}_{obs}$  of  $\hat{\zeta}(t)$  beyond the bands is the area of the shaded region. . . . . 6
  
- 2 Observed type I error  $\hat{\alpha}$  versus its nominal value  $\alpha$  for (B) the bootstrap excursion test, and (C) for the alternative tests based on contiguous bins of the JPSTH. (A) Firing rate functions used for the two simulated neurons. (B, C) Having set the nominal type I error to  $\alpha$  (shown on the x-axis), the observed type I error is the proportion among 1000 simulated data samples for which the p-value satisfies  $p < \alpha$ . The solid line corresponds to perfect agreement  $\alpha = \hat{\alpha}$ , and the bold bands are 95% confidence intervals for  $\hat{\alpha}$ . The plot displays extremely good agreement between  $\hat{\alpha}$  and  $\alpha$  for our bootstrap test in (B), but not for the JPSTH based tests in (C). . . . . 9
  
- 3 Power of the bootstrap-based excursion test compared with the test using two contiguous bins of the JPSTH. Power (probability of rejecting  $H_0$ ) is plotted against maximum percentage excess firing ( $\max \zeta_0(t) - 1$ ). Both tests have the same probability of rejecting  $H_0$  when  $H_0$  holds (type I error)  $\alpha = .05$ , indicated by the coincidence of the two power graphs when the percentage excess firing is zero. . . . . 11
  
- 4 Properties of gamma processes. Left: raster plots of gamma( $q$ ) spike trains with firing rate  $\lambda = 200$  Hz. Right: simulated distributions of the number of spikes in 10 msec windows. The black dots in the middle panel are the theoretical distribution of the counts for a Poisson process; simulated and true distributions match. . . . . 14

- 5    *Type I error of excursion test under the Poisson assumption when (A,B) the data are non-Poisson due to a hard refractory period, and when (C) the data are gamma. (A) Both neurons have the same firing rate (x-axis) and the same refractory period  $d_1 = d_2$ . The horizontal line is at  $\alpha = 5\%$ , the type I error the test is supposed to achieve. Vertical bars show the simulation error in the form of 95% confidence intervals; they were omitted in (B) for clarity. (The vertical bars have been separated horizontally to make them more distinct.) (B) Neuron 1 has firing rate 200Hz and refractory period  $d_1 = 5$  or 10 msec; neuron 2 has firing rate 200, 130 or 50 Hz, and refractory period  $d_2$  (x-axis). (C) Neurons 1 is gamma( $q_1$ ) with rate 200Hz; neurons 2 is gamma( $q_2$ ) with rate 130Hz. The x-axis is on the log scale for clarity. (D) Type I error of excursion test for the Poisson spike trains with hard refractory period of 10 ms in (A), when model (5) is used in the bootstrap simulation. . . . . 16*
- 6    *95% simulation bands for the main diagonal of the JPSTH, plotted as the solid curve, based on 1000 simulated neuron pairs, each having 50 trials. Bands are shown for the raw JPSTH diagonals (without smoothing) and for the smoothed version, using BARS. The bands for BARS are much closer to the true diagonal, indicating much higher accuracy after smoothing. The mean integrated squared errors (MISE) confirm that. . . . . 20*