LETTER Communicated by Jonathan Victor

Testing for and Estimating Latency Effects for Poisson and Non-Poisson Spike Trains

Valérie Ventura
vventura@stat.cmu.edu
Department of Statistics and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

Determining the variations in response latency of one or several neurons to a stimulus is of interest in different contexts. Two common problems concern correlating latency with a particular behavior, for example, the reaction time to a stimulus, and adjusting tools for detecting synchronization between two neurons. We use two such problems to illustrate the latency testing and estimation methods developed in this article. Our test for latencies is a formal statistical test that produces a p-value. It is applicable for Poisson and non-Poisson spike trains via use of the bootstrap. Our estimation method is model free, it is fast and easy to implement, and its performance compares favorably to other methods currently available.

1 Introduction

It is often of interest to determine the response latencies of one or more neurons to repeated presentations of a stimulus. One may be interested in adjusting synchronization tools, most commonly the cross-correlogram (CC) and joint peristimulus time histogram (JPSTH), for trial-to-trial effects such as excitability and latency effects (Brody, 1999a, 1999b; Baker & Gerstein, 2001). Another application consists of correlating neural response latency to a particular behavioral variable such as reaction time, as for example, in Everling, Dorris, and Munoz (1998), Everling, Dorris, Klein, and Munoz (1999), Everling and Munoz (2000), Hanes and Schall (1996), and Horwitz and Newsome (2001). We illustrate both applications in section 5.

Methods to estimate latencies are also varied. For example, Brody (1999b) estimates latencies based on minimizing the peak of the CC shuffle corrector, although he cautions that his estimates are flawed when latencies are not, in fact, present. Baker and Gerstein (2001) rectify this problem by providing three alternative estimation methods, all based on detecting the time at which the firing rate increases from the baseline, which we refer to as change-point methods. Two of their estimates possess good properties but are computationally expensive, and depend on an assumed model for the spike trains. Baker and Gerstein (2001) also provide a diagnostic that indicates whether excitability and latency effects may be present, although no

measure of significance, for example, a p-value, is given for this diagnostic. Latency estimation methods have also been developed for continuous response waveforms like electroencephalograms rather than for point processes, although perhaps they could be applied successfully to smoothed spike trains. For example, Woody (1967) takes the latency of a waveform on a particular trial to be the shift that maximizes the correlation between the shifted waveform and a template. In the statistical literature, Pham, Mocks, Kohler, and Gasser (1987) develop a change-point method for continuous signals based on a maximum likelihood paradigm that involves some modeling assumptions and substantial machinery.

In section 2, we describe a method to estimate latencies, which, like the method of Woody (1967), uses the whole duration of the response of the neuron rather than just the time at which the firing rate increases from baseline, as do change-point methods. It compares well to existing methods in terms of efficiency and computational simplicity. Specifically, our estimates require only calculations of sample means so that they are simple and very fast to obtain, do not require any model assumptions, and have smaller biases and variances than other available methods, based on a variety of simulated data, as shown in section 3. Section 4 concerns statistical inferences about latencies. Because estimating nonexistent effects typically adds random noise to statistical procedures, we first propose a formal statistical test for latency effects; by "formal," we mean that we provide not only a diagnostic test but also a p-value. We also explain how variances can be calculated for the latency estimates. The methods of this section are straightforward for Poisson spike trains but require a more careful treatment otherwise, which is why we deferred its treatment to the end of the article. Finally, we validate our methods based on two real applications in section 5 and conclude in section 6.

2 Latency Estimation

The methods developed in this letter rely on a simple result: the spike times of a Poisson process with firing rate $\lambda(t)$ can be viewed as a random sample from a distribution with density proportional to $\lambda(t)$. Although this result applies to Poisson spike trains only, we show that our estimation method applies generally.

Say that K trials of a neuron were recorded under identical repeated presentations of a particular stimulus or experiment. Let $\lambda(t)$ denote the time-varying firing rate of the neuron. If $\tau_k \geq 0$ denotes the latency of trial k and if we assume that the only effect of a latency is to delay the onset of the response to the stimulus by τ_k, then the firing rate of trial k is $\lambda(t - \tau_k)$; hence, its spike times can be considered a random sample with distribution proportional to $\lambda(t - \tau_k)$. Therefore, assuming that there is no source of variability between trials other than latency effects and the random variability associated with the spike generation mechanism, the spike times combined

over all K trials can be viewed as a random sample with distribution proportional to

$$\sum_{k=1}^{K} \lambda(t - \tau_k). \tag{2.1}$$

The mixture distribution 2.1 has the same "shape" as the peristimulus time histogram (PSTH) of the observed trials.

We take our estimates of $\tau_1, \ldots, \tau_k$ to be the values that minimize the variance $V(\underline{\tau})$ of equation 2.1 with respect to $\underline{\tau} = (\tau_1, \ldots, \tau_K)$, that is,

$$(\hat{\tau}_1, \ldots, \hat{\tau}_K) = \text{argmin} V(\underline{\tau}), \quad V(\underline{\tau}) = \text{var}\left(\sum_{k=1}^{K} \lambda(t - \tau_k)\right). \tag{2.2}$$

To make an analogy from probability densities back to spike trains, this criterion finds the set of latency shifts that make the PSTH the narrowest and therefore the highest.

Note that adding the same arbitrary constant τ_0 to all latencies shifts the PSTH of the spike trains by τ_0, but does not change its shape, which suggests that equation 2.2 will produce latency estimates that are defined relative to one another. However, latency estimates are often used to calculate correlations, for example, with a behavioral variable, which are invariant to a constant shift in either or both variables. In particular, only the relative times between two neuron spikes are needed to calculate their CC.

We first select a time window $[T_1, T_2]$, which includes the period of response of the neuron to the stimulus on all trials. We divide $[T_1, T_2]$ into small enough bins so that at most one spike falls into any bin on each trial; these intervals are centered at times t_j. Let $n_{kj} = 0$ or $n_{kj} = 1$ record the absence or presence of a spike at time t_j for spike train k. Letting τ_k denote the true latency for trial k, the variance to be minimized with respect to all τ_k is

$$V(\underline{\tau}) = E_2(\underline{\tau}) - E_1^2(\underline{\tau}),$$

where $E_1(\underline{\tau})$ and $E_2(\underline{\tau})$ are the first two moments of the distribution of all trials combined, given in equation 2.1. Specifically,

$$E_1(\tau) = \frac{1}{K} \sum_k \sum_j h_{kj} \left(t_j - \tau_k\right) = \frac{1}{K} \sum_k \left(\bar{t}_k - \tau_k\right), \tag{2.3}$$

and

$$E_2(\underline{\tau}) = \frac{1}{K} \sum_k \sum_j h_{kj} (t_j - \tau_k)^2, \tag{2.4}$$

where $h_{kj} = n_{kj}/n_{k\cdot}$ are the n_{kj} standardized by the total number of spikes, $n_{k\cdot} = \sum_j n_{kj}$, for trial k, and $\bar{t}_k = \sum_j h_{kj} t_j$. Setting $\partial V(\underline{\tau})/\partial \tau_i$ to zero gives, for $i = 1, \ldots, K$,

$$\tau_i - K^{-1} \sum_k \tau_k = \bar{t}_i - K^{-1} \sum_k \bar{t}_k, \tag{2.5}$$

with solutions of the form

$$\hat{\tau}_k = \bar{t}_k - \tau_0, \quad \text{for all } k = 1, \ldots, K, \tag{2.6}$$

for an arbitrary τ_0. That is, the latency estimate for trial k is the sample mean $\bar{t}_k$ of the spike times that occurred between T_1 and T_2, shifted by an arbitrary amount τ_0. The latencies are relative rather than absolute, since τ_0 can take any value. Should absolute latencies be needed, all we need is the absolute time position T so that $T + \hat{\tau}_k$ become the absolute latencies; this is illustrated in section 2.1.

Because latency estimates are sample means, smoothing the spike trains first does not affect the latency estimates, but it reduces their variability (see section 3). The benefits of smoothing neural data are discussed more generally in Kass, Ventura, and Cai (2003). We used a kernel smoother so that if $n_{kj} = 1$ or 0 denotes the presence or absence of a spike at time t_j, the smoothed spike train has values

$$n^*_{kj} = (T_2 - T_1)^{-1} h^{-1} \sum_i n_{ki} K\left(\frac{t_i - t_j}{h}\right), \tag{2.7}$$

where the summation is over all the time bins in $[T_1, T_2]$, $K(\cdot)$ is the standard normal kernel, and h is the bandwidth, which controls the amount of smoothness. To estimate the latencies from smoothed spike trains, we apply our estimation procedure with h_{kj} in equations 2.3 and 2.4 replaced by $h^*_{jk} = n^*_{kj}/n^*_{k\cdot}$, where $n^*_{k\cdot} = \sum_j n^*_{kj}$.

Before we illustrate this procedure, recall that it is based on the fact that the spike times of a Poisson process can be considered a random sample with distribution proportional to the firing rate $\lambda(t)$. However, because the resulting latency estimates are the sample means of the spike times, the method is also valid for non-Poisson spike trains, since the center of mass of a distribution can be estimated consistently from either a random (Poisson) or a correlated (non-Poisson) sample from $\lambda(t)$; this is illustrated in sections 2.1 and 3. What the distribution of the spike trains has an effect on are second-order properties such as variances of estimates, confidence intervals, and statistical tests of hypotheses (see section 4).

2.1 Implementation and Illustration. There are several options to implement the estimation procedure. For example, one can slide a window of

length $T_2 - T_1$ along each trial until the sample spike times in the moving window match across trials, or one can fix the estimation window $[T_1, T_2]$ and calculate the mean spike times for each trial in that window. We implemented the latter.

We calculate equation 2.6, shift the trials, and iterate. Indeed, even though equation 2.5 is deterministic, $V(\underline{\tau})$ and its derivatives use spikes in the fixed time window $[T_1, T_2]$; therefore, once the trials have been shifted, a not exclusively different set of spikes contributes to the new $V(\underline{\tau})$ and thus yields new latencies; we use them to adjust the previous estimates. Convergence is achieved when the latency adjustments are negligible, or when the decrease in variance $V(\underline{\tau})$ in equation 2.2 is smaller than some $\varepsilon > 0$. The second criterion is easier to implement since it requires monitoring of only one quantity, and we declare that convergence is reached when the change in $V(\underline{\tau})$ is smaller than 1% of $V(\underline{\tau})$ for three successive iterations.

To illustrate the procedure, we simulated $K = 500$ gamma(8) spike trains with firing rate shown in the middle top panel of Figure 2, with a 20 Hz baseline and a 60 Hz stimulus induced rate, then shifted the trials according to latencies sampled from a uniform distribution on [400, 1000] msec. (More details about simulated samples are provided in section 3.) The top center panel of Figure 1 shows a raster plot of the first 30 spike trains, and the top left panel shows the PSTH of all $K = 500$ trials, with the underlying firing rate superimposed; the effect of the latencies can clearly be seen. The top right panel plots the true versus the estimated latencies obtained from the first iteration of the algorithm; both sets of latencies are recentered around zero to compare relative rather than absolute latencies. The successive rows of panels show the result of the successive iterations, with raster plots and PSTHs based on the spike trains reshifted according to the current latency estimates. The algorithm converged after seven iterations.

We used $\tau_0 = \min(\bar{t}_k)$ in equation 2.6 to ensure that $[T_1, T_2]$ contains the response of the neuron to the stimulus even after the trials are shifted. With that choice, the trial with the smallest estimated latency is not shifted, while the other trials get shifted to align with it, as seen in Figure 1. Hence, to transform relative latencies $\hat{\tau}_k$ into absolute latencies $T + \hat{\tau}_k$, we take T to be the time of onset of response to stimulus based on the PSTH of the shifted trials; T can thus be determined by any change-point method or simply by the naked eye. The resulting estimate of T will be accurate since it is based on a PSTH rather than on a single spike train. In Figure 1, the smallest true latency was $\tau_0 = 418.17$ msec, while we obtained $T = 418.54$ msec by clicking on a computer plot of the PSTH of the shifted trials.

2.2 Simultaneously Recorded Neurons. Assume that N neurons are recorded simultaneously and that the same latency applies to all neurons in a particular trial, up to a constant, so that the latency for neuron i in trial k is $\tau_k + \beta_i$, $i = 1, \ldots, N$, $k = 1, \ldots, K$. Our estimation procedure applies as

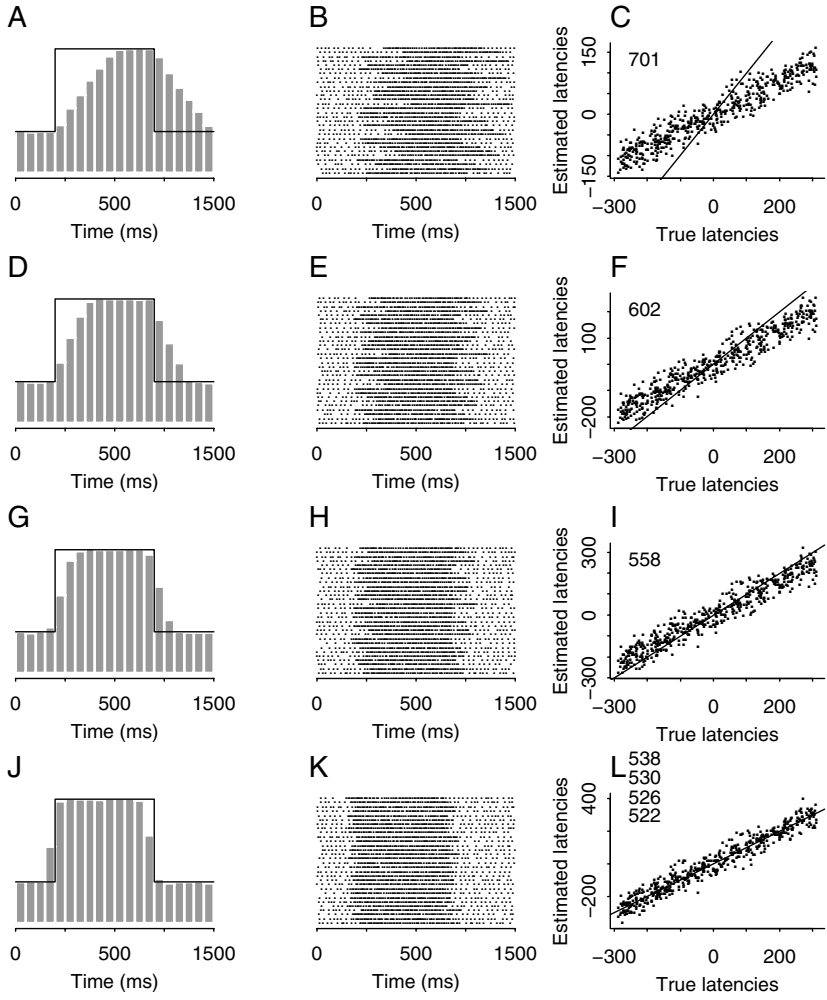


Figure 1: Iterations for latency estimation. Successive rows of panels correspond to successive iterations, until convergence. The data consist of $K = 500$ trials with rate shown in the right panels and latencies uniformly distributed on [400, 1000] msec. The first column shows the PSTHs of the trials shifted by the current latency estimates, and the second column the corresponding rasters; only the first 30 trials were plotted for visibility. The last column shows the centered estimated latencies versus the centered true latencies. The straight line is the first diagonal; the number in the upper left corner is the current value of $V(\tau)$ in equation 2.2. We did not show iterations 4–6, but recorded the successive values of $V(\underline{\tau})$ in the bottom right panel.

before, but with h_{kj} in equations 2.3 and 2.4 replaced by $N^{-1}\sum_{l=1}^{N} h_{kj}^{l}$, where $h_{kj}^{l} = n_{kj}^{l}/n_{k\cdot}^{l}$, and $n_{kj}^{l} = 1$ or 0 indicates the presence or absence of a spike at time t_j in trial k for neuron l. Basically, all this does is combine the spikes of all N neurons for each trial into a "combined" spike train, with rate

$$\Lambda(t) = \sum_{l=1}^{N} \lambda^{l}(t - \beta_l),$$

where $\lambda^{l}(t)$ is the rate of neuron l. This is illustrated in section 5.2 to adjust a CC for latency effects.

3 Procedure Performance and Limitations

In this section, we investigate more fully the properties of our estimates. We assess the effects of the firing rate $\lambda(t)$ and the number of trials K. We illustrate that gains in efficiency can be obtained from smoothing the spike trains and discuss briefly how to choose the estimation window $[T_1, T_2]$. We also show that the latency estimation procedure can readily be applied to spike trains with simple constant excitability covariations in addition to latency effects, but illustrate how more complex excitability effects can bias the estimates.

We measure the efficiency of the estimates using

$$MSE = \left(K^{-1} \sum_{k=1}^{K} [(\hat{\tau}_k - \hat{\tau}_\bullet) - (\tau_k - \tau_\bullet)]^2 \right)^{1/2}, \tag{3.1}$$

where τ_k and $\hat{\tau}_k$ are true and estimated latencies for trial k, and $\tau_\bullet$ and $\hat{\tau}_\bullet$ are their respective means. True and estimated latencies are recentered at zero to compare relative rather than absolute latencies. Equation 3.1 is the square root of what is known in statistics as the mean squared error (MSE); it is a combined measure of bias and variance.

To assess the effect of the firing rate, we use the three rates shown in Figure 2 for a variety of baseline and response to stimulus rates; they are referred to as the step(a, b), block(a, b), and transient(a, b) rates, where a and b are the values of the baseline and of the maximum firing rates. As in Baker and Gerstein (2001), a ranges from 10 to 50 Hz in steps of 10 Hz, and b is two to five times a, in steps of one. For block and transient rates, the duration of the response is 1000 msec, after which the rate returns to baseline. Finally, the latencies are uniformly distributed on [400, 1000] msec. Although Baker and Gerstein used only the step rate, the performances of their estimators also apply to the block rate because they are based on change-point methods. Baker and Gerstein did not consider a transient rate, for which the departure from baseline is perhaps harder to detect.

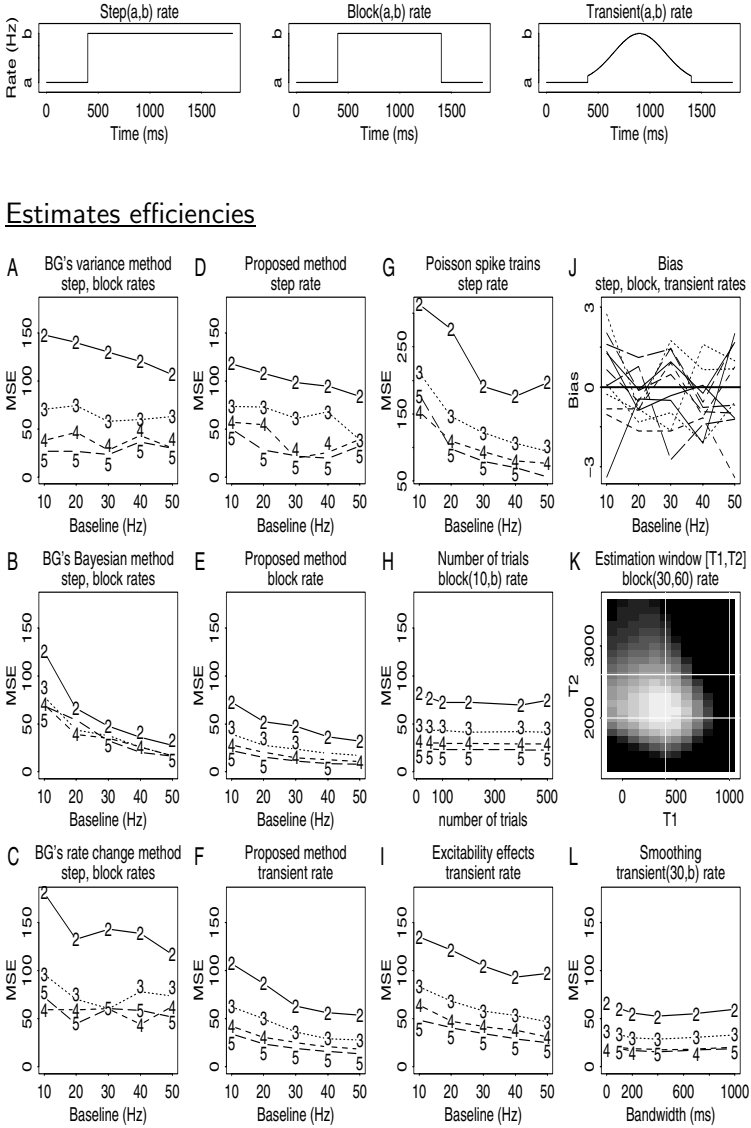


Figure 2: MSE (3.1) of the latency estimates. (A,B,C): Methods in Baker and Gerstein (2001). (All other panels) Proposed method. (K) MSE (3.1) as a function of T_1 and T_2; lighter shades denote lower MSE. (J) Bias for all simulations in *D–F*. All panels but *G* and *H* use 500 gamma(8) spike trains with rate indicated in the title. *G* uses 500 Poisson spike trains. On most panels, the baseline a is on the x-axis, and the maximum rate is $b = k \cdot a$, where k is the plotting symbol.

We use two types of spike generating mechanisms, Poisson and gamma spike trains of order $q = 8$, the latter to match the simulation results in Baker and Gerstein (2001). The former generates spikes independent of the past, while the latter can be used to model spike trains where refractory period effects are present. Recall, however, that our estimation method works generally.

Figure 2D through 2F show equation 3.1, based on $K = 500$ gamma(8) spikes with step, block, and transient rates, from which we conclude that the type of firing rate does have some impact on efficiency. In particular, our method is more effective for firing rates that return to baseline than for sustained rates. Figures 2D and 2E are directly comparable to Figures 2F, 3G, and 4G in Baker and Gerstein (2001), which we reproduced for convenience in Figures 2A through 2C. Figure 2H shows that the efficiency does not depend on the number of trials.

Baker and Gerstein (2001) used the mean of the errors $[\hat{\tau}_k - \tau_k]$ as a measure of the tendency to consistently over- or underestimate the latencies; their methods in Figures 2A through 2C produced biases as large as, in absolute value, 20, 30, and 70, respectively. The equivalent measure for our relative latencies is the mean of $[(\hat{\tau}_k - \hat{\tau}_\bullet) - (\tau_k - \tau_\bullet)]$, which is always zero. We therefore used the median of these errors as our measure of bias, which we plotted in Figure 2J. The bias is close to zero, which suggests that our estimates are randomly scattered around the true relative latencies, as could be seen in Figure 1. We discuss conditions under which estimates can be biased in section 3.1.

Figure 2G suggests that latencies based on Poisson spike trains are less accurate than latencies based on gamma(8) spike trains with the same rate in Figure 2D. This is not surprising since there is more variability in Poisson than in gamma(8) spike times. We discuss further the variances of the estimates in section 4.

Figure 2L shows equation 3.1 as a function of the bandwidth h in equation 2.7 used to smooth the spike trains. It is clear that some efficiency can be gained from smoothing, but that the gain is fairly insensitive to h. However, our experience suggests that convergence is most easily assessed with smaller bandwidths. In the rest of this article, we used $h = 200$ msec.

The efficiencies in Figure 2 all used $[T1, T2] = [200, 2200]$ msec. They can be improved by finer choices of T_1 and T_2. The gray scale in Figure 2K shows equation 3.1, averaged over 1000 simulated data sets, as a function of $T_1 \in [-100, 1000]$ msec and $T_2 \in [1300, 3600]$ msec. We used gamma(8) data with block(30,60) rate; results are qualitatively similar for other spike trains. True latencies are uniform on $[400, 1000]$ msec, so that, based on the PSTH, the response to stimulus starts around 400 msec and reaches its peak around 1000 msec, which is indicated by white vertical lines. The response begins to decline around 1400 msec and returns to baseline around 2000 msec, which is indicated by horizontal white lines. It is clear that our procedure is sensitive to T_1 and T_2. A good choice for T_1 is just before the

firing rate begins to depart from baseline (based on the PSTH). For sustained responses to stimuli (step rate, not shown), we found that the efficiency of our method is not sensitive to the particular choice of T_2, whereas for nonsustained rates (block or transient), a good choice for T_2 is such that the length of $[T_1, T_2]$ is roughly equal to the duration of the response to the stimulus. Finally, even with optimal choices of T_1 and T_2, the efficiencies of the latency estimates are comparable for block and transient rates but not as good for the step or sustained rate, which we observed already in Figures 2D through 2F.

3.1 Limitations. Our latency estimation is based on the assumption that the firing rate is identical across trials except for a time shift. In many contexts, however, the conditions of the experiment or the subject may vary across repeated trials enough to produce discernible trial-to-trial spike train variation beyond that predicted by Poisson or other point processes. We illustrate the consequences of such effects on our latency estimates.

Assume that the state of the neuron varies slowly so that the firing rate of trial k is

$$\alpha_k \lambda(t - \tau_k) \tag{3.2}$$

rather than $\lambda(t - \tau_k)$, where α_k is some positive constant; that is, the firing rate is inflated or deflated by a multiplicative gain α_k on each trial, as pictured in Figure 3A. Because the α_k are constants, firing rates $\alpha_k\, \lambda(t - \tau_k)$ and $\lambda(t - \tau_k)$ are proportional to the same density; hence, they have the same means and thus produce the same estimates of latency. What differs is the number of spike times used for estimation, which affects the variance of the means rather than their average values (see section 4). Figure 2I shows the efficiency of the latency estimates for 500 gamma(8) spike trains with rate 3.2, α_k uniformly distributed on $[0.5, 1.5]$, τ_k uniformly distributed on $[400, 1000]$, and $\lambda(t)$ the transient rate. Figure 2F uses spike trains that are in every way similar but for excitability covariations. We picked the α_k in equation 3.2 so that the total number of spikes across the K trials are comparable for Figures 2F and 2I. This explains why the efficiencies are also fairly comparable.

In practice, excitability effects will likely be more complex than equation 3.2, and depending on the degree of deviation from it, the latency estimates will be biased. A partial solution that sacrifices some efficiency is to reduce the estimation window $[T_1, T_2]$ so that the excitability effects are approximately constant on that window.

This is illustrated in Figure 3, based on gamma(8) spike trains with firing rate for trial k,

$$g_k(t - \tau_k)\, \lambda(t - \tau_k),$$

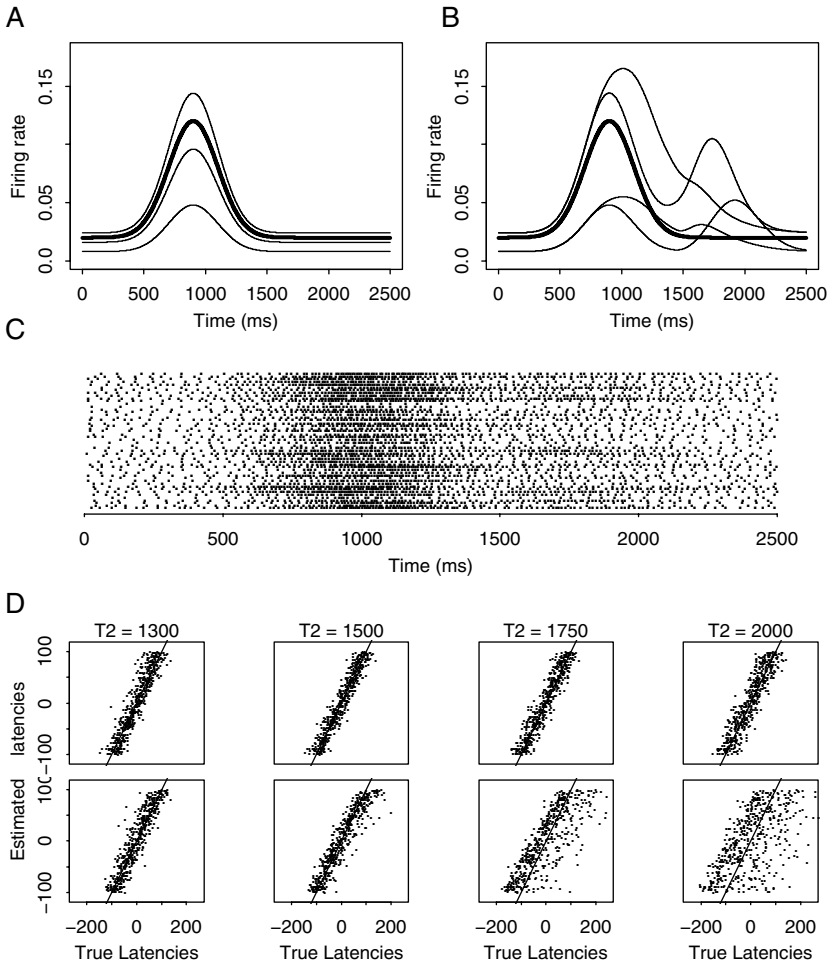


Figure 3: Excitability effects. (A) Firing rates for four trials with multiplicative excitability effects (see equation 3.2). (B) Firing rates for five trials with complex excitability effects. (C) Raster plot for 50 trials with complex excitability effects. (D) True versus estimated latencies for several values of T_2. The first row of panels is for excitability effects in A, and the bottom row for excitability effects in B.

with $\lambda(t)$ the transient(20,120) rate and τ_k uniformly distributed on [400, 600] msec. We generated excitability effects using

$$g_k(t) = \{a_{0k} + a_{1k} \cdot h_1(t, \mu_{1k})\}\{1 + c_k \cdot a_{2k} \cdot h_2(t, \mu_{2k})\},$$

where a_{0k}, a_{1k} and a_{2k} are gain coefficients that are normally distributed with respective means and variances 0.8 and 0.2^2, 0.4 and 0.1^2, and 1 and 0.5^2, and c_r is a Bernoulli random variable with $Pr(c_k = 1) = Pr(c_k = 0) = 0.5$. The time component $h_2(t, \mu_{2k})$ is a normal density with standard deviation 200 and mean μ_{2k} that we take to be normally distributed with mean 1700 and variance 150^2; h_2 adds a later component to the neural response at a random time, for 50% of the trials. The other time component $h_1(t, \mu_{1k})$ is a gamma density function that takes nonzero values when $t > \mu_{1k}$, where $\mu_{1k} = 1800 - 2\tau_k$ depends on the latency; it inflates and lengthens the first peak of the firing rate by random amounts. Figure 3B shows five fairly extreme firing rates from this model, all with latencies $\tau_k = 400$ msec, while Figure 3C shows 50 typical spike trains generated from the model with latencies uniformly distributed on [400, 600] msec.

The estimated latencies from 500 such spike trains are plotted against the true latencies in the bottom panels of Figure 3D, where it is clear that the estimates become severely biased as T_2 becomes larger. For comparison, the top panels in Figure 3D are the corresponding plots from spike trains shown in Figure 3A, with multiplicative excitability effects (see equation 3.2). For T_2 smaller than 1300 msec, latency estimates from trials with either type of excitability effects are comparable.

If excitability effects are so extreme that the firing rate does not have a common component across trials, a change-point latency estimation method may be more efficient.

4 Inference for Latencies

Typical consequences of estimating nonexistent effects on subsequent statistical procedures are loss of statistical power and efficiency. It is therefore important to test if response latencies are constant across trials. We illustrate this below for simulated data, and in section 5 for real data. We also provide standard deviations for the latencies.

To carry out a statistical test, we must choose a test statistic T, determine its distribution under the null hypothesis H_0 that the response latency is constant, which we refer to as the "null distribution of the test statistic," and finally compare this distribution to the observed value t_{obs} of the test statistic, typically via the p-value, to determine if the null hypothesis should be rejected.

Choosing a test statistic is most easily done if we consider once more the analogy between firing rates and distributions. The spike times of trial k can be viewed as a sample from a distribution proportional to $\lambda(t - \tau_k)$. Let μ_k denote its true mean, with estimate the sample mean $\bar{t}_k$ in equation 2.6.

Then the null hypothesis H_0 that all latencies are equal is equivalent to the assumption of equal means in different samples, H_0: $\mu_1 = \cdots = \mu_k = \cdots = \mu_K$, which is routinely dealt with an ANOVA F-test. The ANOVA test statistic is

$$F = \frac{\Sigma_{k=1}^{K} n_k(\bar{t}_k - \bar{t}_\cdot)^2/(K-1)}{\sum_{k=1}^{K}\sum_{j=1}^{n_k}(t_{kj} - \bar{t}_k)^2/(m_\cdot - K)}, \tag{4.1}$$

where t_{kj}, $j = 1, \ldots$ are the spike times of trial k, $\bar{t}_k$ is their mean, $\bar{t}_\cdot$ is the sample mean of the spike times across all K trials, n_k is the number of spikes in trial k, and n is the number of spikes across all K trials.

Large values of F provide evidence against the null hypothesis of constant latency; thus, the p-value is

$$p = Pr(F \geq f_{obs} \mid H_0 \text{ true}), \tag{4.2}$$

where f_{obs} denotes the observed value of equation 4.1 in the sample, and "$\mid H_0$ true" means that the probability is calculated with respect to the null distribution of F. It is commonly assumed that under H_0, the ANOVA F-statistic has a Fisher F-distribution with K and $m_\cdot$ degrees of freedoms, and it is common practice to reject H_0 when equation 4.2 is smaller than 5%, or 1%. This null distribution involves assumptions that we come back to later in this section.

Whatever the outcome of the global test above, it may be of interest to test the latency of a particular trial against a value t_0 or to compare the latencies of any two trials. This can be dealt with in one-sample and two-sample t-tests, respectively; note that a two-sample t-test is equivalent to the F-test (see equation 4.1) applied to $K = 2$ trials. In this article, we consider only the tests of every pair of trials, since choosing a particular point t_0 seems arbitrary; moreover, it seems more relevant to test if particular trials have latencies that differ from those of the other trials, as illustrated in section 5. We report the p-values of all two sample t-tests in a $K \times K$ matrix, where K is the number of trials.

An important point is that the simple excitability effects (see equation 3.2) do not affect either test, because an overall inflated or deflated rate $\alpha_k\, \lambda(t-\tau_k)$ does normalize to the same distribution as the rate $\lambda(t - \tau_k)$. What happens is that excitability becomes a sample size effect that is seamlessly taken care of by the test statistic F via the spike counts n_k in equation 4.1. Of course it is unlikely that excitability effects will be exactly of the simple form of equation 3.2, but they may be sufficiently well approximated by that equation on a small testing window $[T_1, T_2]$. Section 5.2 presents an application where excitability effects are present, yet our testing and estimation procedures appear to work well.

Global and pairwise tests are illustrated in Figure 4A. We simulated 100 Poisson trials with block(20, 60) rate, latency 400 msec for the first 50 trials,

and 700 msec for the other 50 trials. This shows clearly on the raster plot. The p-value (see equation 4.2) for the global test is less than 0.00001, suggesting that latencies are variable across trials. The right-most panel shows the $K \times K$ matrix of the p-values of the pairwise tests, with white, gray, and black cells corresponding to p-values smaller than 1% (strong evidence that the latencies are different), between 1 and 5% (evidence), and above 5% (no evidence), respectively. The two mostly black areas of the matrix correspond to pairs of trials that belong to the same latency group. These areas are not perfectly black because if α denotes the significance level of the test, that is, the probability of erroneously rejecting a true null hypothesis, we should expect to reject approximately a proportion α of tests that have H_0 true. We indeed verified that the proportion of false-positive results (white and gray pixels in the two mostly black areas of the matrix) is just about 5% and the proportion of white pixels about 1%.

The test appears to lack power, since many pairs of latencies that are different are not discovered, as indicated by the dark pixels ($P > 5\%$) in the mostly white areas of the matrix. However, given a fixed testing window $[T_1, T_2]$, the ANOVA F-test is in fact known to be the most powerful test; the power happens to be low in this example because the firing rate is low, which yields small sample sizes, and Poisson trials are quite variable. Figure 4B shows that the same test applied to gamma(8) spike trains with the same rate is more powerful (see section 4.2).

To determine the optimal testing window, we conducted a simulation study similar to that in section 2. The power for the global test was estimated by the proportion of significant tests applied to 1000 simulated samples that have latency effects. We found that the power was sensitive to the testing window and that for all rate types, a good choice for T_1 is just before the firing rate begins to depart from the baseline, and for T_2 approximately when the response to stimulus ends, based on the PSTH of the unshifted trials.

Figure 4A also shows 95% confidence intervals for τ_k obtained as follows. Our estimates $\hat{\tau}_k$ are sample means, so the central limit theorem applies to give

$$\hat{\tau}_k \dot{\sim} N(\tau_k, \sigma_k^2); \tag{4.3}$$

that is, $\hat{\tau}_k$ is approximately normally distributed with mean the true relative latency τ_k, and variance σ_k^2. Hence, a 95% confidence interval for the true τ_k is

$$\hat{\tau}_k \overset{+}{-} 2\sigma_k. \tag{4.4}$$

To estimate σ_k^2 for each trial, we consider once more the spike times to be a random sample from a density proportional to the firing rate. Therefore,

with n_k the number of spikes in $[T_1, T_2]$ for trial k, the estimated variance of $\hat{\tau}_k$ is

$$\hat{\sigma}_k^2 = \frac{S_k^2}{n_k}, \qquad S_k^2 = \frac{\sum_j (t_{kj} - \bar{t}_k)^2}{n_k - 1}, \tag{4.5}$$

where $\bar{t}_k$ and S_k^2 are the usual sample mean and sample variance of the spike times t_{kj}.

The methods proposed in this section so far are fast and straightforward. However, the F-test, t-tests, and confidence intervals' results hold only if the spike times are a random sample from $\lambda(t-\tau_k)$ and if $\lambda(t-\tau_k)$ is proportional to a normal distribution. The second condition is not generally crucial provided the number of spikes in each trial is not too small; this follows from the central limit theorem. The first condition is the most worrisome, since the spike times are random only for Poisson spike trains. For non-Poisson spike trains, we know of no result that specifies the null distribution of F, and thus we will have to either derive it theoretically or obtain an approximation using a bootstrap simulation. The first option is too daunting, if at all possible, so we chose the second.

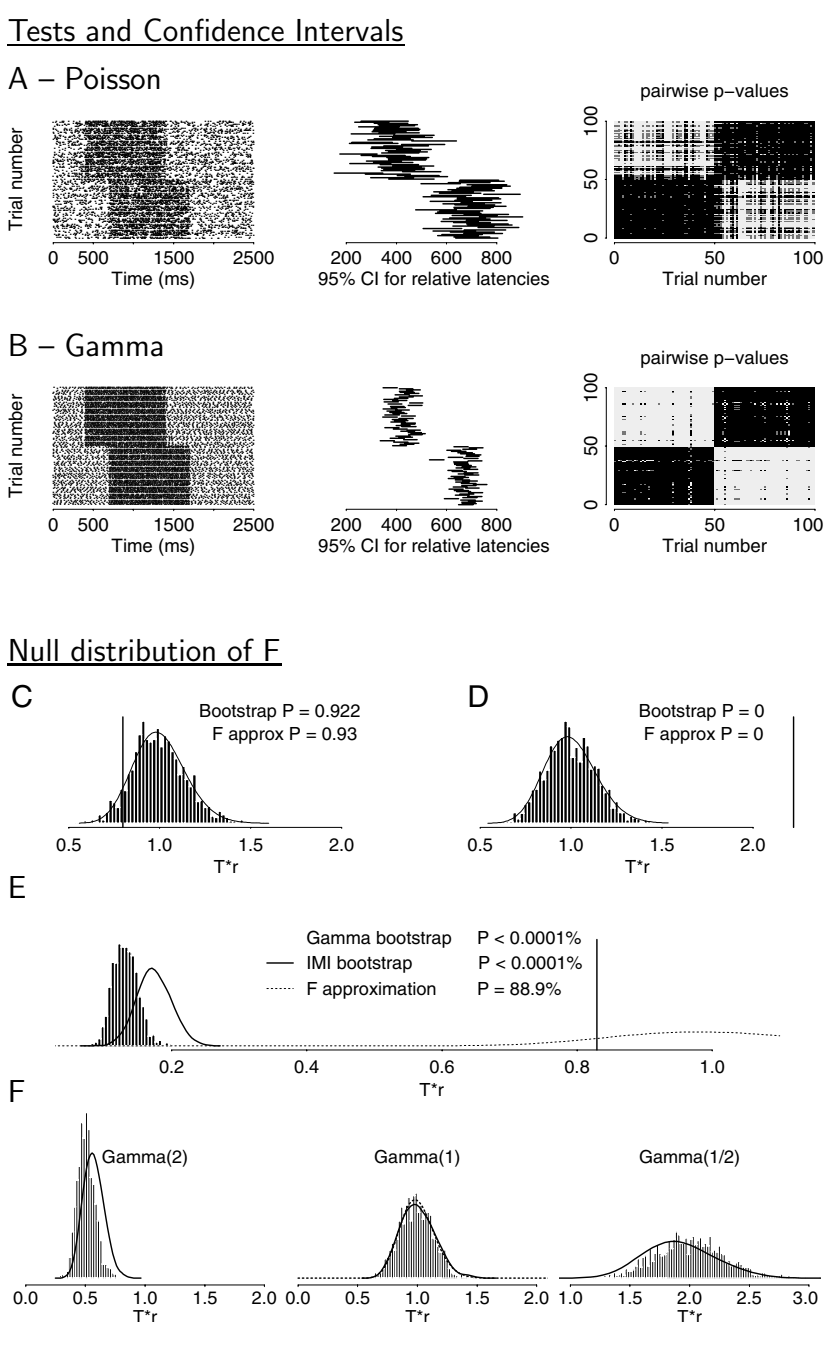
4.1 Bootstrap Inference. In this section, we provide general bootstrap simulation algorithms for tests and confidence intervals, (see Davison and Hinkley, 1997, for a complete bootstrap treatment). Theoretical and bootstrap results are compared in the Poisson case, where we can get both.

Let T denote a test statistic and t_{obs} its value in the sample. For the global test, T is the ANOVA F-statistic (see equation 4.1), and for a pairwise test, T is the two-sample t-statistic, or equivalently the F-statistic (see equation 4.1) evaluated for the two trials under consideration. A general bootstrap testing algorithm follows:

Bootstrap Testing

1. For $r = 1 \cdots R$
   (a) Create bootstrap sample r that satisfies the null hypothesis H_0. Options are described below.
   (b) Calculate t_r^*, the value of the test statistic T in bootstrap sample r.
2. The histogram of the R values of t_r^* approximates the null distribution of T. The bootstrap p-value that approximates the exact p-value in equation 4.2 is:

$$p_{boot} = \frac{1 + \#\{t_r^* \geq t_{obs}\}}{R + 1} \tag{4.6}$$


Tests and Confidence Intervals
A – Poisson
Trial number
0 500 1500 2500
Time (ms)
200 400 600 800
95% CI for relative latencies
pairwise p–values
0 50 100
0 50 100
Trial number
B – Gamma
Trial number
0 500 1500 2500
Time (ms)
200 400 600 800
95% CI for relative latencies
pairwise p–values
0 50 100
0 50 100
Trial number
Null distribution of F
C
Bootstrap P = 0.922
F approx P = 0.93
0.5 1.0 1.5 2.0
T*r
D
Bootstrap P = 0
F approx P = 0
0.5 1.0 1.5 2.0
T*r
E
Gamma bootstrap P < 0.0001%
IMI bootstrap P < 0.0001%
F approximation P = 88.9%
0.2 0.4 0.6 0.8 1.0
T*r
F
Gamma(2)
Gamma(1)
Gamma(1/2)
0.0 0.5 1.0 1.5 2.0
T*r
0.0 0.5 1.0 1.5 2.0
T*r
1.0 1.5 2.0 2.5 3.0
T*r

For Poisson spike trains, bootstrap samples in step 1a can be obtained parametrically or nonparametrically as follows (see Cowling, Hall, & Phillips, 1996, for other sampling options):

Poisson Bootstrap Sample

- Nonparametric. **(i)** Combine the spike times of all K trials. **(ii)** Sample $m. = \sum_k n_k$ spike times from the set of combined spike times, and form the new spike trains by allocating the first n_1 spikes to trial 1, the next n_2 spikes to trial 2, ..., and the last n_K spikes to trial K.
- Parametric. **(i)** Estimate $\hat{\lambda}_0(t)$, the firing rate of the neuron based on the PSTH of the unshifted trials. **(ii)** For $k = 1, \ldots, K$, form trial k by generating n_k spikes from a Poisson process with mean $\hat{\lambda}_0(t)$.

Standard methods to estimate firing rates are gaussian filtering and spline smoothing (see Ventura, Carta, Kass, Olson, & Gettner, 2001).

Figures 4C and 4D show histograms of $R = 1000$ values of t_r^* for two sample of $K = 100$ Poisson spike trains, where T is the F-statistic (see equation 4.1) for the global latency test. Parametric and nonparametric bootstraps produced indistinguishable results. The data in Figure 4C have H_0 true (equal latencies), whereas Figure 4D used data with true latencies uniformly distributed on [400, 1000] msec. The Fisher F-distribution is overlaid; it matches the two histograms almost perfectly, as we would expect, since

Figure 4: *Facing page*. (A) 95% confidence intervals for the latency estimates of the trials in the raster plot, and $K \times K$ matrix of p-values of all pairwise latency tests. Black: $P \geq 5\%$; gray: $1\% < P < 5\%$; white: $P \leq 1\%$. The spike trains are Poisson with block(20,60) rate, latencies 400 msec for the first 50 trials, and 700 msec for the other 50 trials. (B) Same as A but for gamma(8) spike trains. (C, D) Asymptotic F-approximation (solid curves) and bootstrap null distributions (histograms) for the F-statistic (see equation 4.1), with observed values f_{obs} indicated by vertical lines. The spike trains are Poisson with block(20,60) rate, with (C) no latencies, and (D) latencies uniformly distributed on [400, 1000] msec. (E) Bootstrap gamma (histogram), bootstrap IMI (bold curve), and bootstrap Poisson (dotted curve) null distributions for the test F-statistic (see equation 4.1) with observed values f_{obs} indicated by the vertical line, based on gamma(8) spike trains with block(20,60) rate and latencies uniformly distributed on [400, 600] msec. (We use latencies in [400, 600] msec rather than in [400, 1000] msec so that f_{obs} is within the frame of the plot.) (F) Bootstrap gamma (histograms) and bootstrap IMI (curves) null distributions of (see equation 4.1) based on gamma(2), gamma(1), that is, Poisson, and gamma(1/2) spike trains with block(20,60) rate and latencies uniformly distributed on [400, 600] msec. The dotted curve on the middle panel is the asymptotic F-approximation.

the spike times are random, and the number of trials is large enough that the firing rate does not have to be bell shaped. The bootstrap and F-test p-values, equations 4.6 and 4.2, are very close, and for both data sets, we take the appropriate decision: reject H_0 for Figure 4D, with the conclusion that latencies vary across trials, and fail to reject H_0 for Figure 4C.

The dashed curves in Figure 4E are the null distribution of equation 4.1 obtained by Poisson bootstrap and F approximation, this time for gamma(8) rather than Poisson spike trains, with latencies uniformly distributed on [400, 600] msec. Once again, both distributions are practically identical, and thus they are indistinguishable in the plot, with corresponding p-values equations 4.2 and 4.6 equal to 88.9%; we fail to reject H_0, which is the wrong decision. Also plotted as a histogram is the parametric bootstrap null distribution from a gamma(8) rather than a Poisson model, with rate $\hat{\lambda}_0(t)$ fitted to the unshifted spike trains via gaussian filtering. The resulting null distribution is approximately the "correct" one, since the correct model was used. The corresponding bootstrap p-value equation 4.6, is zero so that we now reject H_0, the correct decision. Figure 4E illustrates that the validity of a test (bootstrap or not) requires an appropriate model for the data, from which bootstrap samples can be simulated. Model selection for spike trains is discussed briefly in section 4.2.

Now a few remarks about bootstrap testing are important. A bootstrap sample should be in every way similar to the observed sample—hence the need for an appropriate model. In the context of statistical testing, bootstrap samples should also conform to the hypothetical reality imposed by H_0. Here, H_0 forces us to assume that the trials have a common firing rate, even if we see with the naked eye that they do not. Both our parametric and nonparametric bootstraps did satisfy H_0: the implicitly (combined spikes) and explicitly estimated common firing rate $\hat{\lambda}_0(t)$ ignored any latency effects, since we fitted the same firing rate to all trials.

If the data also contain excitability effects, the bootstrap samples must also contain that extra source of variability. (This issue is developed further in Cai, Kass, & Ventura, 2004.)

The confidence intervals, equation 4.3, hold for Poisson and non-Poisson spike trains since they are based on the central limit theorem, but the estimate of σ_k^2 proposed in equation 4.5 is valid for Poisson data only. We provide a bootstrap alternative that is valid for any spike trains:

Bootstrap Standard Deviations

1. For $r = 1 \cdots R$
   (a) Create bootstrap sample r.
   (b) For each spike train $k = 1, \ldots, K$ in bootstrap sample r, calculate the mean $\bar{t}^*_{kr}$ of the spike times.

2. For $k = 1, \ldots, K$, the bootstrap estimate of σ_k^2 is the sample variance of the $\bar{t}_{kr}^*$,

$$\hat{\sigma}_k^2 = \frac{\sum_{r=1}^{R} (\bar{t}_{kr}^* - \bar{t}_{k.}^*)^2}{R-1}, \quad \text{where } \bar{t}_{k.}^* = R^{-1} \sum_r \bar{t}_{kr}^*.$$

Note that unlike for testing, the model used to simulate bootstrap samples in step 1a should be fitted to the spike trains first shifted according to the latency estimates. Indeed, the final latency estimates are based on shifted trials.

We applied the standard deviation bootstrap algorithm to Poisson and gamma(8) spike trains. In the Poisson case, we obtained bootstrap standard deviations comparable to the analytic result in equation 4.5. The bootstrap standard deviations for gamma(8) spike trains were used to calculate the confidence intervals in Figure 4B.

In this section, we have shown that the bootstrap compares well to asymptotic results in the Poisson case, which gives it credibility when no such asymptotic results are available. For Poisson spike trains, it will be safe to use the theoretical results, unless the numbers of trials or spikes are very small, or the firing rate is very dissimilar in shape to a normal density; in doubt, a bootstrap test is also easily done. For non-Poisson spike trains, all we need to perform bootstrap inference is an appropriate model fitted to the observed data, from which bootstrap samples can be simulated.

4.2 Model Selection for Spike Trains. The quality of any test, bootstrap or not, depends on how well the chosen model fits the data; by quality, we refer to how well the actual significance level of the test matches the nominal one. This is illustrated below.

Model selection, for spike trains or any other data, involves proposing competing models and determining which fits the data best. Reich, Victor, and Knight (1998) introduced the power ratio statistic to test whether spike trains can be completely characterized by an inhomogeneous firing rate. Brown, Barbieri, Ventura, Kass, and Frank (2002) proposed a goodness-of-fit test based on the time rescaling theorem; a goodness-of-fit test determines if a particular model appears to fit, without requiring an alternative model. In the case where two competing models can be fit by maximum likelihood, it is standard to use the likelihood ratio (LR) test; common tests like *t*-tests, and ANOVA *F*-tests are LR tests; Pearson chi-squared tests are asymptotic approximations to LR tests.

The space of models is infinite, so an exhaustive model search is unrealistic, as is the search for the "true" model. A good start is to consider the point process models that have been found to fit some neural data well, for example, inhomogeneous renewal processes, Poisson processes with refractory periods and integrate-and-fire models. (Barbieri, Quirk, Frank, Wilson, & Brown, 2001; Johnson, 1996; Reich et al., 1998). We also like to consider the

inhomogeneous Markov interval (IMI) models of Kass and Ventura (2001) because they include as particular cases inhomogeneous Poisson and homogeneous renewal process models, and thus are well suited to fit data that do not deviate much from these two large classes of models.

To illustrate these ideas, consider once more the gamma(8) spike trains used in Figure 4E. Likelihood ratio tests were performed to compare Poisson, gamma(8), and IMI models; the (true) gamma(8) model was favored over the others ($P < 10^{-3}$). For the sake of illustration, we still performed a parametric bootstrap based on the IMI model, with resulting null distribution overlaid as a bold curve; it is much closer to, although not equal to, the gamma bootstrap null distribution than the bootstrap Poisson distribution was. The discrepancy between the IMI and the correct gamma(8) bootstrap null distributions happens because inhomogeneous renewal processes are not IMI models (only homogeneous renewal processes are). The test based on the IMI bootstrap is conservative, so that the actual significance level is much smaller than the nominal level.

Figure 4F is the same as Figure 4E, but based on simulated gamma(q) spike trains with $q = 2$, 1, and 0.5, which are, respectively, less, as, and more variable than Poisson spike trains; the F approximation was also plotted (dotted curve) in the middle panel. Although the IMI bootstrap null distributions do not quite match their gamma counterparts, these plots suggest that the IMI model yields reasonable inference for data that are close to Poisson or renewal processes.

5 Applications

This section illustrates our latency testing and estimation method based on two examples rather than draws conclusions about the functional characteristics of the types of neurons we used.

5.1 Correlating Neural Activity and Behavior. Individual neurons in the frontal eye field of a rhesus monkey were recorded while the animal performed a memory-guided saccade task or a delayed-response task, as described in Roesch and Olson (2003). Figure 5 shows the PSTH of 20 identical trials of a particular neuron in that experiment. The cue indicating the direction of eye movement is presented for the short period between the two vertical bars. After a waiting period, the central light is turned off at time $t = 0$, at which point the monkey is to execute the eye movement. For each trial, the onset of movement was recorded. We want to investigate if the latency of neural response predicts (or is correlated with) the onset of movement.

Before we apply the latency test, we check if the spike trains are Poisson. The mean versus variance plot in Figure 5A suggests that the deviation from the Poisson assumption is minimal; this is confirmed by the goodness-of-fit test of Brown et al. (2002; not shown). The test of Cai et al. (2004) did

not detect any excitability effects. We thus used the simple latency test of section 4. The global test for latencies suggests strongly that latencies vary across trials, with a p-value smaller than 10^{-8}. Figure 5C shows the overlaid smoothed PSTHs of the original and shifted trials; as expected, the latter is slightly higher.

The latency estimates were robust to the choice of all estimation windows $[T_1, T_2]$ within the guidelines of section 3. Figure 5B shows the latency estimates plotted against the time of onset of movement, with latencies and onsets, respectively, recentered to have sample mean zero. We conclude that for this particular neuron, there exists a strong correlation (0.72, with $P < 0.00001$) between the latency of the neural response to the stimulus and the onset of eye movement.

5.2 Adjusting the Cross-Correlogram for Latency Effects. The raw CC displays the correlation between the spike trains of two neurons at a series of time lags. It is not used directly to assess synchrony, because other correlation sources typically contribute to it. The most common such source is due to modulations in firing rates following some experimental stimulus, although it is easily accounted for by subtracting the shuffle corrector (Perkel, Gerstein, & Moore, 1967). However, the remaining features in the shuffled corrected CC merely indicate that sources of correlation other than stimulus-induced correlations exist between the two neurons. A potential source is synchrony, but other sources include variations in firing rates' amplitudes and in latencies, which Brody (1999a, 1999b) refers to as excitability and latency effects. Baker and Gerstein (2001) provide a list of references that report such effects. Therefore, for the CC to be a useful and reliable tool of synchrony assessment, it must be adjusted for all possible sources of correlation other than synchrony. The customary way to account for latency effects is to estimate the latencies, realign the spike trains, and proceed with the shifted spike trains as one would with usual spike trains (Brody, 1999b; Baker & Gerstein, 2001).

We illustrate our latency testing and estimation on two neurons recorded simultaneously in the primary visual cortex of an anesthetized macaque monkey Aronov, Reich, Mechler, & Victor, 2003, Figure 1B, units 40106.st; 67.5 degrees spatial phase; other spatial phases produced similar results). The data consist of 64 trials shown in Figure **??**B. The stimulus is identical for all trials, and consists of a standing sinusoidal grating that appears at time 0 and disappears at 237 msec.

The mean versus variance plots in Figure **??**A lie somewhat above the diagonal, which suggests more variability than predicted under the Poisson assumption. The goodness-of-fit tests of Brown et al. (2002) in Figure **??**A indeed confirm that the data deviate slightly from Poisson, since the empirical versus model quantiles curves do not lie completely within the 99% joint confidence bands. The IMI model provides a better fit, although the improvement is only marginal. The slight deviation from Poisson could be

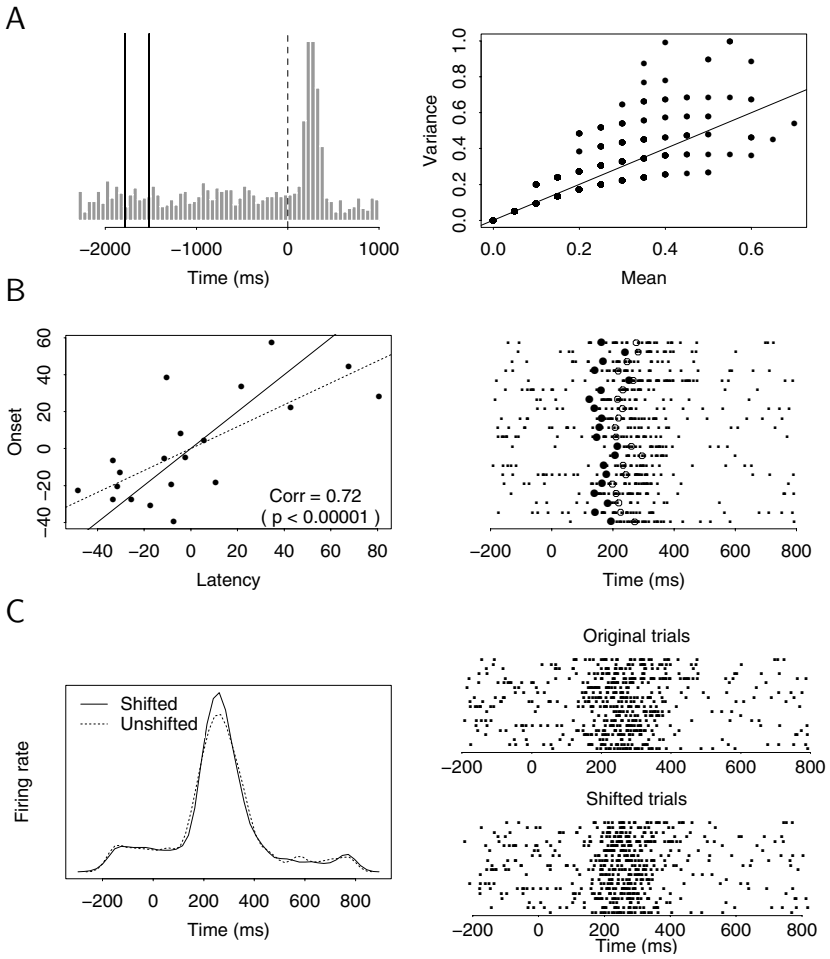


Figure 5: (A) PSTH for a neuron in Roesch and Olson (2003), based on 20 identical trials. The cue indicating the direction of the eye movement is presented between the two vertical bars; the cue for the movement is at time $t = 0$. Mean versus variance plot of the spike trains, based on time bins ranging from 2 to 20 msec. The solid line is the first diagonal. (B) Onset of movement versus neural response latency estimates. The solid line is the first diagonal, and the dotted line the fitted linear regression of onset on latency. Raster plot of the 20 spike trains with filled circles marking the latency estimate and open circles the onset of movement. (C) Overlaid smoothed PSTHs and raster plots of the original and shifted trials.

partly due to excitability effects, which were found to be significant based on the test for excitability in Cai et al. (2004). Although the deviations from the Poisson model may be important from a functional standpoint, they are not large enough from a statistical standpoint to warrant the more complicated bootstrap procedures for non-Poisson spike trains, especially since the evidence for latency effects is overwhelming, as discussed below.

Figure 6B shows the raster plots for the two neurons, from which it appears that the few first and last trials have latencies that differ from those of the other trials. This is confirmed by the global tests for latencies, with a p-value 0.045 for neuron 2, and p-values smaller than 10^{-8} for neuron 1 and for the two neurons combined. Lack of statistical power for neuron 2 is due to the sparseness of the spikes. The value of the test statistic for neuron 1 is $t_{obs} = 3.3$, which is extreme enough as to leave no doubt about the presence of latency effects, under Poisson, gamma or IMI models. Figure 6B also shows the matrix of p-values for all pairs of trials for the two neurons combined, with white and gray pixels corresponding to p-values smaller than 1% and 5%, respectively. If all trials had equal latencies, we would expect about 5% of white and gray pixels, whereas we have about 30%. This confirms the outcome of the global tests. Additionally, there is a clear pattern in the pairwise p-values; the mostly black area in the middle indicates that the latencies of all trials are similar, except for the first few and last trials, as could in fact be seen by the naked eye on the raster plots.

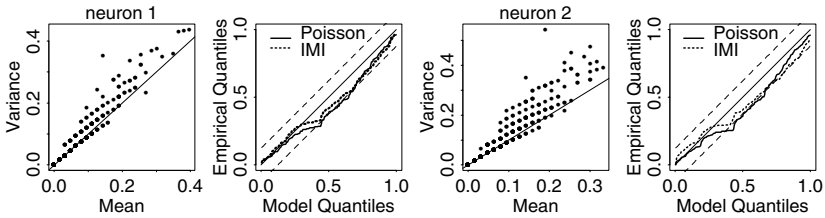
Next, we estimated the latencies for the two neurons separately. The presence of excitability effects is not of overwhelming concern because the estimation windows $[T_1, T_2]$, shown in Figure 6B, are short enough that the effects on these intervals are presumably well approximated by equation 3.2. Figure 6B shows a plot of the latency estimates of one neuron versus the other, which shows that the two sets of latencies covary, with correlation 0.69. To shift the spike trains, we used the latency estimates based on the combined spikes of the two neurons (see section 2.2) because they are less variable.

Finally, Figure 6C shows CC for these data, corrected for the correlation induced by modulations of the firing rates, along with 95% confidence bands. The left-most panel is for the observed spike trains, from which one may conclude that there is synchronous activity at small lags. However, these effects disappear entirely once the CC is adjusted for correlations induced by the latencies. We also produced a CC based on the observed spike trains, after removing the few first and last trials that appear to have different latencies. This CC is qualitatively similar to the CC adjusted for latencies, although the confidence bands are wider because fewer trials are used.

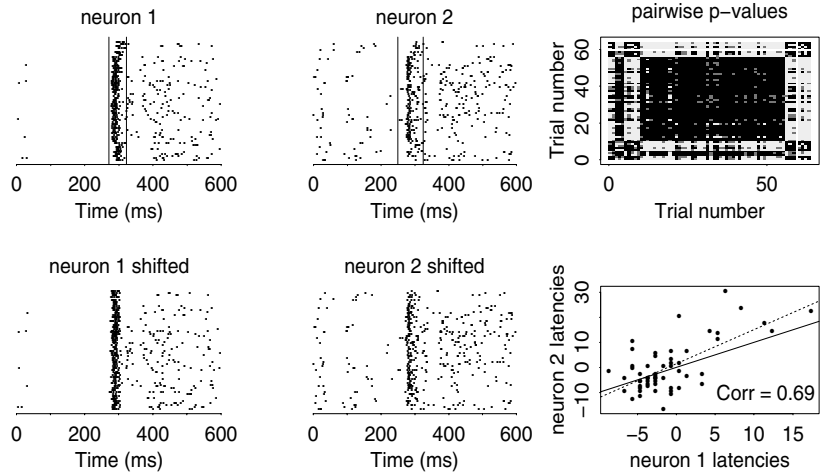
6 Conclusion

We have developed statistical procedures for testing and estimating latency effects in spike trains obtained from identical repeats of an experiment. The

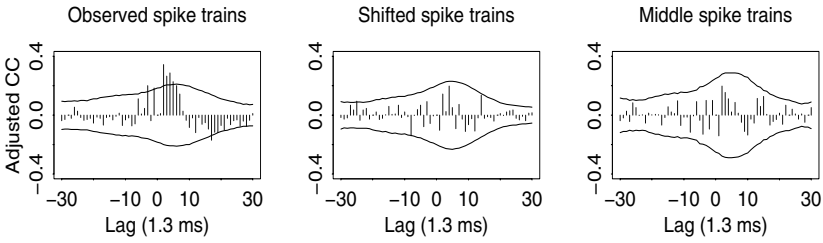
A – Poisson tests



B – Latency testing and estimation



C – Cross-correlograms



main attractions of our methods are their simplicity. Moreover, they appear to be efficient and powerful based on a large number of simulated spike trains. We also applied our methods to two real data sets and obtained results that seem reasonable.

We proposed a formal statistical procedure to test for unequal latencies across trials; by "formal," we mean that we provide not only a diagnostic test but also a p-value. For Poisson spike trains, this test is the usual analysis of variance (ANOVA) test for the equality of several means. For non-Poisson spike trains, we still use the ANOVA F-statistic, although we obtain its null distribution via a parametric bootstrap. We use an inhomogeneous Markov interval (IMI) model to fit non-Poisson spike trains when competing parametric models do not fit the data as well.

Our estimation method consists of finding the set of shifts that minimizes the spread of the resulting PSTH. We show that this minimization criterion is equivalent to finding the shifts so that the means of the shifted spike times are equal in all trials. Therefore, our estimates require calculations only of sample means so that they are simple and very fast to obtain, and they do not require any model assumptions. We applied our method successfully to Poisson and non-Poisson spike trains with various rates, including spike trains that contain simple multiplicative excitability effects.

In situations where more complicated excitability effects are present, our estimates can be biased, so it may be preferable to use a change-point method as, for example, in Baker and Gerstein (2001). But our latency estimates can still be used as starting values in other latency estimation algorithms. Indeed, change-point methods are based on detecting the rate change from baseline, typically on each trial separately, so that the latency estimates thus obtained do not benefit from any information that may be contained in other trials. Our method is based on the differences in firing rate from trial to trial. Including this extra information in change-point estimation procedures is likely to improve them.

Figure 6: *Facing page*. (A) Mean versus variance plots and goodness-of-fit tests for Poisson and IMI models (Brown et al. 2002) for two simultaneously recorded neurons in the primary visual cortex of an anesthetized macaque monkey (Aronov et al., 2003). (B) Raster plots with estimation windows $[T_1, T_2]$ (top), and after the trials are shifted (bottom). Matrix of p-values for all pairwise tests (see section 4). Estimates of latencies for neuron 1 plotted against those of neuron 2, with (0,1) line (solid) and fitted linear regression (dotted line), and sample Pearson correlation. (C) Cross-correlograms adjusted for firing-rate modulation (shuffle corrected) and 95% confidence bands for the observed and the shifted spike trains, as well as for the observed trials that appear to have constant latencies. We used bins of 1.3 msec. The central bin is not plotted because a recording artifact prevents testing of synchrony at lag 0.

Our latency estimates are obtained thus far in a completely nonparametric way. We do not assume a particular model for the firing rate or for the spike generation mechanism; spike trains do not have to be Poisson or gamma. But the absence of a specific model does not preclude assumptions; in particular, our procedure produces meaningful estimates under the basic assumption that the firing rates for all trials are proportional to one another except for a time shift. This begs for two extensions. First, if the basic assumption is met, can we improve our procedure by explicitly making use of the common element, that is, λ, between all trials? If the basic assumption is not appropriate—for example, if more complicated excitability effects exist—can we modify the procedure to allow for this? Our answer to both questions involves a more statistical approach than we have used. A full treatment is beyond the scope of this article, and is, in fact, the topic of a future article that treats the estimation, and adjustment, of latency and excitability effects (Cai et al., 2004).

Acknowledgments

This work was supported by grants N01-NS-2-2346 and 1R01 MH64537-01 from the National Institutes of Health.

References

Aronov, D., Reich, D. S., Mechler, F., & Victor, J. D. (2003). Neural coding of spatial phase in V1 of the macaque monkey. *J. Neurophysiol., 89*, 3304-¬3327.

Baker, S. N. & Gerstein, G. L. (2001). Determination of response latency and its application to normalization of cross-correlation measures. *Neural Computation, 13*, 1351–1377.

Barbieri, R., Quirk, M. C., Frank, L. M., Wilson, M. A., and Brown, E. N. (2001). Construction and analysis of non-Poisson stimulus-response models of neural spike train activity. *J. Neurosci. Methods, 105*, 25–37.

Brody, C. D. (1999a). Correlations without synchrony. *Neural Computation, 11*, 1537–1551.

Brody, C. D. (1999b). Disambiguating different covariation. *Neural Computation, 11*, 1527–1535.

Brown, B. N., Barbieri, R., Ventura, V., Kass, R. E., & Frank, L. M. (2002). The time-rescaling theorem and its applications to neural spike train data analysis. *Neural Computation, 14*(2), 325–346.

Cai, C., Kass, R. E., & Ventura, V. (2004). *Trial to trial variability effects and its effect on time-varying dependence between two neurons.* Manuscript submitted for publication.

Cowling, A., Hall, P., & Phillips, M. J. (1996). Bootstrap confidence regions for the intensity of a Poisson point process. *Journal of the American Statistical Association, 91*, 1516–1524.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their applications*. Cambridge: Cambridge University Press.

Everling, S., Dorris, M. C., Klein, R. M., & Munoz, D. P. (1999). Role of primate superior colliculus in preparation and execution of anti-saccades and pro-saccades. *J. Neurosci. 19*, 2740–2754.

Everling, S., Dorris, M. C., & Munoz, D. P. (1998). Reflex suppression in the anti-saccade task is dependent on prestimulus neural processes. *J. Neurophysiol., 80*, 1584–1589.

Everling, S., & Munoz, D. P. (2000). Neuronal correlates for preparatory set associated with pro-saccades and anti-saccades in the primate frontal eye field. *J. Neurophysiol., 20*, 387–400.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science, 274*, 427–430.

Horwitz, G. D. & Newsome, W. T. (2001). Target selection for saccadic eye movements: Prelude activity in the superior colliculus during a direction-discrimination task. *J. Neurophysiol., 86*, 2543–2558.

Johnson, D. (1996) Point process models of single-neuron discharges. *J. Comput. Neurosci., 3*, 275–299.

Kass, R. E., & Ventura, V. (2001). A spike train probability model. *Neural Computation, 13*, 1713–1720.

Kass, R. E., Ventura, V., & Cai, C. (2003). Statistical smoothing of neuronal data. *NETWORK: Computation in Neural Systems (special issue on Information and Statistical Structure in Spike Trains), 14*, 5–15.

Perkel, D. H., Gerstein, G. L., & Moore, G. P. (1967). Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys. J., 7*, 419–440.

Pham, D. T., Mocks, J., Kohler, W., & Gasser, T. (1987). Variable latencies of noisy signals: Estimation and testing in brain potential data. *Biometrika, 74*(3), 525–533.

Reich, D. S., Victor, J. D., & Knight, B. W. (1998). The power ratio and the interval map: Spiking models and extracellular recordings. *J. Neurosci., 18*(23), 10090–10104.

Roesch, M. R., & Olson, C. R. (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *J. Neurophysiol., 90*, 1766–1789.

Ventura, V., Carta, R., Kass, R. E., Olson, C. R., & Gettner, S. N. (2001). Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics., 1*(3), 1–20.

Woody, C. D. (1967). Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Med. Biol. Engng., 5*, 539–553.

Received August 20, 2003; accepted April 8, 2004.