Statistical Methods for the Assessment of Temporal Structure in the Activity of the Nervous System

by
Asohan Amarasingham
B.A., University of Virginia, 1997
Sc.M., Brown University, 1999

Doctoral dissertation
Submitted in partial fulfillment of the requirements for
the Degree of Doctor of Philosophy
in the Division of Applied Mathematics at Brown University

Ph.D. Advisor: Stuart Geman

Providence, Rhode Island

May 2004

Abstract of "Statistical Methods for the Assessment of Temporal Structure in the Activity of the Nervous System," by Asohan Amarasingham, Ph.D., Brown University, May 2004

One of the fundamental questions in neuroscience concerns the nature of the *neural code*: how the information, or signal, that one neuron communicates to others is embodied in its biophysical processes. A basic controversy in investigations of this topic involves the rate-coding hypothesis. This is roughly the hypothesis that the information conveyed by a neuron in a sequence of action potentials (or spikes) is contained wholly in the number of spikes which occur in coarse temporal intervals on the order of several tens or hundreds of milliseconds. Alternatively, the temporal coding hypothesis holds that the precise location of spikes (for example on millisecond time scales) conveys information. We examine statistical techniques which are employed in the analysis of experimentally-obtained measurements of spike trains, with a view toward elucidating this controversy. A recurring theme which we emphasize is the assumption of repeatability which underlies estimation-based approaches to such analyses and which, while common in general in statistical methodology, is problematic for these investigations in light of the available experimental knowledge. With these issues in mind, we describe statistical models for spike trains which bear on rate versus temporal coding distinctions, and develop hypothesis tests for the purpose of relating the models to experimental data. In particular, we focus on applications to two controversial phenomena in *in vivo* cortical physiology: the variability of spike counts, and the phenomenon of synchronous spiking among simultaneously-recorded neurons. Our goal is to develop definitions and methodology to assess the existence and nature of *fine temporal structure* in the neural records using assumptions which are compatible with the theoretical goals of neuroscience.

This thesis by Asohan Amarasingham is accepted in its present form by the
Division of Applied Mathematics as satisfying the
dissertation requirement for the degree of
Doctor of Philosophy

Date..... .....
Stuart Geman, Ph.D

Recommended to the Graduate Council

Date..... .....
Elie Bienenstock, Ph.D

Date..... .....
Basilis Gidas, Ph.D

Approved by the Graduate Council

Date..... .....
Karen Newman, Ph.D
Dean of the Graduate School

The Vita of Asohan Amarasingham

Asohan Amarasingham was born in Rome, New York on June 3, 1975. He attended St. Paul's School in Concord, New Hampshire and graduated in June 1993. He entered the University of Virginia as an Echols Scholar, graduating in December 1996 with a B.A. in Mathematics and Cognitive Science. He received the Sc.M degree in Cognitive and Linguistic Sciences from Brown University in 1999, and defended this Ph.D. dissertation in the Division of Applied Mathematics on October 14, 2003.

Acknowledgments

Contents

List of Tables

List of Figures

And in the midst of this wide quietness,
A rosy sanctuary will I dress
With the wreath'd trellis of a working brain,
With buds, and bells, and stars without a name
John Keats

Yesterday, I thought I was gathering wood,
and today, when I returned,
my whole house was on fire.
I'd never been taught how scarcely
Nature lays herself bare
in her own light.

Chapter 1

Introduction: the Neural Code

One of the fundamental questions in neuroscience concerns the nature of the *neural code*: how the information, or signal, that one neuron communicates to others is embodied in its biophysical processes, generally, and, in particular, in the sequences of action potentials or *spike trains* that are transmitted through axons, and which apparently form a fundamental basis of activity in the nervous system.

The oldest model for the neural code is the *rate-coding hypothesis*, which holds, roughly, that the *signal* embedded in a neural spike train is an underlying *rate*, varying on coarse time intervals on the order of several tens or hundreds of milliseconds; the rate in these intervals is reflected in the *number* of spikes which they contain. The precise placement of the spikes is random and immaterial. This view is in many ways a dominant paradigm in neuroscientific thought and reflects, in part, a great deal of the history of behavioral neurophysiology, which consists in the main of collecting observations of covariations of (temporally-coarse) firing rates with environmental and behavioral variables, measured by a variety of methods, from single-electrode recordings in anesthetized animals, to modern multi-electrode and imaging studies in awake, behaving subjects. An alternative model for the neural code, whose implications we will explore here, often called the *temporal coding hypothesis*, holds that part of a neural spike train's signal is embedded in the fine, perhaps millisecond-scale, temporal structure of spike positions. Commonly discussed alternatives of this nature include the synchronous firing of spikes across populations of neurons (due to empirical as well as theoretical motivations), a topic we will explore in some depth in this work, and reliable temporal sequences of spike activity within and across multiple neurons (synfire chains [1]).

The distinction between the rate-coding and temporal-coding hypotheses can be somewhat difficult to make precise, certainly because the two models represent points along a continuum, a point of view we share with many authors. Further, distinguishing them experimentally (regardless of the choice of definition) presents its own difficulties. A primary difficulty is that the implications of the presence of covariation between neural phenomena and stimulus aspects for theories of neural coding is somewhat unclear. Just because firing rates in V1, for example, covary with the orientation of a bar of light does not mean that V1 neurons *code* for orientation through firing rates; the covariation could merely be *epiphenomenal*, a reflection of some more central processes. The same goes, practically without saying, for observations of covariation with higher-order temporal structures (such as, for example, reproducible, finely-structured patterns of firing among neurons).

Putting aside the question of coding, however, it is natural to suggest that the *existence* of structure such as firing rates or temporal patterns of some form is a prerequisite for the

use of that structure in coding itself. Hence a great deal of research has been devoted to the search for various forms of structure in neural spike trains, and the subsequent association of such structures with behavioral and environmental events. As Dayan & Abbott have advocated:

"The debate between rate and temporal coding dominates discussions about the neural code. Determining the temporal resolution of the neural code is clearly important, but much of the debate seems uninformative. We feel the central challenge is to identify relationships between the firing patterns of different neurons in a responding population and to understand their significance for neural coding."

The work described here is a contribution to questions that are preliminary, and perhaps prerequisite to, such investigations: that is, how are we to *identify* firing patterns among different neurons in a population from the available experimental evidence? As stated, this is essentially a question of statistical inference.

An essential phenomenon of *in vivo*, particularly cortical, spike trains, is their variability. Present an optimally-tuned stimulus to a neuron (interpreting the word 'optimal' generously) many times, and the spike response usually varies considerably. As a consequence, an important practical problem in the statistical modelling of neural spike trains is the identification of models that usefully capture the variability of the neural response. The most natural and simplest model is the *inhomogeneous Poisson process* [11]. A Poisson process with rate function $\lambda(t)$ is a stochastic point process on [0,T] with the following two properties

- $N(s,t)$, the random variable corresponding to the number of spikes between s and t has expected value
$$E[N(s,t)] = \int_s^t \lambda(u)du. \tag{1.1}$$
where $0 \leq s \leq t \leq T$.
- $N(s,t)$ and $N(u,v)$ are independent if $s \leq t \leq u \leq v$.

These two properties are indeed sufficient to uniquely characterize the process. It turns out then that the distribution of $N(s,t)$ is

$$P(N(s,t)=k) = \frac{e^{-\int_s^t \lambda(u)du}\left(\int_s^t \lambda(u)du\right)^k}{k!}, \tag{1.2}$$

i.e., $N(s,t)$ is a Poisson random variable with rate $\int_s^t \lambda(u)du$. The special case where $\lambda(t) = \lambda$ is a constant is called a *homogeneous Poisson process.*

As a consequence, the Poisson model, particularly with slowly-varying rate functions $\lambda(t)$ has a natural resonance with the rate-coding hypothesis, and it underlies, though often implicitly, a great deal of the analysis and interpretation of spike trains. The rough idea is that the event or signal that the spike train encodes in fact merely determines a slowly-varying rate, and the observed variability in the spike train is in fact due to the resulting (Poisson) noise.

Interpreting variability in this light is tricky, however, as a simple and familiar thought experiment can illustrate. Consider the following idealized model of a neuronal response during a 100 millisecond recording interval in response to, say, a bar of light in an identified receptive field. Suppose that what the neuronal response actually encodes is in fact a far richer state space than the qualities of the bar of light, and involves other contextual elements such as the attentional state of the monkey, the past history of the neuron, the states of other neurons, other aspects of the visual stimulus, etc. (One could, quite

reasonably, compile an unending list of such alternatives.) Now imagine that this larger state space contains exactly 2^{100} alternatives (the specific numbers are not important here), and furthermore that the neuron codes for them *exactly* deterministically, to within, say, 1 ms resolution: there is a one-to-one mapping between the state space and each of the 2^{100} possible spike trains. This neuron is notably a paradigm for the fine-temporal coding hypothesis. Now we present the bar several times. How does the neuron behave? Clearly, it depends on the probability distribution on the state space, or particularly, the probability distribution on the state space *conditioned* on presentation of the bar. For example, if we imagine that conditioned on presentation of the bar, all configurations of the state space are equally likely, then we would in fact observe, for the discretized spike train, a homogeneous Bernoulli process (with rate 1/2), the discrete cousin of the homogeneous Poisson process and the prototypical example of a temporally structureless stochastic process. Of course, from the perspective of physics, this has a familiar ring: it is difficult to meaningfully distinguish randomness induced by *our uncertainty* of the event (to be signalled) from the noise inherent in the (signalling) process.

The problem here, of assuming repeatability in the process underlying the neural response in order to make inferences *about* that process, is a critical issue in behavioral neurophysiology because it defeats a great deal of statistical methodology which has been crafted in general for cases where one assumes some form of repeatability in the data-generating process. On the other hand, the issue has at its heart central questions: *What theories are our data even capable of distinguishing between?*, and *What is the relevance of those theories for neuroscience?*.

A concrete example may serve to cohere these questions. One of the problems addressed in this dissertation concerns the occurrence of synchronous spiking among pairs of simultaneously recorded neurons. We observe, for example, a peak in a cross-correlogram of two neurons at time lag 0 (see Figure 3.4 or 3.6 for an example), indicative of synchronous firing, and ask: is the peak significant, or in other words, is the quantity of synchronous firing more than one would expect from chance? The notion of chance, of course, presupposes a source of variability in the spiking process, and we are thus led to the problem of *modelling* that variability. The rate-coding hypothesis supplies the classical model: the environmental context determines a rate function, and the spike times are generated by a Poisson process with the associated rate function. In this sense, assessing the significance of synchrony could be viewed as an attempt to *test* the rate hypothesis.

Generally, approaches to this analysis have hinged on the assumption of repeatability. For example, suppose we represent the response of two neurons after repeated presentation with a stimulus n times, $X^i(t)$ and $Y^i(t)$, so X^i represents the response of neuron X during trial i, and Y^i represents the response of the neuron Y during trial i. Traditional approaches to assessing synchrony in this case involve (sometimes implicitly) *estimating* the probability of firing at each time, independently for each neuron, by averaging across the n trials [39, 10, 41], under a null hypothesis that i) X and Y are independent, ii) $X(s)$ and $X(t)$ are independent for $s \neq t$, and iii) $Y(s)$ and $Y(t)$ are independent for $s \neq t$. This is, of course, a Poisson-style assumption, unless time is discrete, in which case it is the discrete-analogue, an inhomogeneous Bernoulli assumption (see below). The stimulus simply determines a rate function for each neuron, and thus, conditioned on the stimulus (or, equivalently, the rate), the neurons act independently, and thus their spikes do as well.

Brody [9, 10] suggested these analyses were problematic from the perspective of the fine temporal structure debate, by pointing out some models that were alternative to such a null hypothesis but nevertheless shared the spirit of the rate coding hypothesis: so-called latency and excitability covariations. The models he described were of the mixture-of-Poisson type, in which each neuron has a stimulus-determined rate function but the rate function is

shifted in time by a trial-varying latency parameter, and scaled by a trial-varying gain parameter. However, the trial-varying latency and gain parameters are common to both neurons (hence latency and excitability *covariations*). In such a case, the latency and gain covariations can lead to the detection of significant synchrony, using typical methods, under the standard Poisson null hypothesis. In principle, one could accommodate identifiable sources of variability like these (i.e., latency and scale), but it is of course not clear that this even is a sufficient relaxation. The neuroscientific motivation for this alternative is that slow changes in state (such as could be due to attention, for example) could "masquerade" as synchrony due to fine temporal structure. This is, then, just another version of the sort of difficulties introduced in the thought experiment above. Furthermore, it is not at all clear why such "changes in state" could be adequately accounted for by simply introducing a trial-varying scale and latency parameter.

This can be a recurring problem with the methods employed to analyze spike trains. Often a problem is identified which invalidates a typical approach to spike train analysis (such as, in this case, slow changes in state), and then a generalization of the underlying assumptions is sought which addresses the problem (such as latency and scale parameters). Then a generalized solution is chosen which often seems less to do with neuroscientific issues than statistical convenience (and convenience, in this sense, tends to take the form of familiarity). To provide one of many possible examples of this, in his textbook [29], after defining the notion of a stationary process, and then suggesting its shortcomings as a neural model, Koch writes: "Since true stationarity is very difficult to verify in practice, it is common to use the less stringent concept of a *wide-sense* or *weakly stationary* process..." (p. 352). It is certainly easy to see the problems with using stationary process theory for spike trains, but it is not so easy to see how wide-sense stationary processes (in which not the entire process, but just the mean and autocorrelation are time-invariant) are much more relevant, either. Their main virtue appears to be that they are a generalization that is classically well-understood.

Allowing for abitrarily general models of the sort suggested by these difficulties, on the other hand, leads to identifiability issues. For example, let us model a spike train as a mixture of inhomogeneous Bernoulli processes. An *inhomogeneous* Bernoulli process is the discrete-time analogue of the inhomogeneous Poisson process. A discrete-time stochastic process $X = (X_1, X_2, ..., X_T)$ on $1 \leq i \leq T$ is an inhomogeneous Bernoulli process with rate function $r_1, r_2, ..., r_T$ if

$$P(X_1 = x_1, X_2 = x_2, ..., X_T = x_T) = \prod_{i=1}^{T} r_i^{x_i}(1 - r_i)^{(1-x_i)} \quad \forall (x_1, x_2, ..., x_T) \in \{0,1\}^T. \tag{1.3}$$

The properties of independence (of the X_i's) and the rate function make the analogy to the (continuous-time) Poisson process, and indeed one can obtain the Poisson process by taking suitable limits of finer discretizations of Bernoulli processes. Despite artifacts introduced by discretizations (which are also present in data collection, nevertheless), the Bernoulli process is a convenient model to illustrate more cleanly the essential identifiability problem.

Suppose, now, in the spirit of the above thought experiment, one wishes to test the hypothesis of a *mixture* of inhomogeneous Bernoulli processes with respect to synchrony. Immediately we encounter the following dilemma: *every discrete-time stochastic point process is a mixture of Bernoulli processes.*

To see this, suppose that we have an arbitrary discrete stochastic point process, $Y = (Y_1, ..., Y_T)$, on a time interval of length T. This merely means there is an assignment of probabilities

$$P(Y = \mathbf{y}) = p_{\mathbf{y}} \quad \forall \mathbf{y} \in \{0,1\}^T, \tag{1.4}$$

which additionally satisfies

$$\sum_{\mathbf{y}\in\{0,1\}^T} p_{\mathbf{y}} = 1. \tag{1.5}$$

Each outcome $\mathbf{y}$ of the process in the sample space $\{0,1\}^T$ however is itself a Bernoulli process with rate function $(r_1, ..., r_T) = (y_1, ..., y_T)$: this is just the tautological observation that one can represent any deterministic discrete time point process with a Bernoulli process, simply by restricting the rate function to the exact values of 0 and 1. Thus to each outcome $\mathbf{y}$ we assign a unique rate function $r_{\mathbf{y}} = (y_1, y_2, ..., y_n)$, and we assign to each $r_{\mathbf{y}}$ the mixture probability $p_{\mathbf{y}}$. The resulting mixture process Z, is a mixture of inhomogeneous Bernoulli processes, which has probabilities

$$P(Z = \mathbf{y}) = \sum_{z\in\{0,1\}^T} P(Z = \mathbf{y}|r_{\mathbf{z}})p_{\mathbf{z}} = p_{\mathbf{y}}, \tag{1.6}$$

i.e., Z has the distribution of the original process Y.

This has the following implication: for the discretized spike train, if you give up the repeatability assumption, then modeling the spike trains with Bernoulli processes is tautologically valid. The same observation is even more general, and can include arbitrary sequences of processes which are not independent across trials, for the same reasons. In other words, the model is maximally arbitrary, and if we were to invoke it as a null hypothesis, there would be *no* alternatives, and correspondingly no role for a hypothesis test to play.

These observations suggest that the difficulty of assuming traditional forms of repeatability in the spiking process makes it hard to infer facts about the nature of the spiking process from general observations of *variability* in neural spike trains. The converse is not true: observations of *regularity* in spiking processes, on the other hand, are not necessarily subject to our uncertainties about repeatability, inescapable as they might be. Forms of regularity in spike trains can reveal temporal structure in the processes underlying nervous system activity, and thus offer the promise of shedding light on the rate versus temporal-coding hypotheses. This fact forms the conceptual basis for this dissertation. From this perspective we will examine two controversial phenomena in *in vivo* cortical physiology: the variability of spike counts, and the phenomenon of synchronous spiking among simultaneously-recorded neurons. Our goal is to develop definitions and methodology to assess the existence and nature of *fine temporal structure* in the neural records using assumptions which are compatible with the theoretical goals of neuroscience.

Chapter 2

Spike Count Variability and the Poisson Hypothesis

2.1 Introduction

A statistic commonly employed to assay variability in spike trains is the empirical variance-mean ratio of the spike counts (the Fano factor) across trials [52, 45, 54]. For example, the spike counts of an inhomogeneous Poisson process, a frequently used model, are distributed as a discrete Poisson random variable, for which the mean and variance are identical. In general the spike counts of *in vivo* cortical spike trains (in contrast perhaps to subcortical structures) have been found to be as variable as Poisson processes and perhaps even more so [45, 29]. As Shadlen and Newsome have written: "When an identical visual stimulus is presented for several repetitions, the variance of the neural spike count has been found to exceed the mean spike count by a factor of 1-1.5 wherever it has been measured."

The variability of spike trains bears on theories of neural coding. A range of hypotheses have been offered. At one extreme, the existence and precise location of every spike is significant. At another extreme, spikes are the events of a slow-varying inhomogeneous Poisson process (the "Poisson hypothesis"). Fine-temporal coding would be more likely to yield highly regular spike counts from repeated trials, whereas the Poisson hypothesis, by virtue of the Fano factor, limits the degree of regularity across trials.

Inferring a lack of precision in the neural code from observations of variability is tricky. One line of reasoning is that the existence of variability under identical conditions, when the neuron is presumably signaling the same event, reflects noise (e.g. Poisson noise) in the signaling process itself. However, it is certainly plausible that a significant source of variability is the experimenter's own uncertainty about hidden, contextual variables which the neuron is encoding (and which may be, for example, internal to the brain), such as attention, or the states of other neurons. Furthermore, the variability of overt behaviors that are difficult, but not impossible to measure, like precise eye position, have been shown to contribute to at least some of the variability commonly reported in cortical responses [24]. As Barlow wrote about neural responses in 1972: "their apparently erratic behavior was caused by our ignorance, not the neuron's incompetence."

Regularity, on the other hand, cannot be explained away so easily, and in this light *evidence* of finer temporal structure, particularly deeper in cortex, is especially intriguing. Recently, Muller et al. [37] have observed, from recordings of V1 cells in primates presented with sinusoidal gratings, that near stimulus onset the empirical variance-mean ratio is strikingly smaller than unity (in contradiction to the Poisson hypothesis). Our purpose in this chapter is twofold. First, we present evidence consistent with that of Muller et al., gathered here in cells from anterior IT of primates presented with more complex visual stimuli, of

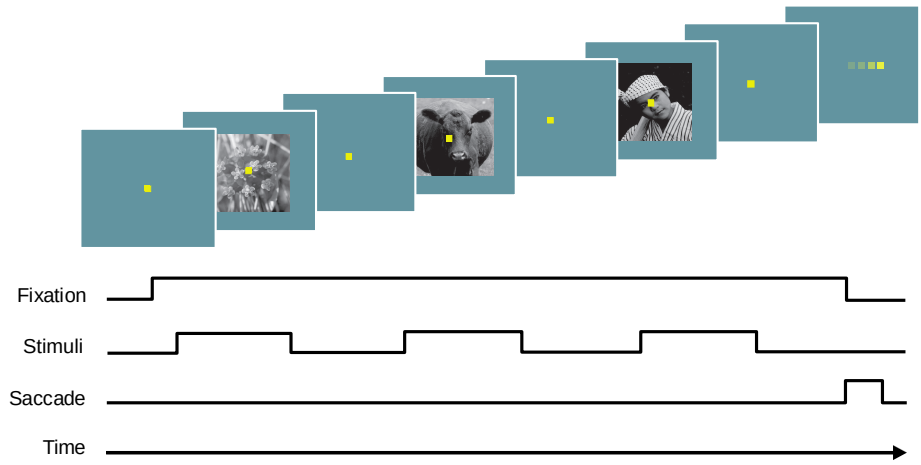

Figure 2.1: Basic fixation task performed by the monkeys. In a single observation period, between three and five individual visual stimuli were flashed on and off as the monkey fixated a spot in front of the images. At the end of the trial, the monkey received juice reward for reacquiring the spot as it jumped from the center of the screen to one of four randomly selected peripheral locations.

low variance-mean ratios particularly near stimulus onset. Second, we derive a simple and exact test of the Poisson hypothesis with respect to the variability of spike counts across trials. Using this test, we are able to reject the Poisson hypothesis, with a notably small amount of data, predominantly near stimulus onset. We are able to argue, furthermore, that these results are not due to the effects of the refractory period *per se*, but rather reflect regularity in the neural response.

2.2 Methods

2.2.1 Subjects and Materials

We use data whose collection has been described previously [46, 47]. In brief, following initial behavioral training, two rhesus monkeys underwent aseptic surgery for the placement of head restraint and a scleral search coil. All surgical procedures were carried out in accordance with the NRC *Guide for the Care and Use of Laboratory Animals*. Following surgery, the monkeys were trained to fixate a small yellow spot (0.25 degrees) appearing in the center of a CRT monitor. After acquiring the spot, between three and five visual images, 4 deg on a side, were flashed behind the spot for either 800 or 1100 ms. Interstimulus intervals (during which only the spot was visible) were set to the same duration as the stimulus presentation duration (Figure 2.1). Visual stimuli were selected from a set of commerically available stock photographs of animals, natural scenes, and man made objects (Corel Corp).

During stimulus presentation, the monkeys were required to maintain fixation within a virtual square region 2 deg on a side. During data collection, the digitized eye position was stored to disk every 5 ms (200hz). In addition to keeping their gaze directed within the virtual window, the monkeys learned to follow the spot as it jumped from the center of the display to a new position. Upon refixation of the spot, they were rewarded with a drop of apple juice.

2.2.2 Single Cell Recordings

Once the animals were trained in the fixation task, a ball and socket chamber housing an 18 gauge guide tube was placed directly above the anterior temporal lobes (AP: +18, ML: +19). Single unit recordings were made by lowering glass coated platinum-iridium electrodes by microdrive (Kopf Model 650) into the lower bank of the superior temporal sulcus (STS) and the lateral convexity of the inferior temporal gyrus, just posterior to the anterior middle temporal sulcus (AMTS). The neural signal was amplified using a Bak A-1 amplifier with remote head stage, and the conditioned signal was fed into a time-amplitude window discriminator. Individual cells were isolated while the animals performed the fixation task. Each cell was tested with at least eight stimuli, but often with many more (up to 80). Note that a significant effort was made to subselect visual stimuli from the relatively large test set that effectively elicited robust neuronal responses, as judged by online rasters.

2.2.3 A Statistical Test

Consider a series of spike trains from a single neuron obtained from n separate trials (each involving, for example, presentation of the same stimulus):

$\{t_1^1, t_2^1, t_3^1, ..., t_{m_1}^1\}, \{t_1^2, t_2^2, t_3^2, ..., t_{m_2}^2\}, ..., \{t_1^n, t_2^n, ..., t_{m_n}^n\},$

where there are m_j spikes in trial j, and t_i^j is the time of occurrence of the i'th spike in the j'th trial, relative to a stimulus onset at time 0.

A *homogeneous* Poisson process of rate λ is a statistical model of the spike train which is characterized by two conditions: i) the events in non-overlapping time intervals are independent, and ii) the expected number of spikes in a time interval of length T is λT. An *inhomogeneous* Poisson process is the nonstationary analogue of this: events in non-overlapping time intervals are independent, but the expected number of spikes is governed by a *time-varying* rate $\lambda(t)$.

Our null hypothesis (H_0) is that $m_1, m_2, ..., m_n$ are independent Poisson random variables. This contains, as a special case, the hypothesis that the recordings come from independent and possibly inhomogeneous Poisson processes.

Let $\hat{\mu}$ be the mean number of counts,

$$\hat{\mu} := \frac{1}{n} \sum_{i=1}^{n} m_i$$

In section 2.3.1 we derive the following exact statistical test, the Poisson Variability Test, for H_0: Reject H_0 at level α if

$$\sum_{i=1}^{n} m_i^2 \leq f(n, \alpha, \hat{\mu})$$

We provide a simple algorithm for computing f, as well as links to tables with pre-computed values.

The test has power towards alternative hypotheses that would predict more repeatable observations of spike counts than the Poisson hypothesis.

2.2.4 Grouping the Tests

To group the tests, we would like to assess the significance of the *number* of rejections of the Poisson variability test in a given epoch across a collection of cell-stimulus pairs, under the null hypothesis that every cell-stimulus pair is independent. One idea is to apply a binomial test to the number of rejections across the population of cell-stimulus pairs,

assuming that the likelihood of rejection is $\alpha = 5\%$, and that the tests are independent. This is conservative, however, especially because one cause for a *failure* to reject the null hypothesis in a single cell-stimulus pair might simply be a lack of data, either in the form of a paucity of trials or a paucity of spikes. Indeed, for some values of n, the number of trials, and of N, the total number of spikes (i.e., $\sum_{i=1}^{n} m_i$), it is impossible to reject the null hypothesis for *any* configuration of the data. This is the case if $P(\sum_{i=1}^{n} X_i^2 \leq k_*) > \alpha$ where

$$k_* := \min_{\substack{m_1,...,m_n \in \mathcal{N} \\ \sum_{i=1}^{n} m_i = N}} \sum_{i=1}^{n} m_i^2, \tag{2.1}$$

where $\mathcal{N}$ represents the non-negative integers and $X_1, X_2, ..., X_n \sim \mathcal{M}(N; 1/n, 1/n, ..., 1/n)$.

The effect is more general: in other cases it is not necessarily impossible to reject the null, but simply less likely than 5%. One can view this formally as an effect of the discreteness of the multinomial statistics, which correspondingly becomes more pronounced as the total number of spikes gets smaller. We can take account of this, by conditioning further, on the number of trials n and the total number of spikes $n\hat{\mu}$ for each cell-stimuli pair, and then grouping the results, as follows.

Let

$$r^*(n, \alpha, \hat{\mu}) := P\left(\sum_{i=1}^{n} X_i^2 \leq f(n, \alpha, \hat{\mu})\right) \quad \text{where } X_1, ..., X_n \sim \mathcal{M}\left\{n\hat{\mu}; 1/n, 1/n, ..., 1/n\right\}. \tag{2.2}$$

Then for a given cell-stimulus pair k, if Y is the rejection event,

$$Y := \{\text{reject} H_0\} := \left\{\sum_{i=1}^{n} m_i^2 \leq f(n, \alpha, \hat{\mu})\right\}, \tag{2.3}$$

then by the definition of f we have

$$P(\text{reject } H_0 | n, \hat{\mu}) = P(Y | n, \hat{\mu}) \leq r^*(n, \alpha, \hat{\mu}) \leq \alpha. \tag{2.4}$$

Analogously for a collection of N cell-stimulus pairs, if we denote Y_i the rejection event for cell-stimulus pair i, $\hat{\mu}_i$ the average number of spikes per trial, n_i the number of trials, then we have

$$P(Y_i | n_i, \hat{\mu}_i) \leq r^*(n_i, \alpha, \hat{\mu}_i) \leq \alpha \quad \forall i. \tag{2.5}$$

Then denoting $\vec{n} = (n_1, ..., n_N)$, we define

$$g(\vec{n}, \hat{\mu}_1, ..., \hat{\mu}_N) := \min\left\{t : P\left(\sum_{i=1}^{n} Z_i \geq t\right) \leq \beta\right\} \tag{2.6}$$

where $Z_1, Z_2, ..., Z_N$ are independent $\{0, 1\}$ (Bernoulli) random variables with $Pr\{Z_i = 1\} = r^*(n_i, \alpha, \hat{\mu}_i)$. The number of rejections under the Poisson variability test is $\sum_{i=1}^{n} Y_i$ and the critical region $\{\sum_{i=1}^{N} Y_i \geq g(\vec{n}, \hat{\mu}_1, ..., \hat{\mu}_N)\}$ tests the hypothesis that the observed number of rejections occured by chance, at level β. Given a particular value for β, the critical value $g(\vec{n}, \hat{\mu}_1, ..., \hat{\mu}_N)\}$ can be computed efficiently by recursion (see the Appendix).

2.3 Results

2.3.1 Derivation of the Poisson Variability Test

Our null hypothesis (H_0) is that $m_1, m_2, ..., m_n$ are independent Poisson random variables. Let us designate the (unknown) rates as $\lambda_1, \lambda_2, ..., \lambda_n$. Define the empirical mean and variance statistics as usual:

$$\hat{\mu} := \frac{1}{n}\sum_{i=1}^{n} m_i \qquad \hat{\sigma}^2 := \frac{1}{n-1}\sum_{i=1}^{n}(m_i - \hat{\mu})^2$$

We seek a hypothesis test under our null which has power when $\hat{\sigma}^2/\hat{\mu}$ is small, or more to the point, such that $\hat{\sigma}^2$ is small, given $\hat{\mu}$. It is equivalent to use $\sum_{i=1}^{n} m_i^2$ in place of $\hat{\sigma}^2$, since they preserve the same order when $\hat{\mu}$ is fixed.

One way to proceed is to form a partition of the sample space based on $\hat{\mu}$. We seek $f(n, \alpha, \hat{\mu})$ such that

$$P\left(\sum_{i=1}^{n} m_i^2 \le f(n,\alpha,\hat{\mu}) \,\middle|\, \hat{\mu}\right) \le \alpha \quad \forall\hat{\mu} \quad \forall P \in H_0. \tag{2.7}$$

Then the event $\{\sum_{i=1}^{n} m_i^2 \le f(n, \alpha, \hat{\mu})\}$ will be an α-level hypothesis test, since:

$$\begin{aligned} P\left(\sum_{i=1}^{n} m_i^2 \le f(n,\alpha,\hat{\mu})\right) &= \sum_{\hat{\mu}} P\left(\sum_{i=1}^{n} m_i^2 \le f(n,\alpha,\hat{\mu}) \,\middle|\, \hat{\mu}\right) P(\hat{\mu}) \le \sum_{\hat{\mu}} \alpha P(\hat{\mu}) \\ &= \alpha \quad \forall P \in H_0 \end{aligned} \tag{2.8}$$

To derive $f(n, \alpha, \hat{\mu})$, it is straightforward to apply the following proposition. We defer the proof and other technical details to the appendix.

Proposition 1. *If $m_1, m_2, ..., m_n$ are independent Poisson random variables with rates $\lambda_1, \lambda_2, ..., \lambda_n$, respectively, then for all r, and for all $\hat{\mu}$, we have*

$$\max_{\lambda_1,\lambda_2,...,\lambda_n} P\left(\sum_{i=1}^{n} m_i^2 \le r \,\middle|\, \hat{\mu}\right) = P\left(\sum_{i=1}^{n} X_i^2 \le r\right)$$

where $X_1, X_2, ..., X_n$ are distributed multinomially with parameters $\{n\hat{\mu};\, 1/n, 1/n, ..., 1/n\}$.

A multinomial distribution generalizes the binomial distribution to a many-sided die, and has the form

$$P(X_1 = m_1, X_2 = m_2, ..., X_n = m_n) = \begin{cases} \binom{N}{m_1\, m_2 \ldots m_n} \prod_{i=1}^{n} p_i^{m_i} & \text{if } \sum_{i=1}^{n} m_i = N \\ 0 & \text{otherwise,} \end{cases}$$

where $p_1, p_2, ..., p_n$ satisfy $p_i \ge 0$ and $\sum_{i=1}^{n} p_i = 1$.

Thus we have

$$\begin{aligned} f(n,\alpha,\hat{\mu}) \quad := \quad & \max\left\{k : P\left(\sum_{i=1}^{n} X_i^2 \le k\right) \le \alpha\right\} \\ & \text{where } X_1, ..., X_n \sim \mathcal{M}(n\hat{\mu}; 1/n, 1/n, ..., 1/n) \end{aligned}$$

There are at least two practical ways to compute the multinomial probability associated with f. It can be computed exactly, as we do in the experiments reported below, by dynamic

programming [6] (see the appendix), provided that $\sum_{i=1}^{n} m_i$ and $\sum_{i=1}^{n} m_i^2$ are not too large. Or it can be computed approximately, by Monte Carlo methods (see the appendix). A table of values can be found in the appendix.

It is worth pointing out that the Poisson variability test covers a more general null hypothesis: $m_1, m_2, ..., m_n$ are *conditionally* independent and Poisson distributed, given the Poisson rates $\lambda_1, \lambda_2, ..., \lambda_n$. Thus the null hypothesis includes: a) trial-varying inhomogeneous Poisson processes (as stated above), but also b) Cox processes, in which a particular inhomogeneous rate function is chosen, randomly, at each epoch for each neuron from an ensemble of possible rate functions.

2.3.2 Analysis of Cell Recordings

Our interest in the trial to trial variability of inferotemporal cells emerged as a consequence of the simple impression that, given an appropriate stimulus, neuronal responses at least *seemed* reliable, especially near the onset of the stimulus (Figure 2.2). Given the prevailing view that cells throughout cortex are invariably irregular, we set out to more rigorously explore this issue.

We divided the neural response into epochs, consisting of disjoint equal-length time intervals, beginning 100 ms after the onset of presentation of a stimulus. We used intervals of length 100 ms and 50 ms in the experiments described below. The pairing of a cell and a stimulus presented repeatedly to the cell we labelled a *cell-stimulus pair*. For a fixed epoch, associated with each particular cell-stimulus pair are the data points $m_1, m_2, ..., m_n$ of spike counts for each of the n presentations of the stimulus to that cell during the epoch. For each cell-stimulus pair, $m_1, m_2, ..., m_n$, (in particular the statistics $\sum_{i=1}^{n} m_i$ and $\sum_{i=1}^{n} m_i^2$), determine the p-value, the minimal value of α that the Poisson variability test will reject, which we compute as per above.

We analyzed 328 cell-stimulus pairs, drawn from recordings from a total of 27 cells (18 from the first monkey and 9 from the second). These cells were selected on the basis of their clear visual response to at least one visual stimulus and based on the quality of the single unit isolation, which was as high as possible to prevent the uncertainty introduced by a noisy signal from contaminating our data. The number of presentations n varied with each cell-stimulus pair, ranging from $n = 2$ to $n = 14$, with a median value of 7.

Using 100 ms epochs, 47 of the 328 pairs were rejected by the Poisson variability test at a significance level of 95% in the 100 to 200 ms period following stimulus onset. In contrast, the number of rejected cell-stimulus pairs decreased rapidly and progressively at subsequent epochs, further away from onset: 19 rejected pairs between 200 to 300 ms, 11 rejected pairs between 300 and 400 ms, and fewer further on. Similar trends are evident for 50 ms epoch partitions, as illustrated in Figure 2.3.

To assess the meaning of the number of rejections, we calculated the significance of the number of rejections of the Poisson variability test under the assumption that the results of the individual cell-stimulus pair tests are independent (as discussed in Methods). Consistently, the significance of the number of rejections is very high for the 100 ms to 200 ms period (significance 1-10^{-33}), and decreases rapidly at subsequently epochs. These and similar trends for the 50 ms epoch partition are also illustrated in 2.3.

2.4 Discussion

2.4.1 Small Samples

The statistical significance of the results is somewhat surprising in light of the small sample sizes, ranging from 2 to 14 per cell-stimulus pair. Indeed, some of the observations are

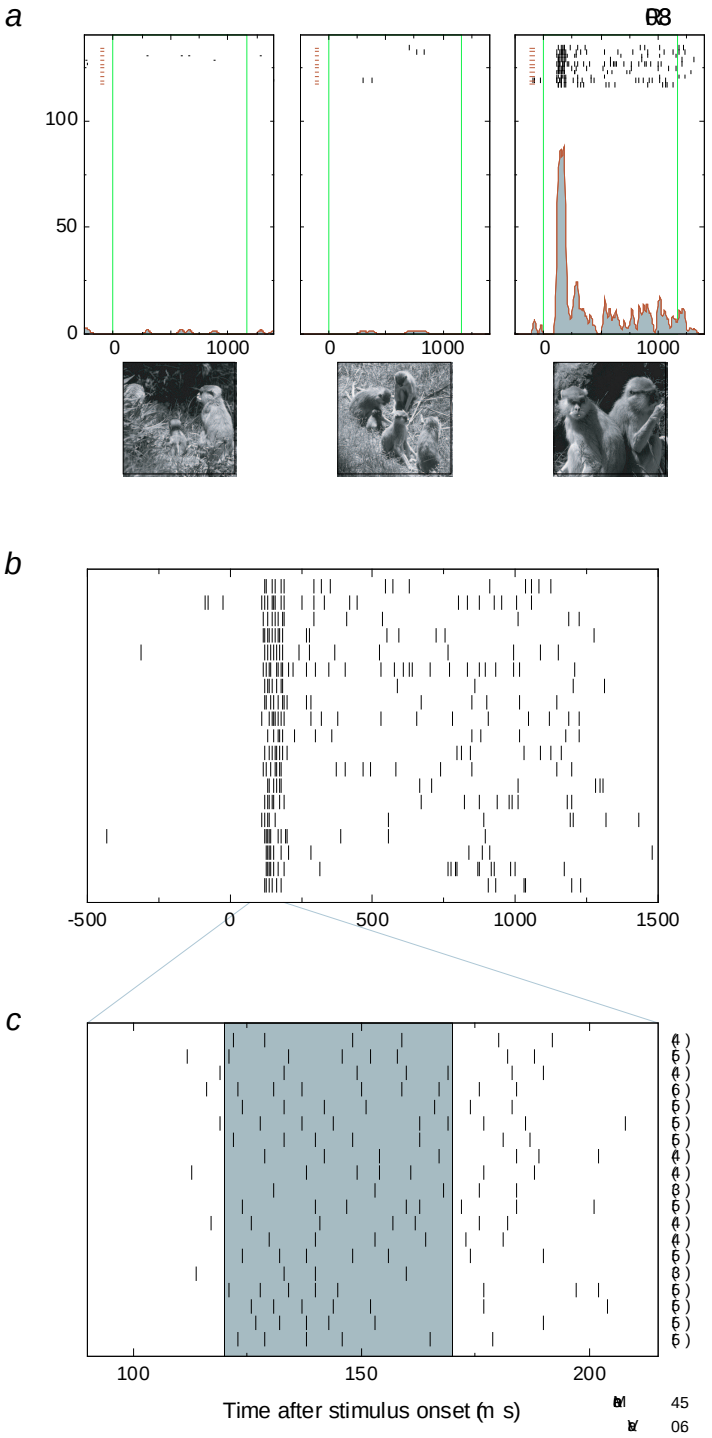

Figure 2.2: Reliable response from a single cell, indicating that particular stimuli can elicit highly regular trial-to-trial spike discharges from a neuron. **A.** Response to three of the stimuli used during testing of this cell. Each plot includes the spike rasters aligned to the onset of the stimulus (time indicated by the left vertical green line). At the bottom of each plot is an estimate of the instantaneous firing rate. **B.** Enlarged view of the aligned spike times for all responses to the most effective stimulus show in **A** (taken from two separate blocks). **C.** A temporally expanded view of the spiking activity near stimulus onset, with the number of spikes occuring in the 50 ms shaded area indicated to the right of each trial. Note that the mean of the counts far exceeds the variance of the counts (ratio: 7.6 times).

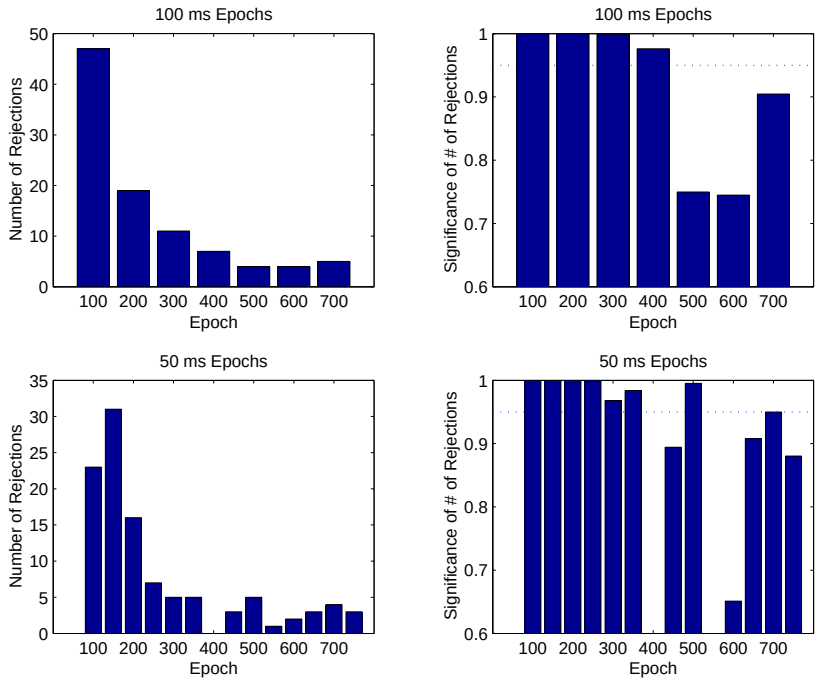

Figure 2.3: Summary of the rejections from the Poisson variability test across a total of 328 cell-stimulus pairs for each epoch. The 800 milliseconds following stimulus onset were partitioned into disjoint intervals of equal length called epochs. Epochs are labelled along the x-axis by the starting time of the interval with which they are associated. The first row summarizes the results from the 100 ms epoch partition, and the second row from the 50 ms epoch partition. The first column illustrates the total number of rejections vs. epoch. The second column graphs the epoch-by-epoch significance of the number of rejections of the Poisson hypothesis, towards excess regularity, under the assumption that the results of the individual cell-stimulus pair tests are independent (see Methods, Section 2.2.4). For the first three 100 ms epochs (100 ms, 200 ms, and 300 ms; first 3 bars, 1st row, 2nd column) the significances are $1 - 10^{-33}, 1 - 10^{-9}$, and $1 - 10^{-3}$, respectively. For the first four 50 ms epochs (100 ms, 150 ms, 200 ms, and 250 ms; first 4 bars, 2nd row, 2nd column) the significances are $1 - 10^{-15}, 1 - 10^{-23}, 1 - 10^{-10}$, and .9992.

so small that rejection of H_0 at 95% is mathematically impossible (see the discussion in Section 2.2.4). For example, even if *every* 100 ms post-stimulus epoch had *exactly* 2 spikes (corresponding to the *most* reliable outcome at 20 Hz), the Poisson variability test would only achieve 95% significance with 4 or more trials. In fact, to take one example, in the analysis of the 100 ms to 200 ms epoch, for 50% of the cell-stimulus pairs it was impossible to reject the null hypothesis on the basis of the total number of trials and spikes alone.

2.4.2 Refractory Period

One of the most natural biophysical objections to Poisson models of neural activity is the refractory period [7], which at the least introduces short-term dependencies in spike train structure. A statistical test which is sensitive to reliability in the spike counts may merely reflect such local structures, particularly among cell-stimuli pairs with relatively high firing rates, since the effect of the refractory period on reliability will increase with the firing rate. One way to examine this interpretation is to compare the results of the variability test across epochs, among cell-stimuli pairs that have the same mean spike counts. If the observed (super-Poisson) reliability is merely due to the refractory period or another local effect imposed on an "otherwise" Poisson process (a Poisson-refractory process, perhaps of the type proposed by Kass and Ventura [28]), then all other things being equal, one would expect the effect to be independent of the time of occurrence relative to stimulus onset. Figure 2.4 provides a scatter plot of firing rate versus p-value, for cell-stimulus pairs separated by epoch. There is evidently a systematic effect of epoch on spike count reliability, which is independent of firing rate, and which cannot be explained by refractory period alone.

2.4.3 Effect of eye position

In an attempt to more carefully characterize the variable response observed in primary visual cortical cells, Gur et al. (1997) found that the use of moving stimuli coupled with precise control of stimulus position on the retina lowered the variance to mean ratio compared to previously reported values. In the present experiment, we did not reposition the stimulus in real time, but we did record eye position throughout each trial. Figure 2.5 shows one measure of the variability due to eye movements during our task, by plotting the standard deviation of the monkeys' horizontal and vertical eye position as a function of time. In general, the variability at each time point was quite low (less than 0.25deg), but there is a notable increase in this variability starting approximately 300 ms after the stimulus appeared. Even during periods of controlled fixation, eye movements are not completely abolished, and we cannot, therefore, rule out the possibility that increased positional uncertainty contributed to higher variability in later temporal epochs.

2.4.4 Behavioral relevance

The present data were collected under conditions where the recorded neural signal was in no obvious way related to the animals' behavior. However, numerous previous studies have demonstrated the importance of inferotemporal cortex in successful execution of complex pattern recognition tasks [32]. The speed with which monkeys can perform these tasks is of special interest, because they give some hints as to when, and for how long, neuronal signals emanating from IT cortex may be integrated to drive recognition behaviors. Using stimuli similar to those employed in this study, for example, Vogels [57] found that monkeys could accurately categorize tree and non-tree stimuli in under 300 ms. More recently, Sheinberg

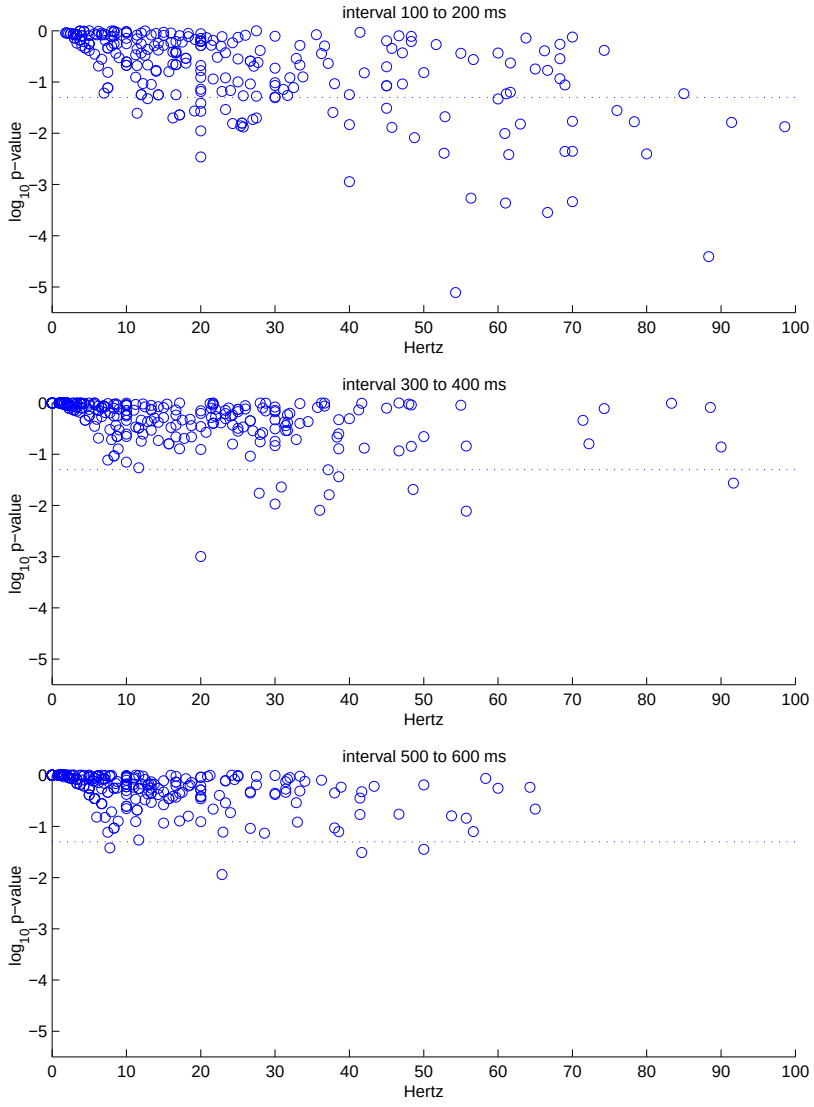

Figure 2.4: A scatter plot of $\log_{10} p$-value vs. firing rate in hertz, across all cell-stimulus pairs, separated in 100 millisecond epochs. The dotted line is the line of 95% significance.

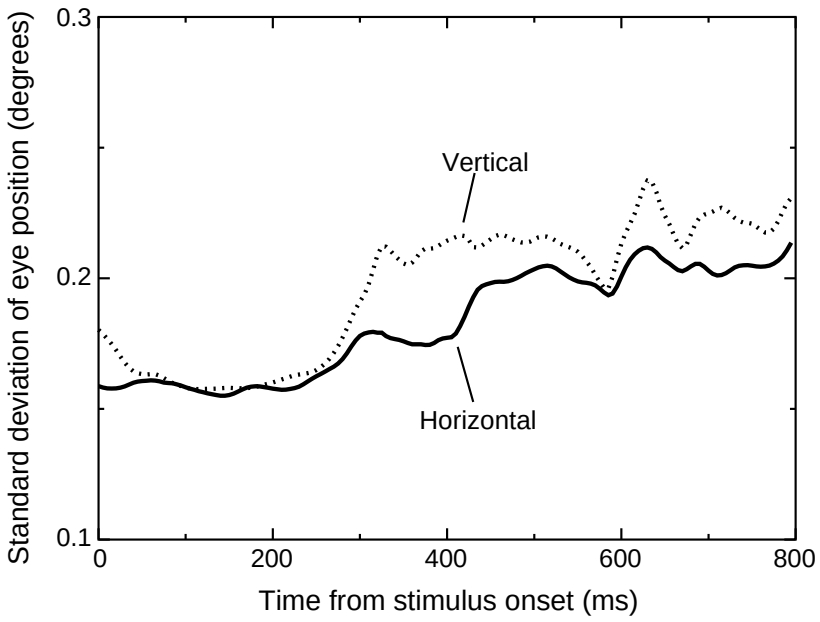

Figure 2.5: Eye position variability. The average standard deviation of both horizontal and vertical eye position for the 800ms period following stimulus onset indicates that although there was relatively little positional uncertainty, the variability did increase later in the trial, and this may have contributed to variability later in the neuronal response.

et al. [48] found that highly similar complex visual images could be discriminated by button response in approximately 425 ms, but that clear evidence for stimulus identity was evident in the monkeys oculomotor behavior approximately 200 ms following stimulus onset. This lower bound of 200 ms is intriguing because it leaves little time for extensive averaging of highly variable signals. Indeed, the average onset latency of selective neuronal responses in IT cortex falls somewhere between 100 ms and 140 ms [42, 58, 47]. Thus, the temporal epoch between initial IT cell activation and discriminatory motor response is quite short, on the order or 100 ms or less. Interestingly, it is precisely during this epoch where we find that individual neurons are most likely to respond reliably to particular visual stimuli.

2.4.5 Models

Identifying a model which usefully captures the variability of a cortical spike train is a key problem in the statistical modelling of neural data. A commonly-invoked model is the slow-varying inhomogeneous Poisson process, which follows in a natural way from a rate-coding viewpoint [45]: the *signal* embedded in a neural spike train is an underlying *rate*, varying on coarse time intervals on the order of several tens or hundreds of milliseconds, and hence the precise placement of spikes is random and irrelevant. We are ignorant of what variables a recorded neuron codes for, however, and it is reasonable that they include internal variables such as attention, the states of other neurons, et cetera which certainly might vary across recording trials, even though the presented stimulus does not.

A Cox process generalizes the Poisson process by making the rate function itself *random*, which seems a reasonable way to incorporate a stochastic and trial-varying dependence of the spike train on contextual variables. Because of this generality a Cox process, equipped perhaps in addition with a refractory period, goes a long way towards providing an arguably universal model of the spike train.

Our null hypothesis H_0 contains Cox processes, and we reject this hypothesis *in favor of reliability*, near stimulus onset. Furthermore, the rejection cannot *only* be due to the (time-independent) refractory period, since the rejection occurs near stimulus onset and not further away even among populations of cell-stimulus pairs with similar average firing rates.

From a pedagogical point of view, it is important to reiterate that the rejection of a null hypothesis implies only that, the rejection of a hypothesis: it in no way identifies an alternative. However, the form of the statistical test, and in particular the distribution of its power among the space of alternative hypotheses, may shed light on the data-generating process. In our case, it is important that the Poisson variability test is targeted towards *reliability* in the spike counts, and in that light it may be worth bearing the following two alternatives in mind. In the first, the spike train is generated by a random rate function, and at the least the *reliability* of the rate function is greater near stimulus onset than later. In the second, the spike train near stimulus onset may be a finely structured temporal process with high precision on many or most of the spikes. Significantly, the distinction between these models can begin to blur.

In any case, there is strong statistical evidence that neuronal responses near the onset of a stimulus cannot in general be well modeled by a simple Poisson process, even allowing for a trial-dependent inhomogeneous rate function.

2.5 Appendix

2.5.1 Proof of Proposition 1

We say the nonnegative integer-valued random variables $X_1, X_2, ..., X_n$ are multinomially distributed with parameters $\{N; p_1, p_2, ..., p_n\}$ if

$$P(X_1 = m_1, X_2 = m_2, ..., X_n = m_n) = \begin{cases} \binom{N}{m_1\, m_2 \ldots m_n} \prod_{i=1}^n p_i^{m_i} & \text{if } \sum_{i=1}^n m_i = N \\ 0 & \text{otherwise,} \end{cases}$$

where $(p_1, p_2, ..., p_n)$ satisfy $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$. In this case we write $X_1, X_2, ..., X_n \sim \mathcal{M}(N; p_1, p_2, ..., p_n)$.

Our goal is

Proposition 1. *If $m_1, m_2, ..., m_n$ are independent Poisson random variables with rates $\lambda_1, \lambda_2, ..., \lambda_n$, respectively, then for all r, and for all $\hat{\mu}$, we have*

$$\max_{\lambda_1, \lambda_2, ..., \lambda_n} P\left(\sum_{i=1}^n m_i^2 \leq r \,\middle|\, \hat{\mu} \right) = P\left(\sum_{i=1}^n X_i^2 \leq r \right)$$

where $X_1, X_2, ..., X_n$ are distributed multinomially with parameters $\{n\hat{\mu}; 1/n, 1/n, ..., 1/n\}$.

We first prove some lemmas about the multinomial distribution which will be useful.

Lemma 2.5.1. *If*

$$X_1, X_2, ..., X_n \sim \mathcal{M}(N; p_1, p_2, ..., p_n),$$

and

$$Y_1, Y_2, ..., Y_k \sim \mathcal{M}\left(N - \sum_{i=k+1}^n m_i \,;\, \frac{p_1}{\sum_{i=1}^k p_i}, \frac{p_2}{\sum_{i=1}^k p_i}, ..., \frac{p_k}{\sum_{i=1}^k p_i} \right),$$

with $k < n$ then

$$P(X_1 = m_1, X_2 = m_2, ..., X_k = m_k | X_{k+1} = m_{k+1}, ..., X_n = m_n) = P(Y_1 = m_1, Y_2 = m_2, ..., Y_k = m_k).$$

Proof.

$$P(X_1 = m_1, X_2 = m_2, ..., X_k = m_k | X_{k+1} = m_{k+1}, ..., X_n = m_n)$$

$$= \frac{P(X_1 = m_1, X_2 = m_2, ..., X_n = m_n)}{P(X_{k+1} = m_{k+1}, ..., X_n = m_n)}$$

$$= \frac{\binom{N}{m_1\, m_2\, ...\, m_n} \prod_{i=1}^{n} p_i^{m_i}}{\sum\limits_{\substack{x_1,...,x_k \\ \sum_{i=1}^{k} x_i = N - \sum_{i=k+1}^{n} m_i}} \binom{N}{x_1 ... x_k\, m_{k+1} ... m_n} \prod_{i=1}^{k} p_i^{x_i} \prod_{i=k+1}^{n} p_i^{m_i}}$$

$$= \frac{\frac{1}{\prod_{i=1}^{k} m_i!} \prod_{i=1}^{k} p_i^{m_i}}{\frac{1}{(N - \sum_{i=k+1}^{n} m_i)!} \left(\sum_{i=1}^{k} p_i \right)^{N - \sum_{i=k+1}^{n} m_i}}$$

by an application of the multinomial formula,

$$= \binom{N - \sum_{i=k+1}^{n} m_i}{m_1 ... m_k} \prod_{i=1}^{k} \left(\frac{p_i}{\sum_{i=1}^{k} p_i} \right)^{m_i}$$

$$= P(Y_1 = m_1, Y_2 = m_2, ..., Y_k = m_k).$$

Corollary 2.5.2. *If $X_1, X_2, ..., X_n \sim \mathcal{M}(N; p_1, p_2, ..., p_n)$,*

$$P(X_3, ..., X_n | X_1 = j, X_2 = k - j) = P(X_3, ..., X_n | X_1 + X_2 = k) \quad \forall\, k,\, j \leq k.$$

Proof. This follows from Lemma 2.5.1, since $X_3, ..., X_n$ depends on X_1 and X_2 only through the sum $X_1 + X_2$.

Lemma 2.5.3. *If X_1, X_2 are multinomially distributed with parameters $\{N; p, 1-p\}$, then for all N and for all r,*

$$\max_p P_{(p,1-p)} \left(X_1^2 + X_2^2 \leq r \right)$$

is achieved by $p = \frac{1}{2}$.

Proof of Lemma. Two observations simplify the problem. The first is that (X_1, X_2) has the same distribution as $(Y, N - Y)$, where Y is binomially distributed with parameters (N, p). (This is essentially the observation that the multinomial distribution generalizes the binomial distribution.) The second is algebraic: for every r, $\exists r''$ such that

$$\left\{ Y^2 + (N - Y)^2 \leq r \right\} = \left\{ \left| Y - \frac{N}{2} \right| \leq r'' \right\}.$$

To see this, define

$$f(Y) = Y - \frac{N}{2}$$

then

$$\begin{aligned}\left\{Y^2+(N-Y)^2\le r\right\} &= \left\{\left(\frac{N}{2}+f(Y)\right)^2+\left(\frac{N}{2}-f(Y)\right)^2\le r\right\}\\ &= \left\{2\left(\frac{N}{2}\right)^2+2f(Y)^2\le r\right\}\\ &= \left\{f(Y)^2\le r'\right\}\\ &= \left\{\left|Y-\frac{N}{2}\right|\le r''\right\},\end{aligned}$$

using $r'=\frac{r-2\left(\frac{N}{2}\right)^2}{2}$, $r''=\sqrt{r'}$. Thus if we take Y as a binomially-distributed random variable with parameters (N,p), it suffices to prove that for all N, and all r

$$\arg\max_p P\left(\left|Y-\frac{N}{2}\right|\le r\right)=\frac{1}{2}.$$

We seek to maximize the binomial probability $g(p)$ with respect to the parameter p:

$$g(p):=\sum_{j=\lceil\frac{N}{2}-r\rceil}^{\lfloor\frac{N}{2}+r\rfloor}\binom{N}{j}p^j(1-p)^{N-j}=\sum_{j=k}^{j=N-k}\binom{N}{j}p^j(1-p)^{N-j}.$$

Differentiating g with respect to p, we obtain:

$$g'(p)=\sum_{j=k}^{N-k}\left[\binom{N}{j}jp^{j-1}(1-p)^{N-j}-\binom{N}{j}(N-j)p^j(1-p)^{N-j-1}\right]$$

Let us define a_j and b_j as follows:

$$a_j:=\binom{N}{j}jp^{j-1}(1-p)^{N-j}\qquad b_j:=\binom{N}{j}(N-j)p^j(1-p)^{N-j-1}$$

i.e., so that $g'(p)=\sum_{j=k}^{N-k}a_j-b_j$. Now using the identity $\binom{N}{j}(N-j)=\binom{N}{j+1}(j+1)$, we can observe

$$\begin{aligned}b_j&=\binom{N}{j}(N-j)p^j(1-p)^{N-j-1}\\ &=\binom{N}{j+1}(j+1)p^j(1-p)^{N-j-1}\\ &=a_{j+1}.\end{aligned}$$

So $g'(p)$ forms a telescoping sum

$$\begin{aligned}g'(p)&=\sum_{j=k}^{N-k}a_j-b_j\\ &=a_k-b_{N-k}\\ &=\binom{N}{k}k\left[p^{k-1}(1-p)^{N-k}-p^{N-k}(1-p)^{k-1}\right]\\ &=\binom{N}{k}k\left[p(1-p)\right]^{k-1}\left[(1-p)^{N-2k+1}-p^{N-2k+1}\right],\end{aligned}$$

assuming $p > 0$. We can conclude by inspection that $g'(p) = 0$ at $p = \frac{1}{2}$, $g'(p) > 0$ for $0 < p < \frac{1}{2}$, and $g'(p) < 0$ for $\frac{1}{2} < p < 1$, by inspection. Thus $g(p)$ attains its maximum at $p = \frac{1}{2}$, establishing the lemma.

Lemma 2.5.4. *If $X_1, X_2, ..., X_n$ are multinomial random variables distributed with parameters $\{N; p_1, p_2, ..., p_n\}$, (respecting $\sum_{i=1}^n p_i = 1$), then for all r, and for all N,*

$$\max_{(p_1,p_2,...,p_n)} P\left(\sum_{i=1}^n X_i^2 \leq r\right)$$

is achieved by $(p_1, p_2, ..., p_n) = (1/n, 1/n, ..., 1/n)$.

Proof. When $p = (p_1, p_2, ..., p_n)$, we write

$$P_p\left(\sum_{i=1}^n X_i^2 \leq r\right)$$

to denote the probability that $X_i^2 \leq r$ when $X_1, X_2, ..., X_n$ are distributed multinomially with parameters $\{N; p_1, p_2, ..., p_n\}$. Define

$$p^* := (1/n, 1/n, ..., 1/n),$$

and fix an arbitrary $p^{(0)} \in \{\vec{p} \in \mathbb{R}^n \,|\, 0 \leq p_i \leq 1 \,\forall i, \sum_{i=1}^n p_i = 1\}$. Then we would like to show that

$$P_{p^{(0)}}\left(\sum_{i=1}^n X_i^2 \leq r\right) \leq P_{p^*}\left(\sum_{i=1}^n X_i^2 \leq r\right)$$

We will construct a sequence of vectors $\vec{p_0}, \vec{p_1}, \vec{p_2}, ...$ such that

$$P_{p^{(j)}}\left(\sum_{i=1}^n X_i^2 \leq r\right) \leq P_{p^{(j+1)}}\left(\sum_{i=1}^n X_i^2 \leq r\right) \quad \forall i \tag{2.9}$$

and

$$\lim_{j\to\infty} p^{(j)} = p^* \tag{2.10}$$

The continuity of $P_p\left(\sum_{i=1}^n X_i^2 \leq r\right)$ in p will then imply

$$P_{p^{(0)}}\left(\sum_{i=1}^n X_i^2 \leq r\right) \leq \lim_{j\to\infty} P_{p^{(j)}}\left(\sum_{i=1}^n X_i^2 \leq r\right) = P_{p*}\left(\sum_{i=1}^n X_i^2 \leq r\right).$$

We construct the sequence $\{p^{(j)}\}$ as follows. Where $p^{(j)} = (p_1^{(j)}, p_2^{(j)}, ..., p_n^{(j)})$, choose

$$\alpha^j := \arg\max_{1\leq k\leq n} p_k^{(j)} \qquad \beta^j := \arg\min_{1\leq k\leq n} p_k^{(j)}$$

Then define $p^{(j+1)}$ via

$$p_k^{(j+1)} := \begin{cases} \frac{p_{\alpha^j}^{(j)} + p_{\beta^j}^{(j)}}{2} & \text{if } k \in \{\alpha^j, \beta^j\}, \\ p_k^{(j)} & \text{otherwise.} \end{cases}$$

First, we establish (2.9). Without loss of generality, we assume $\alpha^j = 1$ and $\beta^j = 2$ (this solely acts to simplify the indexing notation). Observe

$$P_{p^{(j)}}(X_1 + X_2 = k) = P_{p^{(j+1)}}(X_1 + X_2 = k), \quad \forall k, \tag{2.11}$$

since

$$\begin{aligned}
P_p(X_1+X_2=k) &= \sum_{\substack{x_1,\ldots,x_n \\ x_1+x_2=k \\ x_3+\ldots+x_n=N-k}} \binom{N}{x_1\, x_2 \ldots x_n} \prod_{i=1}^{n} p_i^{x_i} \\
&= \sum_{\substack{x_3,\ldots,x_n \\ x_3+\ldots+x_n=N-k}} \frac{N!}{x_3!\ldots x_n!} \prod_{i=3}^{n} p_i^{x_i} \sum_{\substack{x_1,x_2 \\ x_1+x_2=k}} \frac{1}{x_1!x_2!} p_1^{x_1} p_2^{x_2} \\
&= \sum_{\substack{x_3,\ldots,x_n \\ x_3+\ldots+x_n=N-k}} \frac{N!}{x_3!\ldots x_n!} \prod_{i=3}^{n} p_i^{x_i} \cdot \frac{1}{k!} \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} p_1^{j} p_2^{k-j} \\
&\overset{(a)}{=} \sum_{\substack{x_3,\ldots,x_n \\ x_3+\ldots+x_n=N-k}} \frac{N!}{x_3!\ldots x_n!} \prod_{i=3}^{n} p_i^{x_i} \cdot \frac{1}{k!} (p_1+p_2)^k,
\end{aligned}$$

where (a) follows from an application of the binomial formula. Since $p_1^{(j+1)} + p_2^{(j+1)} = p_1^{(j)} + p_2^{(j)}$ and $p_k^{(j+1)} = p_k^{(j)}$ for $3 \le k \le n$, this implies Eq. (2.11). Furthermore,

$$P_{p^{(j)}}(X_3, X_4, ..., X_n | X_1 + X_2 = k) = P_{p^{(j+1)}}(X_3, X_4, ..., X_n | X_1 + X_2 = k) \tag{2.12}$$

since

$$\begin{aligned}
P_{p^{(j)}}(X_3, X_4, ..., X_n | X_1 + X_2 = k) &\overset{(a)}{=} P_{p^{(j)}}(X_3, X_4, ..., X_n | X_1 = 0, X_2 = k) \\
&\overset{(b)}{=} P_{p^{(j+1)}}(X_3, X_4, ..., X_n | X_1 = 0, X_2 = k) \\
&\overset{(a)}{=} P_{p^{(j+1)}}(X_3, X_4, ..., X_n | X_1 + X_2 = k).
\end{aligned}$$

where (a) follows from Corollary (2.5.2), (b) from Lemma (2.5.1).

Equations (2.11) and (2.12) imply

$$P_{p^{(j)}}(X_3, X_4, ..., X_n) = P_{p^{(j+1)}}(X_3, X_4, ..., X_n) \tag{2.13}$$

since

$$\begin{aligned}
&P_{p^{(j)}}(X_3, X_4, ..., X_n) \\
&= \sum_k P_{p^{(j)}}(X_3, , ..., X_n | X_1 + X_2 = k) P_{p^{(j)}}(X_1 + X_2 = k)) \\
&\overset{(a)}{=} \sum_k P_{p^{(j+1)}}(X_3 =, ..., X_n = | X_1 + X_2 = k) P_{p^{(j+1)}}(X_1 + X_2 = k) \\
&= P_{p^{(j+1)}}(X_3, X_4, ..., X_n)
\end{aligned}$$

where (a) follows from (2.12) and (2.11).

Now we can combine these arguments to conclude

$$P_{p^{(j)}}\left(\sum_{i=1}^{n} X_i^2 \le r\right)$$

$$= \sum_{\substack{m_3,\ldots,m_n \\ \sum_{i=3}^n m_i \le N \\ \sum_{i=3}^n m_i^2 \le r}} P_{p^{(j)}}\left(X_1^2 + X_2^2 \le r - \sum_{i=3}^n m_i^2 \middle| X_3 = m_3, ..., X_n = m_n \right) P_{p^{(j)}}(X_3 = m_3, ..., X_n = m_n)$$

$$\overset{(a)}{=} \sum_{\substack{m_3,\ldots,m_n \\ \sum_{i=3}^n m_i \le N \\ \sum_{i=3}^n m_i^2 \le r}} P_{p^{(j)}}\left(X_1^2 + X_2^2 \le r - \sum_{i=3}^n m_i^2 \middle| X_3 = m_3, ..., X_n = m_n \right) P_{p^{(j+1)}}(X_3 = m_3, ..., X_n = m_n)$$

$$\overset{(b)}{\le} \sum_{\substack{m_3,\ldots,m_n \\ \sum_{i=3}^n m_i \le N \\ \sum_{i=3}^n m_i^2 \le r}} P_{p^{(j+1)}}\left(X_1^2 + X_2^2 \le r - \sum_{i=3}^n m_i^2 \middle| X_3 = m_3, ..., X_n = m_n \right) P_{p^{(j+1)}}(X_3 = m_3, ..., X_n = m_n)$$

$$= P_{p^{(j+1)}}\left(\sum_{i=1}^n X_i^2 \le r \right)$$

where (a) follows from (2.13), and (b) follows from an application of Lemma 2.5.1 and Lemma 2.5.3.
Finally we return to Eq (2.10)

$$\lim_{j\to\infty} p^{(j)} = p^*.$$

Note that

$$p_{\alpha^j}^{(j)} = \max_{1\le k\le n} p_k^{(j)} \ge \frac{1}{n}\sum_{l=1}^n p_l^{(j)} = \frac{1}{n} \quad \forall j \tag{2.14}$$

and analogously

$$p_{\beta^j}^{(j)} = \min_{1\le k\le n} p_k^{(j)} \le \frac{1}{n}\sum_{l=1}^n p_l^{(j)} = \frac{1}{n} \quad \forall j. \tag{2.15}$$

Hence we have

$$p_{\alpha^j}^{(j+1)} = \frac{1}{2}\left(p_{\alpha^j}^{F(j)} + p_{\beta^j}^{(j)} \right) \le \frac{1}{2}\left(p_{\alpha^j}^{(j)} + \frac{1}{n} \right). \tag{2.16}$$

So that

$$\begin{aligned}
p_{\alpha^{j+1}}^{(j+1)} &\le \max\left\{ \frac{1}{2}\left(p_{\alpha^j}^{(j)} + \frac{1}{n} \right), \max_{k\neq\alpha^j} p_k^{(j)} \right\} \\
p_{\alpha^{j+2}}^{(j+2)} &\le \max\left\{ \frac{1}{2}\left(p_{\alpha^j}^{(j)} + \frac{1}{n} \right), \frac{1}{2}\left(p_{\alpha^{j+1}}^{(j+1)} + \frac{1}{n} \right), \max_{k\notin\{\alpha^j,\alpha^{j+1}\}} p_k^{(j)} \right\} \\
&\vdots \\
p_{\alpha^{j+n}}^{(j+n)} &\le \max\left\{ \frac{1}{2}\left(p_{\alpha^j}^{(j)} + \frac{1}{n} \right), \frac{1}{2}\left(p_{\alpha^{j+1}}^{(j+1)} + \frac{1}{n} \right), ..., \frac{1}{2}\left(p_{\alpha^{j+n-1}}^{(j+n-1)} + \frac{1}{n} \right) \right\} \\
&\le \frac{1}{2}\left(p_{\alpha^j}^{(j)} + \frac{1}{n} \right),
\end{aligned} \tag{2.17}$$

since $p_{\alpha^{j+1}}^{(j+1)} \le p_{\alpha^j}^{(j)}\ \forall j$, by construction. (2.17) implies $\limsup_{j\to\infty} \max_{1\le k\le n} p_k^{(j)} \le \frac{1}{n}$ (One can see this by considering the first order linear difference equation $y^{(j+1)} = \frac{1}{2}(y^{(j)} + \frac{1}{n})$, and noting $y^{(nj)} \ge p_{\alpha^j}^{(j)}$). Since $\sum_{k=1}^n p_k^{(j)} = 1\ \forall j$, we have

$$\lim_{j\to\infty} \vec{p^j} = \vec{p^*},$$

establishing the lemma.

Proposition 1. *If $m_1, m_2, ..., m_n$ are independent Poisson random variables with rates $\lambda_1, \lambda_2, ..., \lambda_n$, respectively, then for all r, and for all $\hat{\mu}$, we haves*

$$\max_{\lambda_1,\lambda_2,...,\lambda_n} P\left(\sum_{i=1}^{n} m_i^2 \leq r \,\middle|\, \hat{\mu}\right) = P\left(\sum_{i=1}^{n} X_i^2 \leq r\right)$$

where $X_1, X_2, ..., X_n$ are distributed multinomially with parameters $\{n\hat{\mu}; 1/n, 1/n, ..., 1/n\}$.

Proof. *Conditioned* on the event $\{\sum_{i=1}^{n} m_i = n\hat{\mu}\}, X_1, X_2, ...X_n$ is distributed multinomially with parameters $\{n\hat{\mu}; \frac{\lambda_1}{\sum_{i=1}^{n}\lambda_i}, \frac{\lambda_2}{\sum_{i=1}^{n}\lambda_i}, ..., \frac{\lambda_n}{\sum_{i=1}^{n}\lambda_i}\}$:

$$\begin{aligned}\frac{P(m_1 = x_1, ..., m_n = x_n)}{P(\sum_{i=1}^{n} m_i = n\hat{\mu})} &= \frac{\prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}}{\frac{e^{-\sum_{i=1}^{n}\lambda_i}(\sum_{i=1}^{n}\lambda_i)^{n\hat{\mu}}}{(n\hat{\mu})!}} \\ &= \binom{n\hat{\mu}}{m_1\, m_2 \ldots m_n} \prod_{i=1}^{n} \left(\frac{\lambda_i}{\sum_{i=1}^{n}\lambda_i}\right)^{x_i}\end{aligned} \tag{2.18}$$

As a consequence, the proposition follows as a corollary of Lemma 2.5.4.

2.5.2 Dynamic Programming Algorithm

Dynamic programming [6] is a technique for computing functionals of functions on high-dimensional spaces which exploits "conditional independence"-like relations (literally in somes cases such as in Markov random fields) among variables by breaking the computational problem down into smaller subproblem which are then recombined, for the purpose of gains in computational efficiency. The technique is quite general (see [20, 33]), but an elementary example suffices for our purposes. Suppose the function $f : X^N \to \mathbb{R}$, where X is some finite state space, can be decomposed as follows:

$$f(x_1, x_2, ..., x_n) = f_1(x_1, x_2) f_2(x_2, x_3) \cdots f_{n-1}(x_{n-1}, x_n), \tag{2.19}$$

and we would like to compute

$$\sum_{(x_1,x_2,...,x_n)\in X^N} f(x_1, x_2, ..., x_n), \tag{2.20}$$

then one could utilize the decomposition (2.19) as follows:

$$\begin{aligned}&\sum_{x_1,x_2,...,x_n} f(x_1, x_2, ..., x_n) \\ &\quad= \sum_{x_1,x_2,...,x_n} f_1(x_1, x_2) f_2(x_2, x_3) \cdots f_{n-1}(x_{n-1}, x_n) \\ &\quad= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} f_1(x_1, x_2) f_2(x_2, x_3) \cdots f_{n-1}(x_{n-1}, x_n) \\ &\quad= \sum_{x_1}\sum_{x_2} f_1(x_1, x_2) \sum_{x_3} f_2(x_2, x_3) \sum_{x_4} \cdots \sum_{x_n} f_{n-1}(x_{n-1}, x_n)\end{aligned} \tag{2.21}$$

As a consequence, to obtain (2.20) one can work backwards through the sums in (2.21):

$$\begin{aligned} g_{n-1}(x_{n-1}) &:= \sum_{x_n} f_{n-1}(x_{n-1}, x_n) \\ g_{n-2}(x_{n-2}) &:= \sum_{x_{n-1}} f_{n-2}(x_{n-2}, x_{n-1}) g_{n-1}(x_{n-1}) \\ g_{n-3}(x_{n-3}) &:= \sum_{x_{n-2}} f_{n-3}(x_{n-3}, x_{n-2}) g_{n-2}(x_{n-2}) \\ &\vdots \\ g_2(x_2) &:= \sum_{x_3} f_2(x_2, x_3) g_3(x_3) \\ g_1(x_1) &:= \sum_{x_2} f_1(x_1, x_2) g_2(x_2) \end{aligned} \tag{2.22}$$

and then (through (2.21))

$$\sum_{(x_1, x_2, ..., x_n) \in X^N} f(x_1, x_2, ..., x_n) = \sum_{x_1} g_1(x_1). \tag{2.23}$$

The computational savings can be considerable: in this problem the brute-force approach (i.e., enumerating all possibilities and summing directly) would require $O(|X|^N)$ operations, whereas dynamic programming computes the sum $O((n-1)|X|^2)$ operations, with $|X|$ the cardinality of the space X.

Returning to our multinomial probability, we seek to compute

$$P\left(\sum_{i=1}^{n} X_i^2 \leq k\right) = \sum_{\substack{m_1, m_2, ..., m_n \\ \sum_{i=1}^n m_i = N \\ \sum_{i=1}^n m_i^2 \leq k}} \binom{N}{m_1\, m_2\, \cdots\, m_n} \left(\frac{1}{n}\right)^N, \tag{2.24}$$

where $X_1, X_2, ..., X_n \sim \mathcal{M}(N; \frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$. This can be decomposed into a form amenable to dynamic programming: the basic idea is to transform the apparently global constraints $\sum_{i=1}^n m_i = N$ and $\sum_{i=1}^n m_i^2 \leq k$ into local constraints by working directly with the partial sums $\sum_{i=1}^j m_i$ and $\sum_{i=1}^j m_i^2$. If we employ the substitution

$$s_j = \sum_{k=1}^{j} m_k \qquad \tilde{s}_j = \sum_{k=1}^{j} m_k^2, \tag{2.25}$$

we observe that the constraint sets

$$\left\{ m_1, m_2, ..., m_n : \sum_{i=1}^{n} m_i = N, \sum_{i=1}^{n} m_i^2 \leq k \right\} \tag{2.26}$$

and

$$\begin{aligned} B := \{ s_1, s_2, ..., s_n, \tilde{s}_1, \tilde{s}_2, ..., \tilde{s}_n : & s_n = N, \tilde{s}_n \leq k, \\ & s_{j-1} \leq s_j \; \forall j, \\ & \tilde{s}_j = \tilde{s}_{j-1} + (s_j - s_{j-1})^2 \; \forall j \}, \end{aligned} \tag{2.27}$$

are equivalent (with respect to (2.26)). Therefore,

$$\begin{aligned}
&\sum_{\substack{m_1,m_2,\ldots,m_n \\ \sum_{i=1}^n m_i = N \\ \sum_{i=1}^n m_i^2 \le k}} \binom{N}{m_1\, m_2\, \cdots\, m_n} \left(\frac{1}{n}\right)^N \\
&\qquad = \sum_{\substack{s_1,\ldots,s_n \\ \tilde{s}_1,\ldots,\tilde{s}_n \\ s_n = N \\ \tilde{s}_n \le k \\ s_{j-1} \le s_j\ \forall j \\ \tilde{s}_j = \tilde{s}_{j-1} + (s_j - s_{j-1})^2\ \forall k}} \binom{N}{N - s_{n-1}\, s_{n-2} - s_{n-1} \ldots s_2 - s_1\, s_1} \left(\frac{1}{n}\right)^N \\
&\qquad = \sum_{\substack{s_1,\ldots,s_n \\ \tilde{s}_1,\ldots,\tilde{s}_n}} \mathbf{1}_B(s_1,\ldots,s_n,\tilde{s}_1,\ldots,\tilde{s}_n) \binom{N}{N - s_{n-1}\, s_{n-2} - s_{n-1} \ldots s_2 - s_1\, s_1} \left(\frac{1}{n}\right)^N
\end{aligned} \tag{2.28}$$

Now, using the pairwise subconstraints

$$\begin{aligned}
C := \{(s_j, s_{j-1}, \tilde{s}_j, \tilde{s}_{j-1}) : & s_{j-1} \le s_j \\
& \tilde{s}_j = \tilde{s}_{j-1} + (s_j - s_{j-1})^2 \\
& \tilde{s}_j \le k, s_j \le N\},
\end{aligned} \tag{2.29}$$

we have

$$\mathbf{1}_B(s_1,\ldots,s_n,\tilde{s}_1,\ldots\tilde{s}_n) = \mathbf{1}_{\{N\}}(s_n) \cdot \mathbf{1}_C(0, s_1, \tilde{s}_1, 0) \cdot \prod_{i=2}^{n} \mathbf{1}_C(s_j, s_{j-1}, \tilde{s}_j, \tilde{s}_{j-1}), \tag{2.30}$$

(the second indicator function on the righthand side is analogous to implicitly enforcing

$s_0 = 0, \tilde{s}_0 = 0$). And returning to (2.28)

$$
\begin{aligned}
&\sum_{\substack{s_1,\ldots,s_n \\ \tilde{s}_1,\ldots,\tilde{s}_n}} \mathbf{1}_B(s_1,\ldots,s_n,\tilde{s}_1,\ldots,\tilde{s}_n) \binom{N}{N - s_{n-1}\, s_{n-2} - s_{n-1} \ldots s_2 - s_1\, s_1} \left(\frac{1}{n}\right)^N \\
&\quad = \sum_{\substack{s_1,\ldots,s_n \\ \tilde{s}_1,\ldots,\tilde{s}_n}} \mathbf{1}_{\{N\}}(s_n) \cdot \mathbf{1}_C(0, s_1, \tilde{s}_1, 0) \cdot \prod_{i=2}^{n} \mathbf{1}_C(s_j, s_{j-1}, \tilde{s}_j, \tilde{s}_{j-1}) \\
&\qquad\qquad \cdot \binom{N}{N - s_{n-1}\, s_{n-2} - s_{n-1} \ldots s_2 - s_1\, s_1} \left(\frac{1}{n}\right)^N \\
&\quad = \sum_{\substack{s_1,\ldots,s_n \\ \tilde{s}_1,\ldots,\tilde{s}_n}} \mathbf{1}_{\{N\}}(s_n) \cdot \mathbf{1}_C(0, s_1, \tilde{s}_1, 0) \cdot \prod_{i=2}^{n} \mathbf{1}_C(s_j, s_{j-1}, \tilde{s}_j, \tilde{s}_{j-1}) \\
&\qquad\qquad \cdot \binom{N}{s_{n-1}} \binom{s_{n-1}}{s_{n-2}} \cdots \binom{s_2}{s_1} \cdot \left(\frac{1}{n}\right)^N \\
&\quad = \sum_{\substack{s_n \\ \tilde{s}_n}} \sum_{\substack{s_{n-1} \\ \tilde{s}_{n-1}}} \cdots \sum_{\substack{s_1 \\ \tilde{s}_1}} \mathbf{1}_{\{N\}}(s_n) \cdot \mathbf{1}_C(0, s_1, \tilde{s}_1, 0) \cdot \prod_{i=2}^{n} \mathbf{1}_C(s_j, s_{j-1}, \tilde{s}_j, \tilde{s}_{j-1}) \\
&\qquad\qquad \cdot \binom{N}{s_{n-1}} \binom{s_{n-1}}{s_{n-2}} \cdots \binom{s_2}{s_1} \cdot \left(\frac{1}{n}\right)^N \\
&\quad = \left(\frac{1}{n}\right)^N \cdot \sum_{\substack{s_n \\ \tilde{s}_n}} \mathbf{1}_{\{N\}}(s_n) \sum_{\substack{s_{n-1} \\ \tilde{s}_{n-1}}} \binom{N}{s_{n-1}} \mathbf{1}_C(s_n, s_{n-1}, \tilde{s}_n, \tilde{s}_{n-1}) \\
&\qquad\qquad \cdot \sum_{\substack{s_{n-2} \\ \tilde{s}_{n-2}}} \binom{s_{n-1}}{s_{n-2}} \mathbf{1}_C(s_{n-1}, s_{n-2}, \tilde{s}_{n-1}, \tilde{s}_{n-2}) \cdots \sum_{\substack{s_1 \\ \tilde{s}_1}} \binom{s_2}{s_1} \mathbf{1}_C(s_2, s_1, \tilde{s}_2, \tilde{s}_1) \\
&\quad = \cdot \sum_{\substack{s_n \\ \tilde{s}_n}} \mathbf{1}_{\{N\}}(s_n) \sum_{\substack{s_{n-1} \\ \tilde{s}_{n-1}}} \left(\frac{1}{N}\right)^{N - s_{n-1}} \binom{N}{s_{n-1}} \mathbf{1}_C(s_n, s_{n-1}, \tilde{s}_n, \tilde{s}_{n-1}) \\
&\qquad\qquad \cdot \sum_{\substack{s_{n-2} \\ \tilde{s}_{n-2}}} \left(\frac{1}{N}\right)^{s_{n-1} - s_{n-2}} \binom{s_{n-1}}{s_{n-2}} \mathbf{1}_C(s_{n-1}, s_{n-2}, \tilde{s}_{n-1}, \tilde{s}_{n-2}) \\
&\qquad\qquad \cdot \sum \cdots \sum_{\substack{s_1 \\ \tilde{s}_1}} \binom{s_2}{s_1} \left(\frac{1}{N}\right)^{s_2} \mathbf{1}_C(s_2, s_1, \tilde{s}_2, \tilde{s}_1),
\end{aligned}
\tag{2.31}
$$

As in (2.22), one can compute

$$
\begin{aligned}
g_2(s_2,\tilde{s}_2) &:= \sum_{\substack{s_1\\ \tilde{s}_1}} \binom{s_2}{s_1}\left(\frac{1}{N}\right)^{s_2} \mathbf{1}_C(s_2,s_1,\tilde{s}_2,\tilde{s}_1)\\
g_3(s_3,\tilde{s}_3) &:= \sum_{\substack{s_2\\ \tilde{s}_2}} \binom{s_3}{s_2}\left(\frac{1}{N}\right)^{s_3-s_2} \mathbf{1}_C(s_3,s_2,\tilde{s}_3,\tilde{s}_2) g_2(s_2,\tilde{s}_2)\\
&\vdots\\
g_{n-1}(s_{n-1},\tilde{s}_{n-1}) &:= \sum_{\substack{s_{n-2}\\ \tilde{s}_{n-2}}} \binom{s_{n-1}}{s_{n-2}}\left(\frac{1}{N}\right)^{s_{n-1}-s_{n-2}} \mathbf{1}_C(s_{n-1},s_{n-2},\tilde{s}_{n-1},\tilde{s}_{n-2}) g_{n-2}(s_{n-2},\tilde{s}_{n-2})\\
g_n(\tilde{s}_n) &:= \sum_{\substack{s_{n-1}\\ \tilde{s}_{n-1}}} \binom{N}{s_{n-1}}\left(\frac{1}{N}\right)^{N-s_{n-1}} \mathbf{1}_C(N,s_{n-1},\tilde{s}_n,\tilde{s}_{n-1}) g_{n-1}(s_{n-1},\tilde{s}_{n-1})
\end{aligned}
\tag{2.32}
$$

and then

$$
P\left(\sum_{i=1}^n X_i^2 \le k\right) = \sum_{\tilde{s}_n} g_n(\tilde{s}_n). \tag{2.33}
$$

2.5.3 Monte Carlo Estimation

The dynamic programming solution to the calculation of the multinomial probability (2.24), $p = P(\sum_{i=1}^n X_i^2 \le k)$ can be computationally infeasible if $\sum_{i=1}^n m_i$ and $\sum_{i=1}^n m_i^2$ are too large. An alternative way to compute it is via approximation by Monte Carlo methods [17, 26]. The idea is to produce M i.i.d. vector samples $X^{(1)}, X^{(2)}, ..., X^{(M)}$ where $X^{(j)} = (X_1^{(j)}, X_2^{(j)}, ..., X_n^{(j)})$ and $X_1^{(j)}, X_2^{(j)}, ..., X_n^{(j)} \sim \mathcal{M}(N; \frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$ for each j. Then defining

$$
Y^{(j)} := \mathbf{1}_{\{\sum_{i=1}^n (X_i^{(j)})^2 \le k\}}(X^{(j)}), \tag{2.34}
$$

the law of large of numbers implies

$$
\lim_{M\to\infty} \frac{1}{M}\sum_{j=1}^M Y^{(j)} = E\left[Y^{(1)}\right] = P\left(\sum_{i=1}^n X_i^2 \le k\right), \text{(w.p.1)} \tag{2.35}
$$

which provides the estimate

$$
\hat{p} := \frac{1}{M}\sum_{j=1}^M Y^{(j)} \approx P\left(\sum_{i=1}^n X_i^2 \le k\right), \tag{2.36}
$$

with M large. In order to assess the accuracy of the estimate $\hat{p}$, we seek to form α-level *confidence intervals* of tolerance ϵ which satisfy

$$
P(|\hat{p} - p| \le \epsilon) \ge \alpha. \tag{2.37}
$$

One can get a handle on ϵ by invoking a central limit theorem approximation,

$$
\hat{p} = \frac{1}{M}\sum_{j=1}^M Y^{(j)} \sim \text{Bin}\left(p, \frac{p(1-p)}{M}\right) \approx \mathcal{N}\left(p, \frac{p(1-p)}{M}\right). \tag{2.38}
$$

Then

$$P\left(|\hat{p}-p| \leq 3 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{M}}\right) \overset{\sim}{\geq} .99, \tag{2.39}$$

which provides 99% approximate confidence intervals for the Monte Carlo approximation $\hat{p}$. The central limit theorem approximation works well: as Freiberger and Grenander [17] point out the randomness of the error bound "is more nearly a psychological difficulty than a real one." In this problem, it may not be worth the effort of more sophisticated techniques.

2.5.4 Computing g by recursion

$$\begin{aligned} g(\vec{n}, \hat{\mu}_1, ..., \hat{\mu}_N) := \min &\left\{ t : P\left(\sum_{i=1}^{n} Z_i \geq t\right) \leq \beta \right\} \\ &\text{where } Z_1, Z_2, ..., Z_N \text{ areindependent, and} \\ &Z_i \sim \text{Be}(r^*(n_i, \alpha, \hat{\mu}_i))\ \forall i, \end{aligned} \tag{2.40}$$

For simplicity we will denote

$$p_i := r^*(n_i, \alpha, \hat{\mu}_i). \tag{2.41}$$

Then observe that

$$\begin{aligned} P\left(\sum_{i=1}^{j} Z_i = t\right) &= P\left(\sum_{i=1}^{j-1} Z_i = t\right) P\left(Z_j = 0\right) + P\left(\sum_{i=1}^{j-1} Z_i = t-1\right) P\left(Z_j = 1\right) \\ &= P\left(\sum_{i=1}^{j-1} Z_i = t\right)(1-p_j) + P\left(\sum_{i=1}^{j-1} Z_i = t-1\right) p_j \end{aligned} \tag{2.42}$$

and

$$P(Z_1 = 1) = p_1 \qquad P(Z_1 = 0) = 1 - p_1, \tag{2.43}$$

which determines a recursion that can be computed in less than $3N^2$ operations.

2.5.5 The most reliable outcome

One cause for a *failure* to reject the null hypothesis in a single cell-stimulus pair might simply be a lack of data, either in the form of a paucity of trials or a paucity of spikes. Indeed, for some values of n, the number of trials, and of N, the total number of spikes (i.e., $\sum_{i=1}^{n} m_i$), it is impossible to reject the null hypothesis for *any* configuration of the data. This is the case if $P(\sum_{i=1}^{n} X_i^2 \leq k_*) > \alpha$ where

$$k_* := \min_{\substack{m_1,...,m_n \in \mathbb{Z}^n \\ \sum_{m_i} = N}} \sum_{i=1}^{n} m_i^2 = N, \tag{2.44}$$

and $X_1, X_2, ..., X_n \sim \mathcal{M}(N; 1/n, 1/n, ..., 1/n)$. Obtaining k_*, the value of $\sum_{i=1}^{n} m_i^2$ corresponding to the *most reliable outcome* (in mean square sense) consistent with a given n and N, is helpful in the optimizing the code for computing f, the threshold for significant rejection. Rather than use dynamic programming again, we can get this directly.

Lemma 2.5.5. *(The most reliable outcome)*

$$\begin{aligned} k_* &:= \min_{\substack{m_1,...,m_n \in \mathbb{Z}^n \\ \sum_{i=1}^{n} m_i = N}} \sum_{i=1}^{n} m_i^2 \\ &= \left[n\left(\left\lfloor \frac{N}{n} \right\rfloor + 1\right) - N \right] \left\lfloor \frac{N}{n} \right\rfloor^2 + \left(N - n\left\lfloor \frac{N}{n} \right\rfloor \right) \left(\left\lfloor \frac{N}{n} \right\rfloor + 1 \right)^2 \end{aligned} \tag{2.45}$$

Proof. In the continuum, it is straightforward to see that

$$\min_{\substack{x_1,\ldots,x_n\in\mathbb{R}^n \\ \sum_{i=1}^n x_i=N}} \sum_{i=1}^n x_i^2, \tag{2.46}$$

is achieved by

$$x_i = N/n \quad \forall i, \tag{2.47}$$

since any solution which satisfies $\sum_{i=1}^n x_i = N$, can be represented as

$$x_i = N/n + \epsilon_i \tag{2.48}$$

with $\epsilon_1, \epsilon_2, ..., \epsilon_n$ satisfying $\sum_{i=1}^n \epsilon_i = 0$. But then we can expand

$$\sum_{i=1}^n x_i^2 = \sum_{i=1}^n (N/n + \epsilon_i)^2 = N^2/n + \sum_{i=1}^n \epsilon^2, \tag{2.49}$$

which is evidently minimized by $\epsilon_i = 0 \, \forall i$. Thus the minimal $x^* = (x_1^*, ..., x_n^*)$ has $x_i^* = N/n \, \forall i$. (This of course immediately reveals the solution to the integer minimization problem in the case where $N \mod n = 0$, and is consistent with (2.45)).

Now a geometric argument reveals the effect of restricting $m_1, ..., m_n$ to the integer space $\mathbb{Z}^n$. Define the vertices of the hypercube on the integer lattice surrounding the point $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ as $H(x)$:

$$H(x_1, x_2, ..., x_n) = \{(y_1, y_2, ..., y_n) \in \mathbb{R}^n : y_i \in \{\lfloor x_i \rfloor, \lfloor x_i \rfloor + 1\} \ \forall i\}. \tag{2.50}$$

Then

$$\arg\min_{\substack{m_1,\ldots,m_n\in\mathbb{Z}^n \\ \sum_{i=1}^n m_i=N}} \sum_{i=1}^n m_i^2 \in H(x^*), \tag{2.51}$$

To show this, suppose not: then there exists a point $z \in \mathbb{Z}^n$ such that $z \notin H(x^*)$, and $\sum_{i=1}^n z_i^2 < \min_{y\in H(x^*)} \sum_{i=1}^n y_i^2$. But since $\sum_{i=1}^n x_i^2$ is radially symmetric, one can then trace an arc A from z to z' such that $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n z_i^2 \, \forall y \in A$, and such that the ray from x^* to z' intersects a vertex p (of the hypercube) in $H(x^*)$. Therefore, $\sum_{i=1}^n (z_i')^2 < \min_{y\in H(x^*)} \sum_{i=1}^n y_i^2$. But $\sum_{i=1}^n (x_i^*)^2 \leq \sum_{i=1}^n (z_i')^2$ and $\sum_{i=1}^n (z_i')^2 < \sum_{i=1}^n p_i^2$ contradicts the convexity of $f(x) = \sum_{i=1}^n x_i^2$, establishing (2.51).

Thus there exists $m_1^*, m_2^*, ..., m_n^* \in H(x^*)$ such that

$$\sum_{i=1}^n (m_i^*)^2 = \min_{\substack{m_1,\ldots,m_n \\ \sum_{i=1}^n m_i^2}} m_i^2. \tag{2.52}$$

This turns out to identify m^* because the condition $\sum_{i=1}^n m_i^* = N$ uniquely characterizes $m \in H(x^*)$. Essentially, $m_i = \left\lfloor \frac{N}{n} \right\rfloor$, or $m_i = \left\lfloor \frac{N}{n} \right\rfloor + 1$. Let j denote the number of variables among $m_1, ..., m_n$ that take the value $\left\lfloor \frac{N}{n} \right\rfloor$ (so $n - j$ variables take the value $\left\lfloor \frac{N}{n} \right\rfloor + 1$.) Then

$$\sum_{i=1}^n m_i = j \left\lfloor \frac{N}{n} \right\rfloor + (n-j) \left(\left\lfloor \frac{N}{n} \right\rfloor + 1 \right) = N \tag{2.53}$$

Solving for j, we get

$$j = n \left(\left\lfloor \frac{N}{n} \right\rfloor + 1 \right) - N. \tag{2.54}$$

This determines the unique point $m^* \in H(x^*)$ on the hyperplane $\sum_{i=1}^n m_i = N$. Plugging m^* into $\sum_{i=1}^n m_i^2$ finally produces (2.45).

2.5.6 Table of Significance Thresholds

Table 2.1: Significance Thresholds

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
2	1	1	0.000000	1	0.000000
2	2	2	0.000000	2	0.000000
2	3	5	0.000000	5	0.000000
2	4	8	0.000000	8	0.000000
2	5	13	0.000000	13	0.000000
2	6	18	0.000000	18	0.000000
2	7	25	0.000000	25	0.000000
2	8	32	0.000000	32	0.000000
2	9	41	0.000000	41	0.000000
2	10	50	0.000000	50	0.000000
2	11	61	0.000000	61	0.000000
2	12	72	0.000000	72	0.000000
2	13	85	0.000000	85	0.000000
2	14	98	0.000000	98	0.000000
2	15	113	0.000000	113	0.000000
2	16	128	0.000000	128	0.000000
2	17	145	0.000000	145	0.000000
2	18	162	0.000000	162	0.000000
2	19	181	0.000000	181	0.000000
2	20	200	0.000000	200	0.000000
3	1	1	0.000000	1	0.000000
3	2	2	0.000000	2	0.000000
3	3	3	0.000000	3	0.000000
3	4	6	0.000000	6	0.000000
3	5	9	0.000000	9	0.000000
3	6	12	0.000000	12	0.000000
3	7	17	0.000000	17	0.000000
3	8	22	0.000000	22	0.000000
3	9	27	0.000000	27	0.000000
3	10	34	0.000000	34	0.000000
3	11	41	0.000000	41	0.000000
3	12	48	0.000000	48	0.000000
3	13	57	0.000000	57	0.000000
3	14	66	0.000000	66	0.000000
3	15	75	0.000000	75	0.000000
3	16	86	0.000000	86	0.000000
3	17	97	0.000000	97	0.000000
3	18	109	0.044275	108	0.044275
3	19	121	0.000000	121	0.000000
3	20	134	0.000000	134	0.000000
3	21	148	0.038151	147	0.038151
3	22	162	0.000000	162	0.000000
3	23	177	0.000000	177	0.000000
3	24	193	0.033515	192	0.033515
3	25	209	0.000000	209	0.000000

continued on next page

Table 2.1: *continued*

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
3	26	226	0.000000	226	0.000000
3	27	244	0.029883	243	0.029883
3	28	262	0.000000	262	0.000000
3	29	281	0.000000	281	0.000000
3	30	301	0.026961	300	0.026961
4	1	1	0.000000	1	0.000000
4	2	2	0.000000	2	0.000000
4	3	3	0.000000	3	0.000000
4	4	4	0.000000	4	0.000000
4	5	7	0.000000	7	0.000000
4	6	10	0.000000	10	0.000000
4	7	13	0.000000	13	0.000000
4	8	17	0.038452	16	0.038452
4	9	21	0.000000	21	0.000000
4	10	26	0.000000	26	0.000000
4	11	31	0.000000	31	0.000000
4	12	37	0.022030	36	0.022030
4	13	43	0.000000	43	0.000000
4	14	50	0.000000	50	0.000000
4	15	57	0.000000	57	0.000000
4	16	65	0.014683	64	0.014683
4	17	74	0.049922	73	0.049922
4	18	82	0.000000	82	0.000000
4	19	92	0.042683	91	0.042683
4	20	101	0.010671	100	0.010671
4	21	112	0.037348	111	0.037348
4	22	122	0.000000	122	0.000000
4	23	134	0.032809	133	0.032809
4	24	145	0.008202	145	0.008202
4	25	158	0.029294	157	0.029294
4	26	171	0.040802	170	0.040802
4	27	184	0.026230	183	0.026230
4	28	197	0.006558	197	0.006558
4	29	212	0.023771	211	0.023771
4	30	227	0.033428	226	0.033428
4	31	242	0.021589	241	0.021589
4	32	257	0.005397	257	0.005397
4	33	274	0.019790	273	0.019790
4	34	291	0.028036	290	0.028036
4	35	308	0.018171	307	0.018171
4	36	325	0.004543	325	0.004543
4	37	344	0.016808	343	0.016808
4	38	363	0.023952	362	0.023952
4	39	382	0.015569	381	0.015569
4	40	403	0.046352	401	0.003892

continued on next page

Table 2.1: *continued*

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
5	1	1	0.000000	1	0.000000
5	2	2	0.000000	2	0.000000
5	3	3	0.000000	3	0.000000
5	4	4	0.000000	4	0.000000
5	5	6	0.038400	5	0.038400
5	6	8	0.000000	8	0.000000
5	7	11	0.000000	11	0.000000
5	8	14	0.000000	14	0.000000
5	9	17	0.000000	17	0.000000
5	10	21	0.011612	20	0.011612
5	11	26	0.042578	25	0.042578
5	12	30	0.000000	30	0.000000
5	13	35	0.000000	35	0.000000
5	14	41	0.027553	40	0.027553
5	15	46	0.005511	46	0.005511
5	16	53	0.022042	52	0.022042
5	17	60	0.037472	59	0.037472
5	18	67	0.033724	66	0.033724
5	19	74	0.016019	73	0.016019
5	20	81	0.003204	81	0.003204
5	21	90	0.013456	89	0.013456
5	22	99	0.023683	98	0.023683
5	23	108	0.021788	107	0.021788
5	24	117	0.010458	116	0.010458
5	25	128	0.036953	126	0.002092
5	26	137	0.009064	137	0.009064
5	27	148	0.016315	147	0.016315
5	28	161	0.047676	158	0.015227
5	29	172	0.045210	170	0.007360
5	30	183	0.026706	181	0.001472
5	31	196	0.040043	194	0.006519
5	32	209	0.037569	206	0.011920
5	33	222	0.035723	219	0.011239
5	34	235	0.034118	233	0.005459
5	35	250	0.045274	246	0.001092
5	36	263	0.030706	261	0.004913
5	37	278	0.029034	276	0.009089
5	38	293	0.027762	291	0.008635
5	39	310	0.046406	306	0.004209
5	40	325	0.035764	321	0.000842
5	41	342	0.042622	338	0.003835
5	42	357	0.023108	355	0.007159
5	43	374	0.022195	372	0.006841
5	44	393	0.037524	389	0.003344
5	45	410	0.028962	406	0.000669

continued on next page

Table 2.1: *continued*

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
5	46	429	0.034746	425	0.003077
5	47	448	0.047222	444	0.005784
5	48	467	0.045409	463	0.005553
5	49	486	0.030965	482	0.002721
5	50	507	0.048443	501	0.000544
6	1	1	0.000000	1	0.000000
6	2	2	0.000000	2	0.000000
6	3	3	0.000000	3	0.000000
6	4	4	0.000000	4	0.000000
6	5	5	0.000000	5	0.000000
6	6	7	0.015432	6	0.015432
6	7	9	0.000000	9	0.000000
6	8	12	0.000000	12	0.000000
6	9	15	0.000000	15	0.000000
6	10	18	0.000000	18	0.000000
6	11	22	0.020630	21	0.020630
6	12	25	0.003438	25	0.003438
6	13	30	0.014899	29	0.014899
6	14	35	0.028971	34	0.028971
6	15	40	0.032190	39	0.032190
6	16	45	0.021460	44	0.021460
6	17	50	0.008107	50	0.008107
6	18	57	0.031753	55	0.001351
6	19	62	0.006418	62	0.006418
6	20	69	0.013371	68	0.013371
6	21	76	0.015599	75	0.015599
6	22	85	0.048261	82	0.010725
6	23	92	0.037000	90	0.004111
6	24	99	0.017130	97	0.000685
6	25	108	0.030833	106	0.003426
6	26	117	0.033650	115	0.007423
6	27	126	0.030731	124	0.008907
6	28	135	0.029014	133	0.006235
6	29	146	0.047281	142	0.002411
6	30	155	0.035561	151	0.000402
6	31	166	0.041209	162	0.002076
6	32	175	0.021574	173	0.004614
6	33	186	0.019937	184	0.005639
6	34	199	0.044778	195	0.003994
6	35	210	0.031826	206	0.001553
6	36	223	0.046789	219	0.006916
6	37	234	0.028282	230	0.001368
6	38	247	0.035635	243	0.003095
6	39	260	0.039655	256	0.003832
6	40	273	0.031774	269	0.002737

continued on next page

Table 2.1: *continued*

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
6	41	286	0.022612	282	0.001069
6	42	301	0.033803	297	0.004854
6	43	316	0.046221	312	0.009335
6	44	331	0.046786	325	0.002194
6	45	346	0.041612	342	0.010000
6	46	361	0.042653	357	0.009670
6	47	376	0.038243	372	0.007635
6	48	393	0.047000	387	0.003560
6	49	408	0.035057	404	0.006928
6	50	425	0.035715	421	0.007972
6	51	442	0.031869	438	0.007522
6	52	459	0.032860	455	0.007309
6	53	478	0.042820	472	0.005791
6	54	495	0.036653	491	0.009738
6	55	514	0.039513	508	0.005308
6	56	533	0.044377	527	0.006148
6	57	552	0.047071	546	0.005826
6	58	571	0.041306	565	0.005683
6	59	590	0.034080	584	0.004514
6	60	611	0.045867	605	0.007654
7	1	1	0.000000	1	0.000000
7	2	2	0.000000	2	0.000000
7	3	3	0.000000	3	0.000000
7	4	4	0.000000	4	0.000000
7	5	5	0.000000	5	0.000000
7	6	7	0.042839	6	0.042839
7	7	8	0.006120	8	0.006120
7	8	11	0.024480	10	0.024480
7	9	14	0.047211	13	0.047211
7	10	16	0.000000	16	0.000000
7	11	20	0.044160	19	0.044160
7	12	23	0.022711	22	0.022711
7	13	26	0.007029	26	0.007029
7	14	31	0.029122	29	0.001004
7	15	34	0.005021	34	0.005021
7	16	39	0.011477	38	0.011477
7	17	44	0.015484	43	0.015484
7	18	49	0.013272	48	0.013272
7	19	56	0.044831	54	0.007205
7	20	61	0.028020	59	0.002287
7	21	68	0.049218	64	0.000327
7	22	73	0.022015	71	0.001797
7	23	80	0.027753	78	0.004429
7	24	87	0.026636	85	0.006327
7	25	94	0.024290	92	0.005649

continued on next page

Table 2.1: *continued*

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
7	26	103	0.049674	99	0.003147
7	27	110	0.036216	106	0.001012
7	28	117	0.024423	115	0.005000
7	29	126	0.030343	122	0.000838
7	30	135	0.035516	131	0.002155
7	31	144	0.040675	140	0.003182
7	32	153	0.037458	149	0.002909
7	33	162	0.027912	158	0.001646
7	34	173	0.046076	169	0.007194
7	35	182	0.029208	178	0.002741
7	36	193	0.040349	189	0.006166
7	37	204	0.045981	200	0.008258
7	38	215	0.042397	211	0.008243
7	39	226	0.039888	222	0.007727
7	40	237	0.038046	233	0.006753
7	41	250	0.049348	244	0.004355
7	42	261	0.039761	257	0.008588
7	43	274	0.043919	268	0.003822
7	44	287	0.046626	281	0.005205
7	45	298	0.027994	294	0.005256
7	46	313	0.048250	307	0.004968
7	47	326	0.039888	320	0.004374
7	48	339	0.033654	335	0.008345
7	49	354	0.042552	348	0.005690
7	50	367	0.030367	363	0.007481
7	51	382	0.032552	378	0.009169
7	52	397	0.035808	391	0.003554
7	53	412	0.034195	406	0.003381
7	54	429	0.047378	423	0.007858
7	55	444	0.040718	438	0.005798
7	56	461	0.047332	455	0.008726
7	57	476	0.037250	470	0.005262
7	58	493	0.039867	487	0.006506
7	59	510	0.038834	504	0.007849
7	60	527	0.037226	521	0.007491
7	61	546	0.049196	538	0.005669
7	62	563	0.041284	555	0.004191
7	63	582	0.049260	574	0.006370
7	64	599	0.038289	593	0.009310
7	65	618	0.042252	610	0.004782
7	66	637	0.043675	629	0.005807
7	67	656	0.042138	648	0.005568
7	68	675	0.037804	669	0.009825
7	69	696	0.047521	688	0.007607
7	70	715	0.038158	707	0.004791

continued on next page

Table 2.1: *continued*

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
8	1	1	0.000000	1	0.000000
8	2	2	0.000000	2	0.000000
8	3	3	0.000000	3	0.000000
8	4	4	0.000000	4	0.000000
8	5	5	0.000000	5	0.000000
8	6	6	0.000000	6	0.000000
8	7	8	0.019226	7	0.019226
8	8	9	0.002403	9	0.002403
8	9	12	0.010815	11	0.010815
8	10	15	0.023657	14	0.023657
8	11	18	0.032528	17	0.032528
8	12	21	0.030495	20	0.030495
8	13	24	0.019822	23	0.019822
8	14	27	0.008672	27	0.008672
8	15	32	0.034843	30	0.002323
8	16	35	0.011131	33	0.000290
8	17	40	0.024681	38	0.001645
8	18	45	0.034039	43	0.004319
8	19	50	0.034763	48	0.006839
8	20	55	0.031344	53	0.007124
8	21	60	0.027010	58	0.004987
8	22	65	0.019862	63	0.002285
8	23	72	0.037313	68	0.000626
8	24	77	0.021844	75	0.003364
8	25	84	0.029493	82	0.008189
8	26	91	0.033876	87	0.001390
8	27	98	0.038095	94	0.002346
8	28	105	0.037980	101	0.002566
8	29	112	0.031013	108	0.001860
8	30	119	0.022686	117	0.008035
8	31	128	0.037120	124	0.004296
8	32	135	0.023279	133	0.009492
8	33	144	0.031461	140	0.003545
8	34	153	0.038307	149	0.005473
8	35	162	0.038303	158	0.006048
8	36	171	0.035045	167	0.005738
8	37	180	0.032945	176	0.005150
8	38	191	0.047001	185	0.003920
8	39	200	0.038887	196	0.008197
8	40	211	0.047526	205	0.004844
8	41	220	0.033859	216	0.007034
8	42	231	0.035671	227	0.008534
8	43	242	0.038143	236	0.003224
8	44	253	0.038649	247	0.003102
8	45	264	0.033694	260	0.008757

continued on next page

Table 2.1: *continued*

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
8	46	277	0.049012	271	0.006480
8	47	288	0.039003	282	0.004651
8	48	301	0.048346	295	0.007099
8	49	312	0.034695	306	0.004083
8	50	325	0.039058	319	0.005032
8	51	338	0.039139	332	0.006024
8	52	351	0.036986	345	0.006330
8	53	364	0.035191	358	0.005338
8	54	379	0.045318	371	0.003966
8	55	392	0.038512	386	0.007098
8	56	407	0.044445	399	0.004433
8	57	420	0.034960	414	0.006384
8	58	435	0.037384	429	0.008207
8	59	450	0.039356	444	0.008518
8	60	465	0.039029	459	0.007992
8	61	480	0.035928	474	0.007710
8	62	497	0.046149	489	0.006694
8	63	512	0.039373	506	0.009817
8	64	529	0.046664	521	0.007195
8	65	544	0.036159	538	0.008933
8	66	561	0.038934	555	0.009711
8	67	578	0.038913	570	0.005789
8	68	595	0.037875	587	0.005459
8	69	614	0.049446	606	0.009803
8	70	631	0.044271	623	0.008147
8	71	648	0.039455	640	0.006842
8	72	667	0.044343	659	0.008848
8	73	686	0.048556	676	0.006283
8	74	703	0.038006	695	0.006873
8	75	722	0.039416	714	0.007621
8	76	743	0.049637	733	0.007954
8	77	762	0.048554	752	0.007037
8	78	781	0.045363	771	0.005859
8	79	800	0.039198	792	0.008704
8	80	821	0.044603	811	0.006426
9	1	1	0.000000	1	0.000000
9	2	2	0.000000	2	0.000000
9	3	3	0.000000	3	0.000000
9	4	4	0.000000	4	0.000000
9	5	5	0.000000	5	0.000000
9	6	6	0.000000	6	0.000000
9	7	8	0.037935	7	0.037935
9	8	9	0.008430	9	0.008430
9	9	12	0.034656	10	0.000937
9	10	13	0.004683	13	0.004683

continued on next page

Table 2.1: *continued*

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
9	11	16	0.011448	15	0.011448
9	12	19	0.017808	18	0.017808
9	13	22	0.019292	21	0.019292
9	14	25	0.015005	24	0.015005
9	15	28	0.008336	28	0.008336
9	16	33	0.033212	31	0.003176
9	17	36	0.014746	34	0.000750
9	18	41	0.032075	39	0.004082
9	19	46	0.045318	42	0.000528
9	20	49	0.016448	47	0.001563
9	21	54	0.018847	52	0.002837
9	22	59	0.017917	57	0.003468
9	23	66	0.048993	62	0.002954
9	24	71	0.037214	67	0.001751
9	25	76	0.026643	74	0.008104
9	26	83	0.042144	79	0.003679
9	27	88	0.025956	86	0.008924
9	28	95	0.033975	91	0.002862
9	29	102	0.041704	98	0.004904
9	30	109	0.044285	105	0.005958
9	31	116	0.041287	112	0.005895
9	32	123	0.037634	119	0.005333
9	33	130	0.033224	126	0.004370
9	34	139	0.047846	135	0.009995
9	35	146	0.037275	142	0.006627
9	36	155	0.046295	149	0.003390
9	37	162	0.031502	158	0.005473
9	38	171	0.034161	167	0.006896
9	39	180	0.035561	176	0.008048
9	40	189	0.036628	185	0.008711
9	41	198	0.033965	194	0.007998
9	42	207	0.028388	203	0.006253
9	43	218	0.041898	212	0.004548
9	44	227	0.031990	223	0.007998
9	45	238	0.040700	232	0.004909
9	46	249	0.046432	243	0.006912
9	47	260	0.049031	254	0.009040
9	48	269	0.033550	263	0.003984
9	49	280	0.032078	276	0.009730
9	50	293	0.048089	287	0.009111
9	51	304	0.042308	298	0.008247
9	52	315	0.036143	309	0.006418
9	53	328	0.044714	320	0.004261
9	54	339	0.035015	333	0.006672
9	55	352	0.040082	346	0.008844

continued on next page

Table 2.1: *continued*

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
9	56	365	0.044135	359	0.009944
9	57	378	0.045037	370	0.005668
9	58	391	0.043870	383	0.005522
9	59	404	0.041623	396	0.005214
9	60	417	0.038570	411	0.009005
9	61	432	0.047268	424	0.007486
9	62	445	0.041214	437	0.005907
9	63	460	0.046410	452	0.007845
9	64	473	0.037431	467	0.009620
9	65	488	0.039076	480	0.006030
9	66	503	0.040341	495	0.006532
9	67	518	0.040496	510	0.006977
9	68	533	0.038745	525	0.006634
9	69	550	0.048740	540	0.005619
9	70	565	0.043725	557	0.008897
9	71	580	0.037203	572	0.006799
9	72	597	0.042512	589	0.009124
9	73	614	0.046541	604	0.006181
9	74	631	0.048172	621	0.007375
9	75	648	0.048728	638	0.007983
9	76	665	0.049464	655	0.007789
9	77	682	0.047570	672	0.007454
9	78	699	0.043365	689	0.006995
9	79	716	0.039626	708	0.009507
9	80	735	0.045350	725	0.007989
9	81	752	0.039077	744	0.009572
9	82	771	0.042126	761	0.007352
9	83	790	0.044860	780	0.008066
9	84	809	0.045974	799	0.008634
9	85	828	0.044751	818	0.008959
9	86	847	0.043209	837	0.008617
9	87	866	0.041373	856	0.007671
9	88	887	0.047083	875	0.006612
9	89	906	0.042810	896	0.008606
9	90	927	0.046083	915	0.006714
10	1	1	0.000000	1	0.000000
10	2	2	0.000000	2	0.000000
10	3	3	0.000000	3	0.000000
10	4	4	0.000000	4	0.000000
10	5	5	0.000000	5	0.000000
10	6	6	0.000000	6	0.000000
10	7	7	0.000000	7	0.000000
10	8	9	0.018144	8	0.018144
10	9	10	0.003629	10	0.003629
10	10	13	0.016692	11	0.000363

continued on next page

Table 2.1: *continued*

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
10	11	16	0.037921	14	0.001996
10	12	17	0.005389	17	0.005389
10	13	20	0.009341	20	0.009341
10	14	23	0.011442	22	0.011442
10	15	26	0.010298	25	0.010298
10	16	31	0.040211	29	0.006865
10	17	34	0.027927	32	0.003335
10	18	37	0.015256	35	0.001125
10	19	42	0.033262	40	0.005940
10	20	47	0.049072	43	0.001449
10	21	50	0.023533	48	0.004158
10	22	55	0.028935	53	0.007470
10	23	60	0.031862	58	0.009608
10	24	65	0.033103	63	0.009874
10	25	70	0.030884	68	0.008989
10	26	75	0.025088	73	0.007536
10	27	82	0.046537	78	0.005470
10	28	87	0.033031	83	0.003093
10	29	94	0.048790	90	0.007644
10	30	99	0.030656	95	0.003410
10	31	106	0.039476	102	0.005955
10	32	113	0.044126	109	0.007852
10	33	120	0.045380	116	0.009125
10	34	127	0.045971	123	0.009926
10	35	134	0.044049	130	0.009624
10	36	141	0.038137	137	0.008026
10	37	148	0.031264	144	0.006002
10	38	157	0.045061	151	0.004187
10	39	164	0.033722	160	0.007429
10	40	173	0.042472	167	0.004368
10	41	182	0.049654	176	0.006193
10	42	189	0.032351	185	0.008109
10	43	198	0.034916	194	0.009496
10	44	207	0.034372	203	0.009629
10	45	216	0.032074	212	0.009010
10	46	227	0.047529	221	0.008203
10	47	236	0.040890	230	0.006842
10	48	245	0.034522	239	0.004926
10	49	256	0.042180	250	0.008036
10	50	265	0.032841	259	0.005006
10	51	276	0.037063	270	0.006947
10	52	287	0.040911	281	0.008240
10	53	298	0.042828	292	0.008876
10	54	309	0.042240	303	0.009361
10	55	320	0.040183	314	0.009270

continued on next page

Table 2.1: *continued*

# of Trials n	# of Spikes $n\hat{\mu}$	$f(n, .05, \hat{\mu})$	$r^*(n, .05, \hat{\mu})$	$f(n, .01, \hat{\mu})$	$r^*(n, .01, \hat{\mu})$
10	56	331	0.037428	325	0.008192
10	57	344	0.047823	336	0.006804
10	58	355	0.041691	347	0.005508
10	59	368	0.049045	360	0.007899
10	60	379	0.040259	371	0.005628
10	61	392	0.044019	384	0.006976
10	62	405	0.047333	397	0.008261
10	63	418	0.049009	410	0.009223
10	64	431	0.048278	423	0.009306
10	65	444	0.046233	436	0.008846
10	66	457	0.043729	449	0.008320
10	67	470	0.040167	462	0.007363
10	68	485	0.047557	475	0.005877
10	69	498	0.041125	490	0.008240
10	70	513	0.045838	503	0.005968
10	71	528	0.049258	518	0.007416
10	72	541	0.039487	533	0.008347
10	73	556	0.040092	548	0.008866
10	74	571	0.040417	563	0.009227
10	75	586	0.039726	578	0.009146
10	76	603	0.048757	593	0.008370
10	77	618	0.045266	608	0.007283
10	78	633	0.040099	623	0.006204
10	79	650	0.046157	640	0.007999
10	80	665	0.039096	655	0.006210
10	81	682	0.042558	672	0.007300
10	82	699	0.044335	689	0.008370
10	83	716	0.044963	706	0.009036
10	84	733	0.045174	723	0.009113
10	85	750	0.044334	740	0.008829
10	86	767	0.041834	757	0.008352
10	87	786	0.049317	774	0.007598
10	88	803	0.044749	793	0.009837
10	89	820	0.039298	810	0.008335
10	90	839	0.043501	829	0.009860
10	91	858	0.046751	846	0.007678
10	92	877	0.048193	865	0.008359
10	93	896	0.048867	884	0.008804
10	94	915	0.049032	903	0.009076
10	95	934	0.048047	922	0.008960
10	96	953	0.045816	941	0.008404
10	97	972	0.042650	960	0.007538
10	98	993	0.048420	981	0.009765
10	99	1012	0.043401	1000	0.008166
10	100	1033	0.047134	1021	0.009694

Chapter 3

Jitter Methods

3.1 Introduction: The Temporal Structure of the Neural Code

The rate-coding versus temporal-coding debate has been studied by innumerable experts from the whole range of backgrounds which comprise modern investigations of neuroscience. This is reflected in the plethora of definitions and methods for the analysis of temporal structure in spike trains. Indeed, as Dayan & Abbot point out in their textbook, the issue tends to "dominate" [13] discussions of the neural code. The central question is whether *spike timing matters*: that is, whether the precise temporal position of spike times matters in the functioning of the brain. Most thinkers would agree, at least when pressed, that such distinctions in the nature of the neural code lie along a continuum. A first difficulty is defining "precision" and "matters" in a way that are relevant to one another. A second (and clearly more profound) difficulty is relating definitions to physiological experiments. There is a great controversy as to whether progress has been made on any of these fronts.

One method of investigation involves presenting a stimulus repeatedly to a subject, measuring the neural response over many such presentations, and examining the resulting empirical rate: the empirical frequency of firing as a function of time, relative to stimulus onset (or for time-varying stimuli, relative to a fixed time in the stimulus presentation). If the neuron was coding for the identity of stimulus and environmental context precisely in its spike positions, one would expect that the rate would appear to be changing very rapidly: spikes are laid down with great precision. This is, for example, referred to as temporal coding in [13]. If, on the other hand, the spikes are not laid down precisely, one would expect the empirical rate function to be slow-varying, or flatter; this would be consistent with rate-coding. Clearly, the central problem raised in the introduction, that of assuming *repeatability* across trials in such experiments is a danger here. Again, it is certainly possible that the neuron signals, in whole or in part, something that varies across trials (such as attention, or the states of other neurons), but that it signals them with high precision: across trials, the empirical rate could well appear flat, despite the ostensible temporal coding. To make an analogy with hypothesis testing: the failure to reject the null hypothesis (in this case, of *coarse* temporal structure) has to be interpreted with care; it is not necessarily evidence *for* the null hypothesis. Interpretation has to do, of course, with the power of a test. Thus it is an interesting problem to discover means of identifying temporal structure in the neural response even in cases where the empirical rate function, averaged across trials, varies slowly.

A particularly elegant way of beginning to conceptualize the temporal coding debate has been suggested in the form of a thought experiment due to Elie Bienenstock, which we take as a guiding principle in the investigations described here. The experiment is

as follows. Suppose we had a magical pharmacological agent that we could distribute throughout the brain that had the effect of randomly *jittering* the individual spikes on the order of k milliseconds. That is, wherever spikes would have occurred naturally (without the drug), *with the drug* they now appear randomly, uniformly distributed, say, in a k-millisecond interval around their "naturally occurring" position. Now, the question is: how big does k have to be before one ceases to think? If k is indeed very small, on the order of microseconds for instance, then few would suppose brain functioning would suffer. On the other hand, if k is on the order of seconds or hundreds of milliseconds then the rates, in the conventional sense, would clearly be disrupted, and by this thinking brain activity would correspondingly be severely affected. More moderate values of k, on the order of milliseconds or tens of milliseconds, are where the controvery lies: this is the region of temporal coding.

Bienenstock's thought experiment does suggest a preliminary *operational* definition of the temporal scale of neural activity,[1] but there is, however, and unfortunately, no apparent way to conduct this experiment in practice. It does, on the other hand, lead to a profitable way of thinking about statistical approaches to studying the spike train similar in purpose to the approach sketched above (involving the empirical rate function), but more flexible in the alternatives which it gives power to.

For example, the thought experiment certainly suggests a data-analytic method : randomly and independently perturb (or *jitter*) each spike in a recorded train, using perturbations on the order of L milliseconds, and compute the value of some statistic on the jittered train.[2] Roughly, jittering the train should have the effect of preserving the coarse temporal structures of the trains (or those structures coarser than L milliseconds, such as, for example, spike counts measured in L millisecond intervals.) This leads to the following strategy for assessing the existence of fine temporal structure: repeating the jittering process many times, one can tabulate a *distribution* on the value of the statistic for jittered versions of the spike train, and then compare the value of the statistic on the original, *unjittered* spike train. If the original statistic is then atypical with respect to the jittered trains (for example, if it lies in a *tail* of the tabulated distribution), then one is led to suspect that the presence of *fine temporal structure* is necessary to account for the observed spike train. Specifically, one concludes that the data is incompatible with temporal structure that is (L-millisecond) coarse.

We develop the statistical meaning of this approach, *the jitter method* [12], in what follows.

3.2 The Jitter Method

3.2.1 Method

We will be working with *discretized* spike trains (with typical discretizations on the millisecond scale), consisting of T bins. Define

$$s_j(t) := \begin{cases} 1, \text{ if there is a spike in bin } t \text{ of the spike train of neuron } j \\ 0, \text{ otherwise.} \end{cases}$$

We will partitition each of the L spike trains into N disjoint windows with a fixed length of J bins each. With the windows numbered sequentially from 1 to N, if we label the set

[1] Preliminary, at the least because it certainly ignores the secondary effects of local temporal structures in the spike trains (for example, jittering the spike trains literally could lead to violations of the absolute refractory period), which we will address later.

[2] See also [44], Figure 3.20, for a similar application of this idea.

of bins associated with window k as W_k, we further define

$$m_l(k) := \sum_{t \in W_k} s_l(t),$$

i.e., $m_l(k)$ is the number of spikes in window k, for neuron l. The family of random variables containing all the window spike counts we call $\mathcal{M}$,

$$\mathcal{M} := \{m_1(1), m_1(2), ..., m_1(N), ..., m_L(1), m_L(2), ..., m_L(N)\}.$$

The algorithm proceeds as follows:

- For each original spike train s_i, generate M surrogate spike trains $s_i^{(1)}, s_i^{(2)}, ..., s_i^{(M)}$ by sampling M times, independently and uniformly from

$$\left\{ x(t) \in \{0,1\}, 0 \leq t \leq T \mid \sum_{t \in W_k} x(t) = m_i(k), \forall\, 1 \leq k \leq N \right\},$$

  i.e, from the space of spike trains which have the same total number of spikes in each window as the original train. Note that this is not *literally* jittering the train, but a natural analogue of it.

- Given any function $f(s_1, ..., s_L)$ which is a function of L discrete spike trains, compute the statistic

$$Y = f(s_1, ..., s_L),$$

  for the original spike trains as well as

$$Y_i = f\left(s_1^{(i)}, ..., s_L^{(i)}\right),$$

  for the "surrogate" spike trains.

- From $Y_1, Y_2, ..., Y_M$, form the *order statistics* $Y_{(1)}, Y_{(2)}, ..., Y_{(M)}$. (i.e., where $Y_{(i)}$ is the i'th ordered value of $Y_1, Y_2, ..., Y_M$, ranked from lowest to highest). Then reject H_0 of (window length) J if

$$\{Y > Y_{(\theta)}\},$$

  where $\theta := \lceil (1-\alpha)(M+1) \rceil$. The parameter α will turn out to be the *level* of a hypothesis test, under a suitably formulated null hypothesis, which we describe below. In practice, one can just compute the p-value of this test. In this case the p-value is

$$p = \frac{M - k + 1}{M + 1}, \tag{3.1}$$

  where

$$k := \min\{j : Y > Y_{(j)}\}. \tag{3.2}$$

  The data can reject the null hypothesis at significance $1 - \alpha$ if $p < \alpha$.

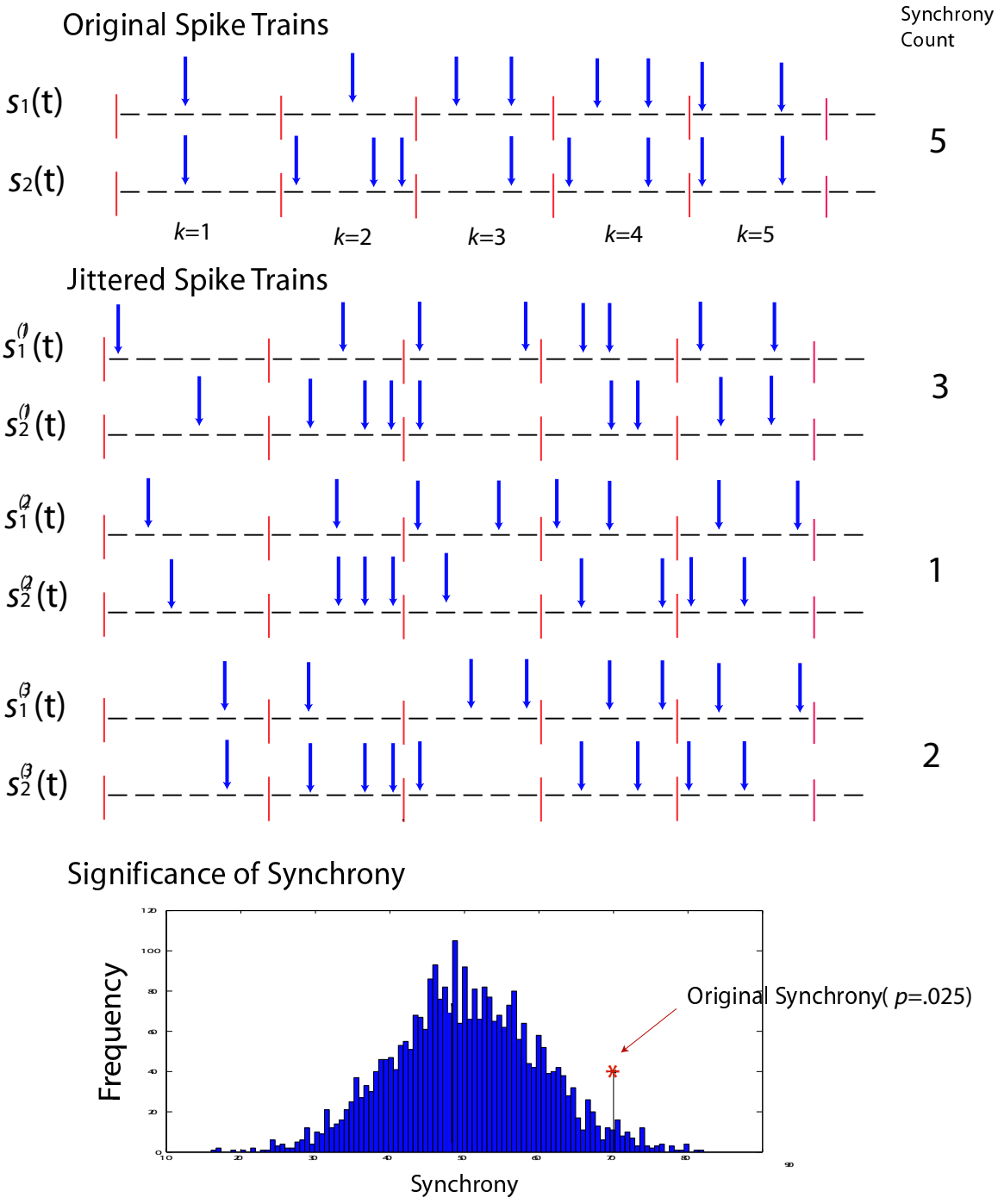

Figure 3.1: A pictorial description of the jitter method, using the number of synchronous co-occurrences of spikes as the jitter statistic f. Here windows of length J=5 bins are used. 5 windows are shown. We have the counts $\mathcal{M}$ in the original spike trains: $m_1(1) = 1, m_1(2) = 1, m_1(3) = 2, m_1(4) = 2, m_1(5) = 2$ and $m_2(1) = 1, m_2(2) = 3, m_2(3) = 1, m_2(4) = 2, m_2(5) = 2$. The randomly sampled surrogate spike trains preserve the counts $\mathcal{M}$. Finally, a distribution of the synchrony counts is tabulated (here presuming many more windows which are not shown), and the significance of the synchrony from the original distribution assessed from the tabulated distribution.

3.2.2 Statistical Interpretation

A *Bernoulli process* is a discrete-time stochastic process composed of a sequence of independent Bernoulli random variables. That is, a discrete-time stochastic process $X = (X_1, X_2, ..., X_T)$ on $1 \leq t \leq T$ is an inhomogeneous Bernoulli process with rate function $r(t) = (r_1, r_2, ..., r_T)$ if

$$P(X_1 = x_1, X_2 = x_2, ..., X_T = x_T) = \prod_{i=1}^{T} r_t^{x_t}(1 - r_t)^{(1-x_t)} \qquad \forall x_1, x_2, ..., x_T \in \{0, 1\}^T. \tag{3.3}$$

The model is natural, and one can obtain the Poisson process by taking suitable limits of finer and finer discretizations of Bernoulli processes. When the rate function $r(t)$ is a constant ($r(t) = r$), we call this a *homogeneous Bernoulli process.*

Null Hypothesis. The jitter method described above tests the null hypothesis, H_0, that, *conditioned* on the rate functions $r_1(t), r_2(t), ..., r_L(t)$ for L neurons, the discrete spike trains are samples from independent Bernoulli processes that are *homogeneous* (i.e., of constant rate) *within each window*:

$$\begin{aligned} H_0 &: P(s_1(1) = x_1(1)..., s_1(T) = x_1(T), ..., s_L(1) = x_L(1)..., s_L(T) = x_L(T)) \\ &= \prod_{l=1}^{L} \prod_{t=1}^{T} r_l(t)^{x_l(t)}(1 - r_l(t))^{1-x_l(i)} \\ &\text{where } x_l(t) \in \{0, 1\} \, \forall \, t, l \\ &\text{and } p_l(t) = p_l(t') \text{ if } t, t' \in W_k \text{ (for some } k), \, \forall 1 \leq l \leq L \end{aligned}$$

This includes the special case where the spike trains are themselves inhomogeneous Bernoulli processes with constant rates in every window. But because the assumption of independence is conditional, it also includes, for example, mixtures of Bernoulli processes in which the rate functions themselves are random (and possibly dependent), but are chosen randomly from an ensemble of rate functions which all share the property of constancy within the windows. Thus the null hypothesis evades the assumption of statistical repeatability, while still capturing the notion of *coarse* temporal structure with the homogeneous windows.

Now we demonstrate the validity of the test, beginning with a preliminary lemma.

Lemma 3.2.1. *If $X_0, X_1, ..., X_M$ are independent and identically-distributed random variables then*

$$P\left(X_0 > X_{(k)}\right) \leq \frac{M - k + 1}{M + 1},$$

where $X_{(k)}$ is the k'th order statistic of the random variables $X_1, ..., X_M$.

Remark. A straightforward way to understand this lemma is to consider the case where X_1 is a continuous random variable. Then considering the space Π of permutations of $\{0, 1, ..., M\}$,

$$\sum_{\pi \in \Pi} P(X_{\pi_0} > X_{\pi_1} > ... > X_{\pi_M}) = 1,$$

since ties occur with zero probability. Furthermore, given any $\pi, \pi' \in \Pi$, symmetry alone implies that

$$P(X_{\pi_0} > X_{\pi_1} > ... > X_{\pi_M}) = P(X_{\pi'_0} > X_{\pi'_1} > ... > X_{\pi'_M}).$$

From there, it is a transparent counting argument to

$$P\left(X_0 > X_{(k)}\right) = \frac{M-k+1}{M+1},$$

a stronger result for the continuous case. Generalizing this argument for non-continuous random variables introduces logistical details associated with the occurrence of ties, which we defer to the appendix.

Corollary 3.2.2. *If $P \in H_0$, then*

$$P(\{reject\, H_0\}) := P(\{Y > Y_{(\theta)}\}) \leq \alpha$$

Proof. By virtue of the null hypothesis, $Y_0, Y_1, Y_2, ..., Y_M$ are *conditionally* independent and identically-distributed, conditioned on $\mathcal{M}$. Hence an application of Lemma 3.2.1 gives

$$\begin{aligned} P(Y_0 > Y_{(k)}) &= \sum_{\mathcal{M}} P\left(Y_0 > Y_{(k)} | \mathcal{M}\right) P(\mathcal{M}) \\ &\leq \sum_{\mathcal{M}} \left(\frac{M-k+1}{M+1}\right) P(\mathcal{M}) \\ &= \frac{M-k+1}{M+1}. \end{aligned} \tag{3.4}$$

Plugging in $k = \theta := \lceil (1-\alpha)(M+1) \rceil$ into (3.4) yields the corollary. In addition, the inequality (3.4) justifies the p-value defined by (3.1).

3.2.3 Interpreting the Null Hypothesis

The idea of the jitter method, and its null hypothesis, is to explore ways of identifying finer temporal structure in the spike train (i.e., fast changes in rates), in cases where it is not compelling to assume *repeatability* in the spike-generating process, for our usual reasons. As we state in the introduction, if it were reasonable to assume repeatability, the most natural inference technique would be to measure the empirical rate function across trials, and then observe how quickly the rate function varies. This is inference via estimation. What about when the process is not repeating? An alternative way to define coarse-temporal structure statistically is to assume the spike counts in intervals are a *sufficient statistic*. A statistic T is *sufficient* with respect to a class of distributions $\mathcal{C}$ when the distribution of the data, conditioned on the sufficient statistic, is the same for all distributions in $\mathcal{C}$: i.e.,

$$P_1(X|T(X)) = P_2(X|T(X)) \, \forall P_1, P_2 \in \mathcal{C}. \tag{3.5}$$

In such a situation, the value of $T(X)$ alone is therefore sufficient to make inferences about the distribution governing X, with respect to the class $\mathcal{C}$. For our null hypothesis H_0 the class of processes such that *conditioned on the rate*, the process is an inhomogeneous Bernoulli process with constant rates (within pre-specified windows), the data is distributed identically (in fact, uniformly) for all processes, given $\mathcal{M}$. Thus $\mathcal{M}$ forms a sufficient statistic for the null hypothesis, and this is a statistical analogue of saying *'only the counts matter'*.

The symmetry implied by a sufficient statistic for a class of distributions $\mathcal{C}$ has practical implications for testing the null hypothesis $H_0 = \mathcal{C}$ [49, 31]. This is useful because in general the problem of testing *composite hypotheses*, in which the null hypothesis contains many distributions, is difficult. Sufficient statistics reduce an otherwise composite hypothesis testing problem to a simple one, in the following way. Given a sufficient statistic, one can seek a critical region $f(T(X))$ for each value of $T(X)$ such that

$$P(X \in f(T(X))|T(X)) \leq \alpha \quad \forall P \in H_0 \tag{3.6}$$

By virtue of *sufficiency*, for a given value of $T(X)$, this is a problem of a *simple hypothesis*, since the conditional distribution of the data given $T(X)$ is the same for all distributions in the null hypothesis. The function f then gives rise to a *conditional test*, to test the general hypothesis H_0. That is, the event $\{X \in f(T(X))\}$ is then an α-level hypothesis test:

$$P(X \in f(T(X))\,) = \sum_{T(X)} P(X \in f(T(X))|T(X)\,)P(T(X)) \leq \sum_{T(X)} \alpha P(T(X)) = \alpha. \quad (3.7)$$

This is the mathematical essence underlying the data-analytic intuition of the jittering procedure. $\mathcal{M}$, the counts in the windows, forms a sufficient statistic of the null hypothesis, and jittering is akin to Monte Carlo estimation of the conditional distribution $P(\cdot|\mathcal{M})$.

As a toy example, consider an 18-bin model of a short spike train. Let us model the spike train as a pattern of two spikes, spaced 5 bins apart, with its location uniformly distributed on the interval. Now we will place 2 5-bin windows in the center 10 bins of the 18 bins (we take this slightly cumbersome route of embedding the windows within a larger model to avoid edge effects, which would become asymptotically negligible as we lengthen the model). If you were to sample T times from this model (corresponding to T trials), the empirical rate function on the 10 bins would converge to a constant as T increases. On the other hand, this clearly seems like an example of temporal structure: the precise location of one spike predicts (in fact, determines) the precise location of the other. Now, if we employ as our jitter statistic the following function

$$f(s) := \begin{cases} 1, & \text{if there are two spikes in the train 5 ms apart} \\ 0, & \text{otherwise.} \end{cases} \quad (3.8)$$

Then clearly the jitter method will quickly lead to reliable rejection of the null hypothesis, whereas using the empirical rate function would not.

As this example illustrates, however, clearly the choice of f is crucial for the rejection. Some choices of the function f would not lead to rejection. For example, it is not difficult to construct an f which essentially mimics the procedure of measuring the rate of variation of the empirical firing frequency; thus jittering can be seen as a more general procedure. To do this, let $r_i = \frac{1}{T}\sum_{j=1}^{T} s_j(i)$, the average number of spikes in bin i across trials, where we index i along the 10-bin window, $1 \leq i \leq 10$. Thus $r_1, r_2, ..., r_{10}$ is the empirical rate function. Then employ as f a measure of the variation across r. For example

$$f(s) = \frac{1}{10}\sum_{i=1}^{10} (r_i - \bar{r})^2\,, \quad (3.9)$$

where $\bar{r} = \frac{1}{10}\sum_{i=1}^{10} r_i$, i.e., here f measures mean square variation. Because of the law of large numbers, for sufficiently large T, $f(s)$ will be near zero. On the other hand, *jittered* versions of the train will also be near zero, since the jittered trains themselves will *also* have flat empirical rate functions. The empirical rate function does not reveal the temporal structure of this model.

In this regard, the *failure* to reject the null hypothesis via the jitter method, as usual, is not necessarily evidence *for* the null hypothesis: it may be that an alternative choice of the jitter statistic f would lead to a rejection. Since the choice of candidate functions f is virtually unbounded, judicious choices for the function f is an important issue. This can be seen as a virtue: since any f can be used with essentially no change in the method, it is possible, by suitable choice of f, to adapt the method to produce tests of the coarse temporal structure hypothesis which are well-adapted to alternatives of scientific interest.[3] Furthermore,

[3] In this respect, the method shares a theme with other modern developments in statistics, such as the bootstrap [15], which take advantage of computing methods to widen the scope of available inferences.

the form of the function f can then be suggestive, in cases where the null hypothesis is rejected, of the types of alternatives which could have generated the data.

For one final point of clarification, we return to the statement of the null hypothesis: *conditioned on their rate functions, the discretized spike trains are Bernoulli processes that are homogeneous within each window.* This can be a little confusing. It is clear from the statement that Bernoulli processes with window-constant rate functions are in the null hypothesis. It is also clear, that mixtures of Bernoulli processes drawn from random but window-constant rate functions are also in the null hypothesis. However, rate functions are not literally *observed*, hence the possible confusion. Taken literally, the statement has to do with *representations* of discrete point processes, which takes us back to the notion to the universality of the Bernoulli representation.

That is, to repeat, *every collection of discrete-time stochastic point processes is a collection of mixtures of Bernoulli processes* (refer to the discussion leading to Equation 1.6). Therefore, one can represent arbitrary discrete point processes with a hierarchical model in which the spike train is a random sample from a random rate function. However, such a representation is not necessarily unique. One can make this point with a model of the simplest such process: a two bin spike train, represented by $s(1)$ and $s(2)$. Such a model is specified by three probabilities p_{10}, p_{01}, p_{11}:

$$\begin{aligned} P(s(1)=1, s(2)=0) &= p_{10} \\ P(s(1)=0, s(2)=1) &= p_{01} \\ P(s(1)=1, s(2)=1) &= p_{11} \\ P(s(1)=0, s(2)=0) &= 1 - p_{10} - p_{01} - p_{11} \end{aligned} \tag{3.10}$$

Now, let us suppose that $p_{00} = .25, p_{11} = .25, p_{01} = .25, p_{11} = .25$. This is, of course, a Bernoulli process with rate $\frac{1}{2}$, and therefore is a candidate for the null hypothesis (when placed in a window, et cetera). However, one could in principle *represent* it with a mixture-Bernoulli model with non-constant rate functions. This is essentially the exercise that gives rise to (1.6): Just use the following four rate functions:

$$\begin{aligned} r_a &= (0,0) \\ r_b &= (1,0) \\ r_c &= (0,1) \\ r_d &= (1,1), \end{aligned} \tag{3.11}$$

and mix them with probability .25. Then we have the same process, but the rate functions r_b and r_c are, of course, not constant. This example extends to arbitrarily long time domains. Therefore to be (perhaps overly) explicit, the null hypothesis H_0 contains discrete point processes which merely *can* be represented as mixture-Bernoulli processes with window-constant rates.

3.3 Results

A statistic f can be drawn from any of the virtually unbounded class of functions of multidimensional spike processes, and by virtue of Corollary 3.2.2, the jitter method will test a null hypothesis of 'coarse temporal structure'. In our case, one would like to choose a statistic that has *power* under a presumed alternative distribution: $P(\{\text{reject } H_0\})$ should be high for an alternative distribution P of interest (i.e., a distribution, for example, that is plausible under a neuroscientific theory of interest, or which could be generated by a plausible neural mechanism). Such a choice of f is most likely to reveal temporal structure insomuch as the assumptions giving rise to its use are true.

The jitter method has been applied to two different problems involving multi-neuronal cortical recordings. The original application was to assessing the significance of synfire chains, the second application was to the problem of identifying significant synchrony.

3.3.1 Synfire Chains

The jitter method was first introduced in [12], where it was applied in the context of synfire chains. Synfire chains are the repeated, temporally-precise sequential activation of neuronal subpopulations which have been proposed to occur in cortex [1, 2]. To determine whether synfire chains are detectable in neurophysiological records, it is necessary to rule out the hypothesis that the observable temporally-precise repetitions of neuronal firing could have arisen in the absence of temporal structure. Hence the jitter method was developed.

The statistic f employed was the maximum number of repetitions of spatiotemporal patterns of a given *complexity*. For example, a pattern of complexity 4 is: neuron 1 firing, then 15 milliseconds later neuron 7 firing, then 3 milliseconds later neuron 3 firing, and then neuron 1 firing after another 10 milliseconds. Fixing an upper bound w, on the total time it takes for a pattern of complexity k to occur, there are a finite number of possible k-complexity patterns in a set of spike trains in discrete time. For each such pattern, one can count the number of repetitions of the pattern in the neuronal records, and then optimize over patterns to produce f, the maximum number of repetitions of spatiotemporal patterns of complexity k.

The method was applied to simultaneous recordings of five neurons obtained from primate Supplemental Motor Area (SMA) in monkeys performing reaching tasks. Details on the experiments, as well as methods for efficiently computing f can be found in [12]. The null hypothesis of window size J=6 milliseconds was rejected at a significance level of 97.5%, and the more restrictive null hypothesis of J=15 was rejected at significance level of 99.5%.

3.3.2 Synchrony

In [27], the jitter method is applied to the question of whether the synchronous firings observed in a pair of cortical spike trains are in excess of what one would expect from a rate hypothesis. The spike trains were discretized into bins of 1-2 milliseconds. One train from the pair was (arbitrarily) designated as the reference train, and windows (with lengths ranging from 2 to 8 milliseconds) were centered around the spikes in the reference train. These windows were then used as the jittering windows for spikes in the target train. Synchrony, defined as the number of target neuron spikes falling in the center of the reference windows was used as the jittering statistic f, and the jitter method was used to test the hypothesis that, *conditioned* on the reference train, the spikes in the target train were a Bernoulli process with constant rates in the reference-anchored windows.

The pairwise test was applied to a total of 224 cell pairs obtained from several simultaneous recording sessions from M1 of macaque primates performing reaching tasks. In order to accommodate multiple testing, the rejection or failure of rejection of each pairwise synchrony test was itself treated as a statistic, and the binomial test was used to assess the general (but not quite appropriate, due to pairwise redundancy) null hypothesis that each cell pair was a pair of independent processes drawn from the null hypothesis of the pairwise jitter synchrony test. The null hypothesis was rejected ($p < 0.01$) for all three of the testing parameters: { 1 ms discretization, 2 ms windows }, {2 ms discretization, 4 ms windows}, and {2 ms discretization, 8 ms windows}.

3.4 Discussion

We have proposed a technique for defining and inferring the *time scale* of the brain, from observations of temporal structure in neural spike trains. The definition is the *null hypothesis*: that the spike train is a Bernoulli process whose rate function is *constant* in windows of length L. The technique is the jitter method, a hypothesis test, which provides a method for calculating the *significance* of a chosen statistic under the null hypothesis, and which we propose to use as a basis for inference. The choice of statistic is, of course, crucial, and depends on scientific criteria. The utility of a hypothesis test depends on the ability of a statistic to *distinguish* (via its distribution) the null hypothesis from an alternative distribution *of interest*. This is classically the issue of the *power* of a test. The choice, or design, of a statistic makes any hypothesis test just as much about the null hypothesis, as about the *alternative* hypotheses that the test is optimal and hence targeted towards.

The size of the window J is a proxy for the 'time scale' of nervous system activity. Roughly, the null hypothesis expresses the idea that the spike processes are ultimately random when viewed at a resolution finer than $\{J \times \text{bin size}\}$. Thus rejecting such a null makes a statement about the temporal resolution of nervous system activity. This can be viewed as well from the perspective of sufficient statistics: the size of the window is the resolution at which "the spike counts matter". There is a degree of artifice in this approach, however. The essence of coarse temporal structure is not necessarily *constancy* but really *slow variation* in the likelihood of spiking across time. In this sense the null hypothesis merely offers an approximation to the notion of coarse temporal structure in terms of slow variation in rates: roughly, a process is in the null hypothesis when it is well-approximated by a process that has constant rates in windows of size J. Furthermore, the pre-selection of window locations is arbitrary.

This approximation introduces a problem in the form of an interaction between the quality of approximation (i.e., the degree of slow variation) and the amount of data available to the test. Given enough data, a hypothesis test could in principle reject an arbitrarily small deviation from the null hypothesis. An example is the elementary problem of testing the hypothesis $p \leq .5$ for a Bernoulli random variable with probability p. Via the law of large numbers, we know that, given enough samples, a binomial test for $p \leq .5$ will reject the hypothesis for samples with $p = .5 + \epsilon$, for *any* $\epsilon > 0$. Because of this confound between the amount of deviation from the null and the amount of data, many statisticians eschew such *sharp* hypothesis tests in favor of confidence intervals. The same sort of problem arises for permutation tests such as the jitter method. Since we are often dealing with great amounts of data in the form of long spike trains, this concern warrants serious consideration.

Furthermore, the mere *existence* of fine temporal structure in the spiking processes can already be taken for granted on the grounds of biophysical knowledge. Due to intrinsic membrane properties involving the recovery of neuronal polarization following an action potential, there is a period of time of duration approximately 1 to 3 milliseconds following a spike, the *absolute refractory period*, during which the neuron is incapable of spiking at all. Furthermore, there is in addition a *relative refractory period* of longer extension, a consequence of the same recovery, in which more synaptic input is in general required to fire a spike than at equilibrium, and which is likely to introduce a local dependence among the spikes. In addition some cortical neurons possess specialized ionic channels which cause them to regularly fire not single spikes, but short, finely-timed *bursts* of spikes in response to sufficient input. All of these phenomena are stereotyped forms of *fine temporal structure* which are local, but perhaps not as interesting from the perspective of the timing of the neural code. Nevertheless, they do represent deviations from the null hypothesis which could increase the likelihood of rejection by jittering.

This line of thought leads us to seek a generalization of the jitter null hypothesis: non-repeating processes in which the rate of change of the underlying rate functions are *parametrically* bounded (in analogy to the approach of confidence intervals), and a method in which rejection implies temporal structure over and above *local* fine temporal structure of the sort attributable to refractory periods and bursts.

However, it is not clear how to solve these two problems generically in the framework of the jitter method. The goal of the remaining development of these ideas is to preserve the themes of the jitter method (particularly, with respect to the issues of repeatability), while taking into account these general problems, in the context of the specific alternatives of processes exhibiting synchronous spiking.

3.5 Related Work

Temporal and rate coding is the subject of a vast literature. The observation that *rates* of spiking are correlated with behavioral and environmental variables occurred as early as the pioneering work of Adrian [4]. Since then, it has played the role of a dominant paradigm, and similar sorts of correlations have been observed in nearly every system and preparation investigated in behavioral neurophysiology. It also underlies the thinking of much of neural network research in which the signals sent from one model neuron to another are presumed to be continuous variables. Notions of temporal coding tend to be of more recent vintage, though temporal coding has long been acknowledged to exist in early sensory systems where the neural responses tend to be very direct reflections of sensory input, which themselves contain precise temporal structures [55]. Commonly-cited and striking examples are often drawn from the auditory system where the detection and communication of very precise temporal events are built into the problem which the brain has to solve, for example the problem of detecting the very fine differences between sounds detected at the two ears for the purpose of computing the location of a sound-emitting source in space. In such systems neurons which detect coincidences for functional purposes on millisecond time scales are routinely found. Deeper in the brain, particularly in cortex, the subject of temporal codes is controversial, and it is hard to paint a clear picture of the schools of thought in the field. Perhaps the central question is whether temporal codes signal more abstract, or processed, events than raw sensory signals contain. In this light, the work of Laurent and colleagues [30, 53] in insect olfaction has probably been the most compelling (some might say, the only) case for temporal structures in coding, but the issue is fraught with so many methodological difficulties that it is hard to interpret the lack of evidence as a negative conclusion.

Some of the early work involving the recurrence of temporally-precise statistical structures arose from Abeles' thinking on synfire chains [1]. This was the motivation for the original proposals of the jitter method. Methods for assessing the significance of repeated observations of spatiotemporal patterns with respect to null models of chance are discussed in [3], and further applications can be found in [2, 43, 56]. Oram and colleagues [38] argued that assumptions of repeatability and the Poisson assumptions in these methods confounded the conclusions, and suggested that more intricate models threw a negative light on conclusions from these studies. The fitting process in these methods are a little involved however, and, though their critique is certainly valid, it can be a little difficult to identify the null hypothesis in their proposals, and how restrictive it is. Other approaches to this problem includes the papers [18, 19], where the authors consider the significance of the number of recurrences of a *particular* spatiotemporal pattern of neural firing, using assumptions of a similar flavor.

The case of synchrony is the center of a great deal of work on temporal coding. One clear line of thought begins with the theories of von der Malsburg [59, 60], which proposed

synchronous co-occurrence among neurons as a way to solve the *binding* problem, the problem of representing the *relationships* among the features present in a stimulus, in neural networks which employ distributed representations [16]. Experimental results consistent with these ideas were presented in [21]: here the responses of two distant V1 cells (from an anesthetized cat) whose receptive fields had similar orientation selectively but were spatially distinct exhibited synchronous co-activity when a coherent line of light which crossed the receptive fields of both cells was presented, and not when the cells were activated by two independent lines that were identical in the receptive fields proper. Moreover, the amount of activity in both situations was similar, indicating that the effect was independent of rate coding. These experiments spurred a flurry of subsequent work (see [51] for a review), and in turn highlighted a need for analytical methods which could precisely detect the the occurrence of synchronous activity to guide experimental work. As Wolf Singer concluded in a review article [50] on this topic in 1994: "As temporal relations such as synchrony can only be evaluated by simultaneously recording the activity of different cells, critical tests require the application of multielectrode recordings, and it will most likely also be necessary to develop new analytical methods to reliably detect transient temporal relations among the activities of widely distributed groups of neurons. For the brain, a synchronous discharge of several thousand cells is likely to be a highly significant event – even if it occurs only once – but for the experimenter, such episodes may pass undetected as long as only the responses of a few neurons can be examined at a time."

Methods for analyzing cross-correlations between two spike trains were introduced in [41]. A basic idea is to correct for the effects of the stimulus-induced rate on each spike train, estimated by averaging across trials, separately in order to predict its effect on the joint behavior of the two spike trains. Suppose we represent the response of two neurons after repeated presentation with a stimulus n times, $X^i(t)$ and $Y^i(t)$, so X^i represents the response of neuron X during trial i, and Y^i represents the response of the neuron Y during trial i. In [41], a permutation test is suggested: randomly permute the responses $Y^1, Y^2, ..., Y^n$, and compute the cross-correlation between the responses $X^1, ..., X^n$ and the permuted Y responses. If the permuted cross-correlation differs significantly from the original cross-correlation between X and Y, then the difference can be attributed to interactions *within* trials, independently of the stimulus-dependent responses. Implicitly, this tests the null hypothesis that $Y_1, ..., Y_n$ are exchangeable, hence this is an assumption of repeatability. Rather than use the permutation idea directly, Perkel applied assumptions of independence, within and across the trains, to assess the significance of the difference between the original and the permuted records. This is a rate-coding type of assumption: the stimulus determines a rate, and hence conditioned on the stimulus (or equivalently, the rate), the neurons act independently, and hence their spikes are as well. Formally similar tests operating under the same hypotheses are presented in [39] and [5].

Brody [9, 10] suggested these methods were problematic from the perspective of the fine temporal structure debate, by pointing out some models that were alternative to such a null hypothesis but nevertheless shared the spirit of the rate coding hypothesis: so-called latency and excitability covariations. The models he described were mixture-of-Poisson style, in which each neuron has a stimulus-determined rate function but across trials the rate function is shifted in time by a trial-varying latency parameter, and scaled by a trial-varying gain parameter. However, the trial-varying latency and gain parameters are common to both neurons (hence latency and excitability *covariations*). In such a case, the latency and gain covariations can lead to the detection of significant synchrony, using these methods, under the straight Poisson null hypothesis.

Grün et al. [22] and [23] develop methods for detecting coincident activity among many cells in multi-neuronal records. The central assumption of the null hypothesis in [22] is that

the spiking process is a homogeneous Bernoulli process for every spike train, and that all spike trains are independent. Under this assumption, the rate can be estimated from the frequency of the spike counts, and the distribution of the number of occurrences of coincident activity of any particular subassembly can be computed. The computed distribution can then be used to evaluate the significance of coincident activity from the multi-neuronal records. In [23], the problem of nonstationarity in the spike trains is addressed by using a sliding window and assuming only that the process is homogeneous Bernoulli in the smaller sliding window: this roughly amounts to approximating nonstationary processes as 'locally' stationary. Gütig [25] invests this framework with a permutation test. The null hypothesis also is that the neurons consist of independent, homogeneous Bernoulli processes, but rather than estimating the Bernoulli rates (as in [22],[23]), one computes the probability of coincidence counts *conditioned* on the spike counts in both trains, in order to evaluate the significance of the observed coincidence count. This conditional distribution is independent of the rate, under the Poisson model, and as in the jitter method. This is very much akin to the jitter method, specialized for the case of synchrony in which Monte Carlo estimation is not necessary, but with large jittering windows. They show that the permutation test is uniformly more powerful than the rate-estimation method of [22] for alternatives consisting of pairs of homogeneous Poisson processes with non-zero correlation *across* the trains. Notably, regarding the problem of correlations across time, Gütig writes "The important task of overcoming the difficulties introduced by serially correlated [i.e., temporally-correlated] spike trains is the subject of ongoing research."

Another approach to identifying higher-order correlations is presented in [35]. Here log-linear models of the joint probability distribution of multiple spike trains are employed to test hypotheses about interactions among the trains, but the trains are assumed to be homogeneous in time.

In [40], the authors attempt to overcome the problems with trial-to-trial variability (i.e., the assumption of repeatability [9]) to locally estimate the firing rates from spike trains in each trial individually, with a view towards detecting significant synchrony. The idea is to work with a null hypothesis in which the spike trains are Poisson rate-governed but that the rates are not repeatable. To accomplish this, they estimate the rate function in a given trial from the interspike interval function: smoothing the interspike interval function while allowing for large discontinuities in the rate functions in periods of spike bursting. There are a number of parameters in the approach, however, and it appears difficult to clearly identify the null hypothesis.

3.6 Pattern Jitter: Asymmetric Methods

We want to generalize the jitter method in two directions. In the first, we want to capture the notion that coarse temporal structure is not so much about *constant* rates in pre-specified windows as about *slow variation* in the rates. In the second, we want to allow for local fine temporal structures (bursts, refractory periods) in the null hypothesis itself, precluding them from being a source of rejection.

To see how to approach this generalization, one can look at the jitter method from the following perspective. Its essence is the implication of the null hypothesis that, within every window, *conditioned on the counts in the window*, all configurations of spikes with that number of spikes are equally likely. As a consequence of this, under the null hypothesis, the Monte Carlo surrogate spike trains $s^{(1)}, s^{(2)}, ..., s^{(N)}$ are independent and identically distributed samples from the original distribution of the process, conditioned on the counts in all windows. As a consequence of that fact, the validity of the test is then guaranteed by Lemma 3.2.1.

We can re-cast the two problems we seek to solve in the framework of this perspective. First, the problem with local temporal structures is that it is not true that, *conditioned* on the count in a window, all configurations of spikes with that count are equally likely: the particular pattern of spikes in the window may modify its relative likelihood with respect to other patterns of spikes with the same count. The obvious example is the absolute refractory period: for a given spike count, some configurations of spikes in the windows will actually violate the absolute refractory period which would have probability zero under a valid model. Bursts and the relative refractory period would give rise to similar phenomena. Secondly, if the rate of spiking across a window was not constant, but simply slowly varying, then the relative locations of spikes conditioned on the counts would be affected by this structure as well.

A natural way to accommodate these observations is to *refine* the partition of the outcome space of spike trains from counts more finely into *patterns* of spikes. This would have the effect of preserving the local temporal structure of spikes in a window. Furthermore, we could assign as a parameter a bound on the relative likelihood of the locations of patterns, *conditioned* on their identity. The bound, by limiting how precisely patterns of spikes are placed within a window, would thus limit the temporal precision of the process, and serve as the measure of temporal structure in the process.

We formalize such a null hypothesis below, but there is no obvious way to build tests of this null hypothesis for arbitrary statistics, as we can using Monte Carlo-like methods in the standard jitter method. Rather, we develop a specialized test designed with synchrony-based statistics, which has power for finely-structured point processes that exhibit synchronous spiking. We call this method the *pattern jitter* method.

3.6.1 Null Hypothesis

This suggests the modified null hypothesis of the *pattern jitter* method. A *window-division* is a division of the discrete time domain into N windows of length J bins, with the condition that every window is separated from its neighboring windows by a S bin interval. We refer to these conditions as the *window-separation constraints,* and we define the following random variables.

Notation.

$$\ell_i := \text{the location of the first spike in window } i$$

$$A_i := \text{the pattern of spikes in window } i\text{, irrespective of their temporal location}$$

$$\mathcal{A} := (A_1, A_2, ..., A_N)$$

$$_i\ell := (\ell_1, \ell_2, ..., \ell_{i-1}, \ell_{i+1}, \ell_{i+2}, ...\ell_N)$$

Null Hypothesis. *We distinguish two discrete point processes separately as a* reference *process and a* target process*. A probability distribution P on the reference and target processes is in the null hypothesis H_0 if, conditioned on the reference process, for any window-division on the target process satisfying the window-separation constraints, the target process satisfies*

$$\frac{1}{\Delta} \leq \frac{P(\ell_i = j|_i\ell, \mathcal{A})}{P(\ell_i = j+1|_i\ell, \mathcal{A})} \leq \Delta \quad \forall 1 \leq j < j_{\max}(A_i), \forall 1 \leq i \leq N, \forall_i\ell, \forall\mathcal{A} \tag{3.12}$$

where $j_{\max}(A_i)$ is the maximal location of ℓ_i associated with A_i, (i.e., the maximal value of the locations of the first spike which keeps the pattern in the window.)

Remarks.

- The windows are separated so that locally stereotyped patterns such as bursts from one window will not interact with one another in such a way as to complicate the independence relations across windows. Hence the size of the separation parameterizes the length of 'local' fine temporal structures, such as bursts, which are permitted under the null hypothesis.

- Note that the choice of location for the windows is arbitrary, so long as the windows satisfy the (S bin) separation property. The *choice* of window locations within this constraint, might affect the *power* of a test (as described below), i.e, how it behaves under alternative distributions.

- Bernoulli processes which are of constant rate in the windows satisfy the conditions of the null hypothesis with $\Delta = 1$. Thus this null hypothesis generalizes that of the jitter method (in the direction of temporal precision).

- Δ therefore serves as a restriction on the temporal structure of the process, over and above *local* temporal structure (parameterized by the degree of separation of windows).

- Interpreting the independence assumptions can cause some confusion. In this regard, it is useful to keep in mind the idea of the original jitter method, of which this is a generalization (i.e., the jitter null hypothesis is contained here as a special case of $\Delta = 1$.) There, the idea is that the rate functions (and hence their sufficient statistic, the counts) themselves might have *any particular* dependence: across time, across trains, et cetera, *but over and above* this dependence, there is no structure in the spike times. The same idea holds in spirit here, but for patterns (which determine the counts) more specifically. Since the conditions of the hypothesis hold *conditioned* on $\mathcal{A}$, the patterns in windows across time might have any possible dependence. But, *over and above* that dependence, there is no structure in the control of spike positions apart from the structure that the parameter Δ permits.

3.6.2 Method

In the jittered trains, the reference train is preserved, and patterns in the target train are translated randomly in a weighted manner determined by Δ. The "synchrony" count here is not literally a synchrony count, but the number of windows in the target train in which there is a spike in the center bin (i.e., synchronous with the spike in the reference train that "anchors" the window). The tabulated distribution of these synchrony counts among the pattern-jittered trains is used to assess the significance of the synchrony counts in the original train with respect to the Δ null hypothesis. Actually, the pattern-jittered trains here are metaphorical: the actual distribution is computed analytically.
Algorithm.

- From the discretized *reference* spike train, identify N spikes that are separated by at least $S + 2 \cdot \lfloor \frac{J}{2} \rfloor$ bins each (in practice J is taken to be odd), by proceeding from the onset of the train forwards, and ignoring spikes that are within $S + 2 \cdot \lfloor \frac{J}{2} \rfloor$ bins from their neighboring previous spike. Here $\lfloor x \rfloor$ is the largest integer that is less than or equal to x. Then create N time windows of length J bins in the identically-aligned and discretized *target* spike train centered at the identified spikes in the reference train. Define ℓ_i and A_i as in the notations from the i'th window of the target train, i.e., where ℓ_i is the location of the first spike in the window relative to the onset of the window, and A_i characterizes the pattern of spikes in the target train.

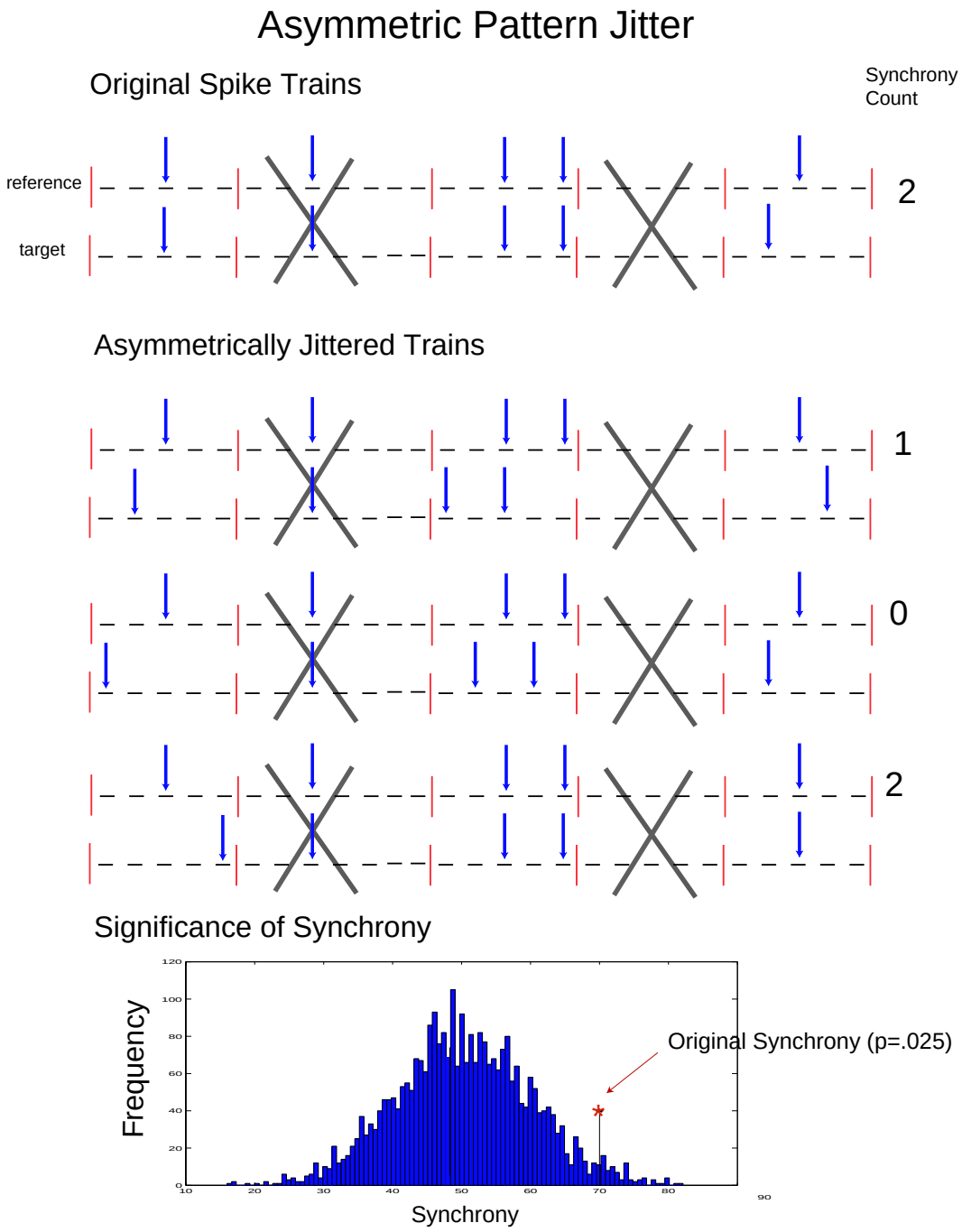

Figure 3.2: Pictorial description of the asymmetric pattern jitter. Here $J = S = 5$ bins. X's mark out regions of the spike train which are ignored on account of the S-bin separation requirement. There are 3 windows (indexed by k). In the target train we have $A_1 = \{1\}, \ell_1 = 3, A_2 = \{101\}, \ell_2 = 2, A_3 = \{1\}, \ell_3 = 1$. In the jittered trains, the reference train is preserved, and patterns in the target train are translated randomly in a weighted manner determined by Δ (i.e., preserving $\mathcal{A} = (A_1, A_2, A_3)$. The "synchrony" count here is not literally a synchrony count, but the number of windows in the target train in which there is a spike in the center bin (i.e., synchronous with the spike in the reference train that "anchors" the window). The tabulated distribution of these synchrony counts among the pattern-jittered trains is used to assess the significance of the synchrony counts in the original train with respect to the Δ null hypothesis. Actually, the pattern-jittered trains here are metaphorical: the actual distribution is computed analytically.

- Define

  $S(A_i) :=$ the values of ℓ_i for which A_i will place a spike in the center bin of window i, (3.13)

  and for each window i, compute

$$p_i^* = \max_{\substack{q_1,\ldots,q_{j_{\max}(A_i)} \\ q_i \geq 0 \, \forall i \\ \sum_i q_i = 1 \\ \frac{1}{\Delta} \leq \frac{q_i}{q_{i+1}} \leq \Delta}} \sum_{j \in S(A_i)} q_j, \tag{3.14}$$

  by linear programming [34].

- From $p_1^*, p_2^*, ..., p_N^*$, compute the (reverse) cumulative probability distribution

$$f(t) = P\left(\sum_{i=1}^{N} Y_i \geq t\right), \tag{3.15}$$

  where each Y_i is distributed as a Bernoulli random variable with parameter p_i^*, and $Y_1, Y_2, ..., Y_N$ are independent. An algorithm for efficiently computing f is described in the appendix of chapter 2.

- Count the number of windows Y in which there is a spike in the center bin of the target train, and reject H_0 of Δ if

$$\{Y \geq \inf\{t : f(t) \leq \alpha\}\}. \tag{3.16}$$

3.6.3 Validity of the Test

The following lemma is the essence of the proof of significance.

Lemma 3.6.1. *If $X_1, X_2, ..., X_n$ are Bernoulli random variables and*

$$P(X_i = 1 |_i X) \leq p_i \quad \forall i, {}_i X$$

Then

$$P\left(\sum_{i=1}^{n} X_i \geq k\right) \leq P\left(\sum_{i=1}^{n} Y_i \geq k\right) \quad \forall k,$$

where $Y_1, Y_2, ..., Y_n$ are independent Bernoulli random variables with $Y_i \sim Be(p_i)$.

Proof. The proof is by induction. The case $n = 1$ is self-evident. Accordingly, assume that the lemma holds for $n - 1$. Observe that

$$\begin{aligned} P\left(\sum_{i=1}^{n} X_i \geq k\right) =& P\left(\sum_{i=1}^{n-1} X_i \geq k \,\middle|\, X_n = 0\right)(1 - P(X_n = 1)) \\ &+ P\left(\sum_{i=1}^{n-1} X_i \geq k - 1 \,\middle|\, X_n = 1\right) P(X_n = 1), \end{aligned}$$

and, analogously,

$$\begin{aligned} P\left(\sum_{i=1}^{n} Y_i \geq k\right) =& P\left(\sum_{i=1}^{n-1} Y_i \geq k\right)(1 - p_n) \\ &+ P\left(\sum_{i=1}^{n-1} Y_i \geq k - 1\right) p_n, \end{aligned}$$

using the same decomposition but noting the independence of the Y_i's. Now, conditioning on $X_n = 0$, the hypotheses of the lemma are satisfied for $X_1, X_2, ..., X_{n-1}$ (i.e., the case $n-1$), and hence

$$P\left(\sum_{i=1}^{n-1} X_i \geq k \middle| X_n = 0\right) \leq P\left(\sum_{i=1}^{n} Y_i \geq k\right), \tag{3.17}$$

and by the same reasoning,

$$P\left(\sum_{i=1}^{n-1} X_i \geq k-1 \middle| X_n = 1\right) \leq P\left(\sum_{i=1}^{n-1} Y_i \geq k-1\right). \tag{3.18}$$

Furthermore, the set relation $\{\sum_{i=1}^{n} Y_i \geq k\} \subseteq \{\sum_{i=1}^{n} Y_i \geq k-1\}$ implies

$$P\left(\sum_{i=1}^{n-1} Y_i \geq k\right) \leq P\left(\sum_{i=1}^{n-1} Y_i \geq k-1\right) \tag{3.19}$$

Finally, we also have

$$P(X_n = 1) = \sum_{{}_iX} P(X_n = 1|{}_iX)P({}_iX) \leq \sum_{{}_iX} p_n P({}_iX) = p_n. \tag{3.20}$$

Equations (3.17)-(3.20) yield

$$\begin{aligned}
P\left(\sum_{i=1}^{n} Y_i \geq k\right) &= P\left(\sum_{i=1}^{n-1} Y_i \geq k\right)(1-p_n) \\
&\quad + P\left(\sum_{i=1}^{n-1} Y_i \geq k-1\right) p_n \\
&\geq P\left(\sum_{i=1}^{n-1} Y_i \geq k\right)(1 - P(X_n = 1)) \\
&\quad + P\left(\sum_{i=1}^{n-1} Y_i \geq k-1\right) P(X_n = 1) \\
&\geq P\left(\sum_{i=1}^{n-1} X_i \geq k \middle| X_n = 0\right)(1 - P(X_n = 1)) \\
&\quad + P\left(\sum_{i=1}^{n-1} X_i \geq k-1 \middle| X_n = 1\right) P(X_n = 1) \\
&= P\left(\sum_{i=1}^{n} X_i \geq k\right),
\end{aligned} \tag{3.21}$$

concluding the induction and the proof.

Corollary 3.6.2. (Validity of the Test) *If $P \in H_0$, then*

$$P(\{\textit{reject } H_0 \textit{ at } \Delta\}) \leq \alpha.$$

Proof. Define

$$g_i := \mathbf{1}_{\{\ell_i \in S(A_i)\}} \tag{3.22}$$

the indicator of synchrony in window i. We will show that

$$P\left(\sum_{i=1}^{n} g_i \geq k \middle| \mathcal{A}\right) \leq P\left(\sum_{i=1}^{n} Y_i \geq k\right) \quad \forall \mathcal{A}, \tag{3.23}$$

where $Y_1, ..., Y_n$ are independent and $Y_i \sim \text{Be}(p_i^*)$. Observe that

$$P(g_i = 1|_i\ell, \mathcal{A}) \leq p_i^* \quad \forall_i\ell, \mathcal{A} \tag{3.24}$$

by the definition of p_i^*, Equation (3.14), and that ${}_ig$ is a function of ${}_i\ell$ and $\mathcal{A}$, which we will denote by h:

$${}_ig = h({}_i\ell, \mathcal{A}). \tag{3.25}$$

As a consequence,

$$\begin{aligned} P(g_i = 1|_ig, \mathcal{A}) &= \frac{P(g_i = 1, {}_ig, \mathcal{A})}{P({}_ig, \mathcal{A})} \\ &= \frac{\sum_{\{{}_i\ell : h({}_i\ell, \mathcal{A}) = {}_ig\}} P(g_i = 1|_i\ell, \mathcal{A}) P({}_i\ell|\mathcal{A})}{P({}_ig|\mathcal{A})} \\ &\leq \frac{\sum_{\{{}_i\ell : h({}_i\ell, \mathcal{A}) = {}_ig\}} p_i^* P({}_i\ell|\mathcal{A})}{P({}_ig|\mathcal{A})} \\ &= p_i^*. \end{aligned} \tag{3.26}$$

Now an application of Lemma 3.6.1 gives Eq (3.23), as desired, and plugging in $k = \inf\{t : f(t) \leq \alpha\}$ yields the corollary.

3.7 Pattern Jitter: Symmetric Method

The pattern jitter method described in the previous section has one obvious (at minimum, aesthetic) flaw, asymmetry. The internal statistical structure of the *reference* train has no bearing on the null hypothesis. Intuitively, this means that the process underlying the reference train can vary very (in fact, arbitarily) quickly, independently of the likelihood of rejecting the null hypothesis. In effect, this is because the target train can depend on the precise location of the spikes in the reference train, which is in itself a form of fine temporal structure. This is a weaker hypothesis than we wish to test; the natural hypothesis is that both spike trains are bounded in their temporal precision. Furthermore, the outcome of the test depends on the choice of reference and target: swap the trains, and the p-value changes. Both of these characteristics seem unappealing.

To extend the analogy of the asymmetric pattern jitter, the natural thing would be to write down the same null hypothesis, but for both spike trains. In analogy to the one train case, the natural assumption is that locations ℓ of patterns in different windows are independent, given the patterns $\mathcal{A}$. Then, there is the problem of where to *place* the windows: if you choose one train as a "reference" train as before, to anchor the windows, then a dependence is introduced. For example, if a local pattern in one spike train consisted of a single spike, then anchoring the window around that spike in fact *determines* the location of the pattern ℓ in the center of the window: there is no room for temporal variation. Therefore, it is necessary to choose the windows in fact *independently* of both spike trains in order to accommodate a symmetric null hypothesis. This means in practice, fixing the windows in advance of the analysis. This introduces a loss of power, since spikes will be ignored as a consequence; this is akin to "throwing away" data. On the other hand, compared to the asymmetric pattern jitter method, this loss of power is balanced by the

gain in power due to restricting the rate of change of *both* processes, rather than just one. Given enough data, the more restrictive symmetric null hypothesis would then be more powerful, but it is an empirical question if the amount of data that we have is enough. For the data sets we examined in these investigations (described below), the symmetric test was more powerful in the sense of giving rise to rejection of larger Δ, and these are the results we report.

3.7.1 Null Hypothesis

We partition the discrete time domain into N windows of length J, with the condition that every window is separated from its neighboring windows by a S bin interval, and define the following random variables. There are $2N$ windows in the two spike trains: windows 1 to N correspond to the first spike train, and windows N+1 to 2N correspond to the second spike train.

Notation.

$$\ell_i := \text{the location of the first spike in window } i.$$

$$A_i := \text{the pattern of spikes in window } i, \text{ irrespective of their temporal location}$$

$$\mathcal{A} := (A_1, A_2, ..., A_{2_N})$$

$$_i\ell := (\ell_1, \ell_2, ..., \ell_{i-1}, \ell_{i+1}, \ell_{i+2}, ...\ell_{2N})$$

Null Hypothesis. *A probability distribution P is in the null hypothesis H_0 if it satisfies*

$$\frac{1}{\Delta} \leq \frac{P(l_i = j|\mathcal{A})}{P(l_i = j+1|\mathcal{A})} \leq \Delta \quad \forall 1 \leq j < j_{\max}(A_i), \forall 1 \leq i \leq 2N, \forall \mathcal{A}, \tag{3.27}$$

and

$$\ell_1, \ell_2, ..., \ell_{2N} \text{ are conditionally independent given } \mathcal{A}, \tag{3.28}$$

where $j_{\max}(A_i)$ is the maximal location of ℓ_i associated with A_i, (i.e., the maximal value of the locations of the first spike which keeps the pattern in the window.)

Remark. Actually, a slightly more general null hypothesis is also valid:

$$\frac{1}{\Delta} \leq \frac{P(l_i = j|_i\ell, \mathcal{A})}{P(l_i = j+1|_i\ell, \mathcal{A})} \leq \Delta \quad \forall 1 \leq j < j_{\max}(A_i), \forall 1 \leq i \leq 2N, \forall_i\ell, \forall \mathcal{A}, \tag{3.29}$$

and

$$P(\ell_i|\mathcal{A}, _i\ell) = P(\ell_i|\mathcal{A}, _{i,\alpha(i)}\ell) \quad \forall_i\ell, \forall \mathcal{A}, \tag{3.30}$$

where $\alpha(i) = (N + i) \mod 2N$ (i.e., $\alpha(i)$ is the window corresponding to window i in the alternate spike train.)

3.7.2 Method

Algorithm.

- In the discrete time domain and starting with the first bin, demarcate N windows with a width of J bins, with each window separated from its preceding neighbor by S bins (i.e., bins 1 to J compose the first window, $J + S + 1$ to $2J + S$ compose the second window, et cetera).

  Define ℓ_i and A_i as in the notations from the i'th window of the target train, i.e., where ℓ_i is the location of the first spike in the window relative to the onset of the window, and A_i characterizes the pattern of spikes in the target train. Note that here the window i and the window $N + i$ correspond to the windows occupying the same position in time for the two spike trains.

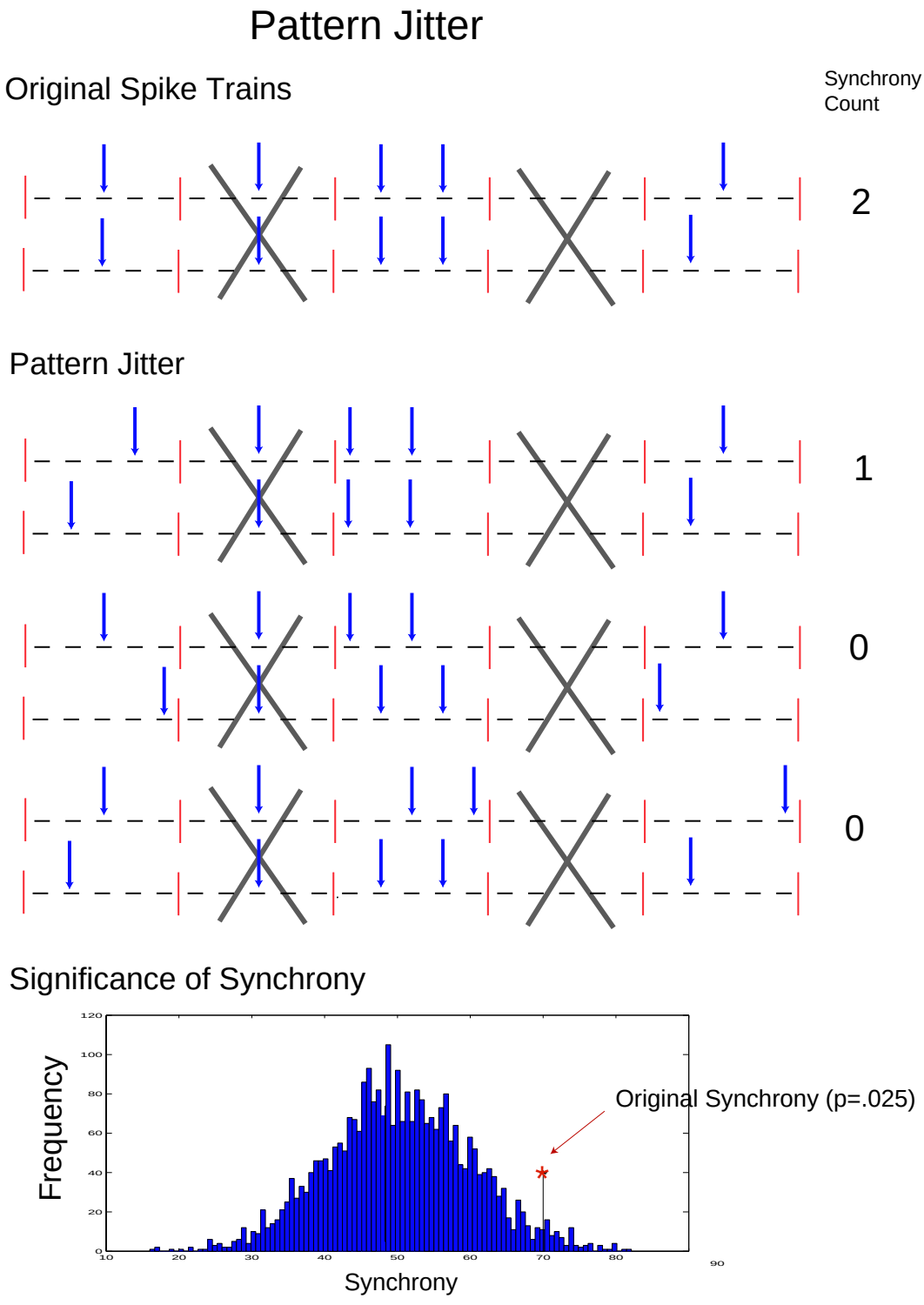

Figure 3.3: Pictorial description of the symmetric pattern jitter method. Here $J = S = 5$ bins. X's mark out the region of separation between windows. The synchrony count here is simply the number of windows that contain at least one synchronous spike. The method is similar to that illustrated in Figure 3.2, but patterns in *both* trains are translated randomly in the jittered trains, in a weighted manner determined by Δ. The tabulated distribution of these synchrony counts among the pattern-jittered trains is used to assess the significance of the synchrony counts in the original train with respect to the Δ null hypothesis. The pattern-jittered trains here are metaphorical: the actual distribution is computed analytically.

- Define

$$S(A_i, A_{N+i}) := \text{the values of } (\ell_i, \ell_{N+i}) \text{ for which } (A_i, A_{N+i}) \text{ will induce a least one pair of synchronous spikes among the spikes in windows } i \text{ and } N+i. \tag{3.31}$$

and for each window i restricted to $1 \le i \le N$, compute

$$p_i^*(A_i, A_{N+i}) = \max_{\substack{x_1,\ldots,x_{j\max(A_i)} \\ y_1,\ldots,y_{j\max(A_{N+i})} \\ x_i \ge 0\ \forall i \\ \sum_i x_i = 1 \\ \frac{1}{\Delta} \le \frac{x_i}{x_{i+1}} \le \Delta \\ y_i \ge 0\ \forall i \\ \sum_i y_i = 1 \\ \frac{1}{\Delta} \le \frac{y_i}{y_{i+1}} \le \Delta}} \sum_{(j,k) \in S(A_i, A_{N+i})} x_j y_k. \tag{3.32}$$

Equation (3.32) is a quadratic programming problem, and hence more difficult to compute than the analogous maximization problem in the asymmetric pattern jitter method, Equation (3.14), which reduces to linear programming. On the other hand, (3.32) is equivalent to maximizing a convex function, hence it is sufficient to maximize over the extremal points of its constraint set, which can be accomplished at least by enumeration, if J is not too big, in general. This is the method we adopt, facilitated by the construction of a table to reduce re-computation. It is an open question whether there is a more efficient method of determining the maximum. The explicit solution for the case of one spike in each window, $p_i^*(1,1)$, however, is derived in Lemma 3.8.5 below.

- From $p_1^*, p_2^*, ..., p_N^*$, compute the (reverse) cumulative probability distribution

$$f(t) = P\left(\sum_{i=1}^{N} Y_i \ge t\right), \tag{3.33}$$

where each Y_i is distributed as a Bernoulli random variable with parameter p_i^*, and $Y_1, Y_2, ..., Y_N$ are independent. This is as in the asymmetric pattern jitter (see the appendix of Chapter 2 for an efficient method for computing f).

- Count the number of windows Y in which there is (at least) one synchronous spike among the pair of neurons, and reject H_0 at Δ if

$$\{Y \ge \inf\{t : f(t) \le \alpha\}\}. \tag{3.34}$$

Corollary 3.7.1. (Validity of the Symmetric Test) *If $P \in H_0$, then*

$$P(\{\textit{reject } H_0 \textit{ at } \Delta\}) \le \alpha.$$

Proof. In analogy to Corollary 3.6.2, the proof is essentially an application of Lemma 3.6.1. Define

$$g_i := \mathbf{1}_{\{(\ell_i, \ell_{N+i}) \in S(A_i, A_{N+i})\}} \tag{3.35}$$

the indicator of synchrony between windows i and $N+i$. We will show that

$$P\left(\sum_{i=1}^{n} g_i \ge k \,\middle|\, \mathcal{A}\right) \le P\left(\sum_{i=1}^{n} Y_i \ge k\right) \quad \forall \mathcal{A}, \tag{3.36}$$

where $Y_1, ..., Y_n$ are independent and $Y_i \sim \text{Be}(p_i^*)$. Equations (3.27) and (3.28), along with the definition of p^*, imply that

$$P(g_i = 1|_i\ell, \mathcal{A}) \leq p_i^* \quad \forall_i\ell, \mathcal{A} \tag{3.37}$$

Further, ${}_ig$ is a function of ${}_i\ell$ and $\mathcal{A}$, which we will denote by h:

$${}_ig = h({}_i\ell, \mathcal{A}), \tag{3.38}$$

and hence the same partition argument of Equation (3.26) yields

$$P(g_i = 1|_ig, \mathcal{A}) \leq p_i^* \quad \forall 1 \leq i \leq N. \tag{3.39}$$

Finally an application of Lemma 3.6.1 gives Equation 3.36, and plugging in $k = \inf\{t : f(t) \leq \alpha\}$ gives the corollary.

3.7.3 Experimental Results

The symmetric pattern jitter method was applied to data from array-based recordings of neurons in primates, provided by Nicholas Hatsopoulos, Department of Anatomical & Organismal Biology, University of Chicago. A total of 6 data sets were obtained from three monkeys (*Macaca fascicularis* and *Macaca mulatta*), performing reaching tasks.

Experimental Methods

A silicon-based electrode array developed at the University of Utah was used to record neural discharge from multiple sites in the arm area of primary motor cortex (M1) and supplementary motor area (SMA) (see [36] for more details concerning the electrode array). During a recording session, signals from up to fifty electrodes were amplified and recorded digitally onto disk at either 20 or 30 kHz per channel (Datawave Technologies, Longmont, CO & Bionics Technologies, Salt Lake City, UT). Only waveforms that crossed a threshold (1.5 ms in duration) were stored and spike-sorted off-line. Autocorrelation functions were computed to verify single unit isolation.

Two reaching tasks were performed in separate experiments by the animals, called a *center-out* task and a *binding* task. In the *center-out* task, animals moved a two-joint manipulandum in the horizontal plane to direct a cursor from a central hold position to one of two (left or right) or eight possible (radially-positioned) targets that were displayed on a computer monitor in front of the monkey. A trial was composed of three epochs: a hold period during which time the monkey had to maintain the cursor at the hold position for 0.5 s, a random 1-1.5 s instructed delay period during which one of the radial targets appeared but movement was withheld, and a movement period initiated by target blinking (time to movement onset 365 ms). The symmetric pattern jitter method was applied to center-out task data from every trial for each unique neuronal pair, starting from the instruction period to the end of movement.

In the *binding task*, animals moved a two-joint manipulandum in the horizontal plane to direct a cursor from the bottom target through an intermediate target or via a point, and then to one of two possible final targets, either to the left or to the right. In addition, a control condition was added in which the monkey was instructed to move to the intermediate target and stop at that target. This condition was included to insure that the animal was paying close attention to the visual cues indicating which movement to perform. Data from this condition was not analyzed and, therefore, is not further referred to in this work. All five kinds of trials (i.e. {leftward & rightward} $\times$ {bound & unbound}, plus the control) were intermingled randomly throughout the experiment. Two kinds of instruction signal

were used in separate experiments. In some experiments (visually-guided), the intermediate and final targets appeared in green while all other targets were white. In other experiments (non-visually guided), all nine targets appeared in the same color. Two circles appeared in the middle of the screen either in blue, signaling a leftward sequence, or in yellow, signaling a rightward sequence. This was deemed non-visually guided because the monkey did not simply move to a different colored target, but had to learn a rule or association linking the color of the circles to the final target (either left or right) that had to be reached. Data were analyzed for each unique neuronal pair for every trial, starting from the instruction period to the end of the second movement.

In addition, there was a final data set (*no task*) in which the animal was performing no task, but was kept in position looking at a computer monitor.

Data Analysis

Individual experiments produced as many as 28 isolated, simultaneously-recorded spike trains. Spike trains were discretized into 1 ms bins (the original, post-isolated data was recorded at a maximum resolution of .25 milliseconds). Then for each unique pair of neurons (e.g., for the 28-neuron recording, there are 378 unique pairs), the *maximal* Δ that can be rejected at 95% significance was determined to a resolution of .01, using window sizes J=10 bins=10 milliseconds and separations S=10 bins=10 milliseconds. Similarly, we also computed the maximal Δ that can be rejected at 99% significance. Δ was varied from its minimal value of 1 (corresponding to the constant rate null hypothesis). Hence, we interpret the output of this procedure for a specific neuron pair as a (statistical) lower bound on the minimal temporal resolution that could support the observed synchrony, as defined by the null hypothesis.

Some neuronal pairs examined in this way exhibited very fine temporal structure. Figures 3.4, 3.5, and 3.6 show the cross-correlations taken from three neuron pairs drawn from these data sets. For these neuron pairs, the null hypothesis of temporal structure with Δ values of 1.15, 1.18, and 1.29, with 95% confidence. The pair exhibited in Figure 3.6, rejecting Δ=1.29, is the highest value of Δ which was rejected in all three data sets. Taken as is, this is by itself evidence for very fine temporal structure in cortical spike trains.

One could raise the objection, of course, of "fishing", on the basis of this evidence alone. Given *enough* samples, one could find particular examples of data that rejected the null hypothesis, even were the null hypothesis true. Hence we should consider ways of grouping the results of the tests. One way to do this is to apply a binomial test to the number of rejections at 95% confidence, assuming that the likelihood of rejection is 5%, and that the tests are independent. The notion of independence is problematic here, because some neurons are redundant in the pairs (For example, consider we have the test results for neurons pairs 1 & 2, and 2 & 3; then the results of the test for 1 & 3 are not independent of the first two tests). It is not clear how to account for this redundancy, on the other hand, in a simple way. Nevertheless, the binomial test does seem to present a useful rule of thumb for assessing the meaning of multiple tests here (which we suspect, furthermore, is on the conservative side). A table of results from the binomial test is provided in Table 3.1. The binomial test suggests that a Δ hypothesis of 8-10%/ms variation can be rejected on the basis of these data sets.

From a neural coding perspective, however, the example of the extraordinary pairs (for example the pair in Figure 3.6, where 29% per millisecond rate changes are rejected) may be the most relevant, despite concerns about fishing, for the following reasons. The alternative hypothesis of temporal coding which we really have in mind is that *some* pairs of neurons exhibit synchrony, that is indicative of temporal structure, and others do not. This is because insomuch as synchrony is a phenomenon presumed to be involved in *coding*, the

Table 3.1: Binomial Test Results for Symmetric Pattern Jitter : maximum Δ rejected

			Significance Level Criterion for Pattern Jitter	
			95%	99%
Set 1	Binomial Test	95%	1.07	1.08
	Significance Level	99%	1.05	1.05
Set 2	Binomial Test	95%	1.08	1.05
	Significance Level	99%	1.07	1.05
Set 3	Binomial Test	95%	1.09	1.10
	Significance Level	99%	1.09	1.08
Set 4	Binomial Test	95%	None	None
	Significance Level	99%	None	None
Set 6	Binomial Test	95%	1.09	None
	Significance Level	99%	None	None
Set 8	Binomial Test	95%	1.08	1.02
	Significance Level	99%	1.08	1.01
All Sets	Binomial Test	95%	1.10	1.10
	Significance Level	99%	1.08	1.08

theoretical thinking generally is that the occurrence of synchrony among a pair of neurons circumstantially signals a *relationship* between the pair. However, from the perspective of the null hypothesis, we have to assume that all neurons have the same amount of temporal structure (say, 10%/ms variation in rates), in order to reject that amount of structure. This is the function of the binomial test. Therefore, if the alternative we have in mind is actually true, the degree of temporal structure we will be able to reject is certainly less than the amount present in those *some* neurons that have extraordinary synchrony, because it will be mixed together with those neurons that are *not* firing synchronously. As a consequence, grouping to account for fishing will not characterize the alternative hypothesis which we have in mind adequately.

On the other hand, a rather remarkable, though perhaps mostly informally acknowledged, phenomenon in multi-neuronal recording is the persistent recurrence of synchrony in the form of discontinuities at time lag 0 of the cross-correlation function (as in Figure 3.6), in contrast to discontinuities at other time lags, which seem to occur markedly less frequently. This, for example, is certainly the case in the six data sets which were examined in this study. If this phenomenon, the predominance in the cross-correlations of discontinuities at 0, is reflective of a general principle, which appears plausible, then it becomes slightly less plausible to attribute instances of significant synchrony to fishing. That is, if temporal structure in the form of the precise timing of one spike train, with respect to another, were simply due to the (say 5%) *noise* in the null hypothesis, it is not obvious why precise timing would occur only at time lag 0. In that light, perhaps the degree of rejection in those cases of high synchrony alone (such as e.g., 29% per millisecond) might be more relevant for understanding the temporal precision the brain is, or is capable of, using in its spike trains.

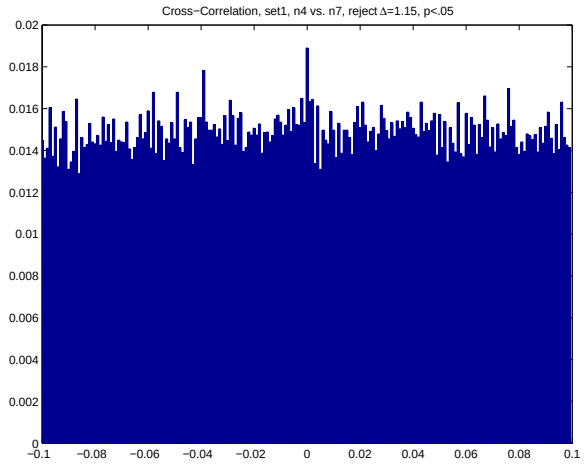

Figure 3.4: Given two discretized binary spike trains $s_1(t)$ and $s_2(t)$ for our purposes here we define the cross-correlation function $C(\tau) = \frac{\sum_t s_1(t)s_2(t+\tau)}{\sum_t s_1(t)}$ The cross-correlation function $C(\tau)$ here is drawn from two neurons in a 2-direction *center-out* task. The data is drawn from 304 trials from the instruction to the end of movement, for a total 710 seconds of data. Spike trains were discretized in bins of 1 millisecond. The x-axis τ is expressed in units of seconds. For this neuron pair, using $J = S = 10$ bins=10 milliseconds, the symmetric pattern jitter method rejected $\Delta = 1.15$ with 95% confidence.

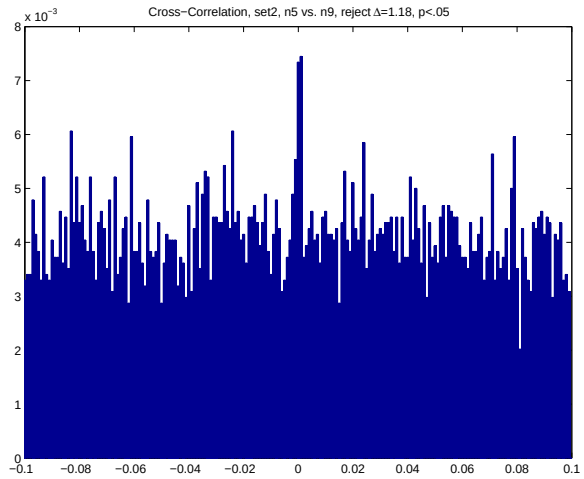

Figure 3.5: Cross-Correlation $C(\tau)$, as in Figure 3.4, here drawn two neurons in a 4-direction *center-out* data set. The data is drawn from 400 trials from the instruction to the end of movement, for a total 902 seconds of data. Spike trains were discretized in bins of 1 millisecond. The x-axis τ is expressed in units of seconds. For this neuron pair, using $J = S = 10$ bins=10 milliseconds, the symmetric pattern jitter method rejected $\Delta = 1.18$ with 95% confidence.

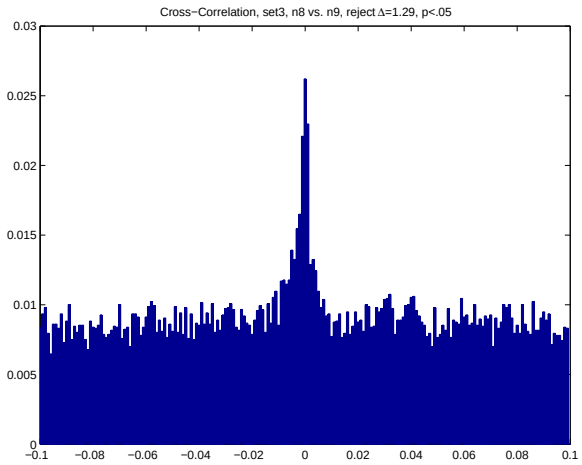

Figure 3.6: Cross-Correlation $C(\tau)$, as in Figure 3.4, here drawn from two neurons in the *no-task* task. The data is drawn a total 3605 seconds of data. Spike trains were discretized in bins of 1 millisecond. The x-axis τ is expressed in units of seconds. For this neuron pair, using $J = S = 10$ bins=10 milliseconds, the symmetric pattern jitter method rejected $\Delta = 1.29$ with 95% confidence.

3.7.4 Connections to Bernoulli Processes

The jitter method originated as a method for testing a spike train null hypothesis of locally *constant* rate Bernoulli processes, via a permutation test, with an arbitrary choice of statistic (or, perhaps in other words, an arbitrary choice of power). The constant rate assumption was a practical concern, since arbitrarily small deviations from 'constancy' could lead to rejection, given enough data. This assumption is appropriately relaxed in the pattern jitter method, but an alternative and natural way to relax that assumption would be to consider a null hypothesis which bounds the maximal percentage change in the rate function directly. This handles the primary statistical challenge in evaluating temporal structure from the spike train: allowing for differences in 'rate' at different times in the records. This would not be able to account for local temporal structures as the pattern jitter method does, but since it is perhaps an easier null hypothesis to think about it is interesting to wonder what happens if this approach were pursued. This leads to the following question: what adjustment in the jitter method (in particular, in the calculation of significance) needs to be made to accommodate a changing rate function with a bounded rate of change? It turns out that the required adjustment can be made arbitrarily large by increasing the rate of the process: there is no way to bound the adjustment without requiring an upper bound on the rate function itself. However, in the *low rate* limit, this null hypothesis is related to the null hypothesis of the pattern jitter method. This is useful to note because the low rate limit is the practical one for spike trains: even if a neuron is spiking at 100 Hz, the probability of spiking in a 1 millisecond bin is only 0.1. One way to express this is via a consistency-type theorem: in the limit of large data and small rates, the delta boundary of the pattern jitter method is the same boundary as the delta boundary of inhomogenous Bernoulli processes. Explicitly, we mean:

Definition. (Temporally-Bounded Inhomogeneous Bernoulli Process Models). *We model a pair of spike trains* $(\mathrm{X_t}, \mathrm{Y_t})$ *by pairs of independent inhomogeneous Bernoulli processes. Given a pair of rate functions,* $\mathrm{r(t)}$, *and* $\mathrm{s(t)}$, *let us denote their associated inhomogeneous Bernoulli processes by the measure* $\mathrm{P_{r(t),s(t)}}$. *Explicitly, under* $\mathrm{P_{r(t),s(t)}}$, *the*

probability that the neuron X *spikes at time t,* $\mathrm{P}(\mathrm{X_t} = 1) = \mathrm{r(t)}$*, and the probability of* X *spiking at one time is independent of its firing at other times. Similarly,* $\mathrm{P}(\mathrm{Y_t} = 1) = \mathrm{s(t)}$*, and the probability that neuron* Y *spikes at one time is independent of its firing at other times. Further, the processes* $\mathrm{X_t}$ *and* $\mathrm{Y_t}$ *are independent. Then we define the class of such processes:*

$$\begin{aligned}\mathcal{P}_T(\gamma, r^*) := \Big\{ P_{(r(t),s(t))} : &\frac{1}{\gamma} \leq \frac{r(t)}{r(t+1)} \leq \gamma \ \forall 1 \leq t \leq T-1, \quad \sup_t r(t) \leq r^* \\ &\frac{1}{\gamma} \leq \frac{s(t)}{s(t+1)} \leq \gamma \ \forall 1 \leq t \leq T-1, \quad \sup_t s(t) \leq r^* \Big\},\end{aligned} \tag{3.40}$$

i.e., $\mathcal{P}_T(\gamma, r^*)$ *is the class of Bernoulli processes whose rate functions cannot change faster than* γ*, and are bounded above by* r^**.*

Theorem 3.7.2. *(Pattern Jitter Consistency) Under the symmetric pattern jitter method,*

$$\lim_{r^* \downarrow 0} \lim_{T \to \infty} \sup_{P \in \mathcal{P}_T(\gamma, r^*)} P(\text{reject} H_0 \text{ at } \Delta) = \begin{cases} 0 & \text{if } \gamma < \Delta, \\ 1 & \text{if } \gamma > \Delta. \end{cases} \tag{3.41}$$

We present the proof of this theorem in the appendix.

3.7.5 Simulations of Bernoulli Processes

The inhomogeneous Bernoulli process is a flawed model for the spike train but nevertheless its simplicity makes it useful for examining the effectiveness of the pattern jitter method in handling the basic problem of normalizing for the effects of baseline rate on observed synchrony. We partitioned a discrete time domain into bins corresponding to 1 ms, with windows of length J=10 bins separated by S=10 bins, as in the analysis, and constructed rate functions that started at 25 Hz, 50 Hz, and 100 Hz, respectively, at every window, and then decreased the rate function at a rate of 10% per millisecond. This corresponds to maximizing the expected total number of synchronous spikes for a fixed expected number of total spikes. We extended the rate functions to 40,000 windows, comparable to the amount of time analyzed from many of the data sets above. Then we sampled 2 Bernoulli processes from these rate functions, and applied the pattern jitter method to the resulting model spike trains. 50 model spike trains were sampled for each rate function. Figure 3.7 shows the results as an empirical distribution of the maximal Δ rejected by the pattern jitter method, for the three types of rate functions. The interpretation of Δ as a statistical upper bound on the maximal rate of change of Bernoulli processes appears to remain a good approximation even at high firing rates.

3.8 Appendix

Lemma 3.2.1 *If* $X_0, X_1, ..., X_N$ *are independent and identically-distributed random variables then*

$$P\left(X_0 > X_{(k)}\right) \leq \frac{N-k}{N+1},$$

where $X_{(k)}$ *is the k'th order statistic of the random variables* $X_0, X_1, ..., X_N$*.*

Proof. The proof is a matter of setting up partitions. With $a = (a_0, a_1, ..., a_N) \in \{0,1\}^{(N+1)}$, write

$$\Omega_a := \{\omega : X_{(i)} = X_{(i+1)} \text{ if and only if } a_i = 1, \forall\, 0 \leq i \leq n\},$$

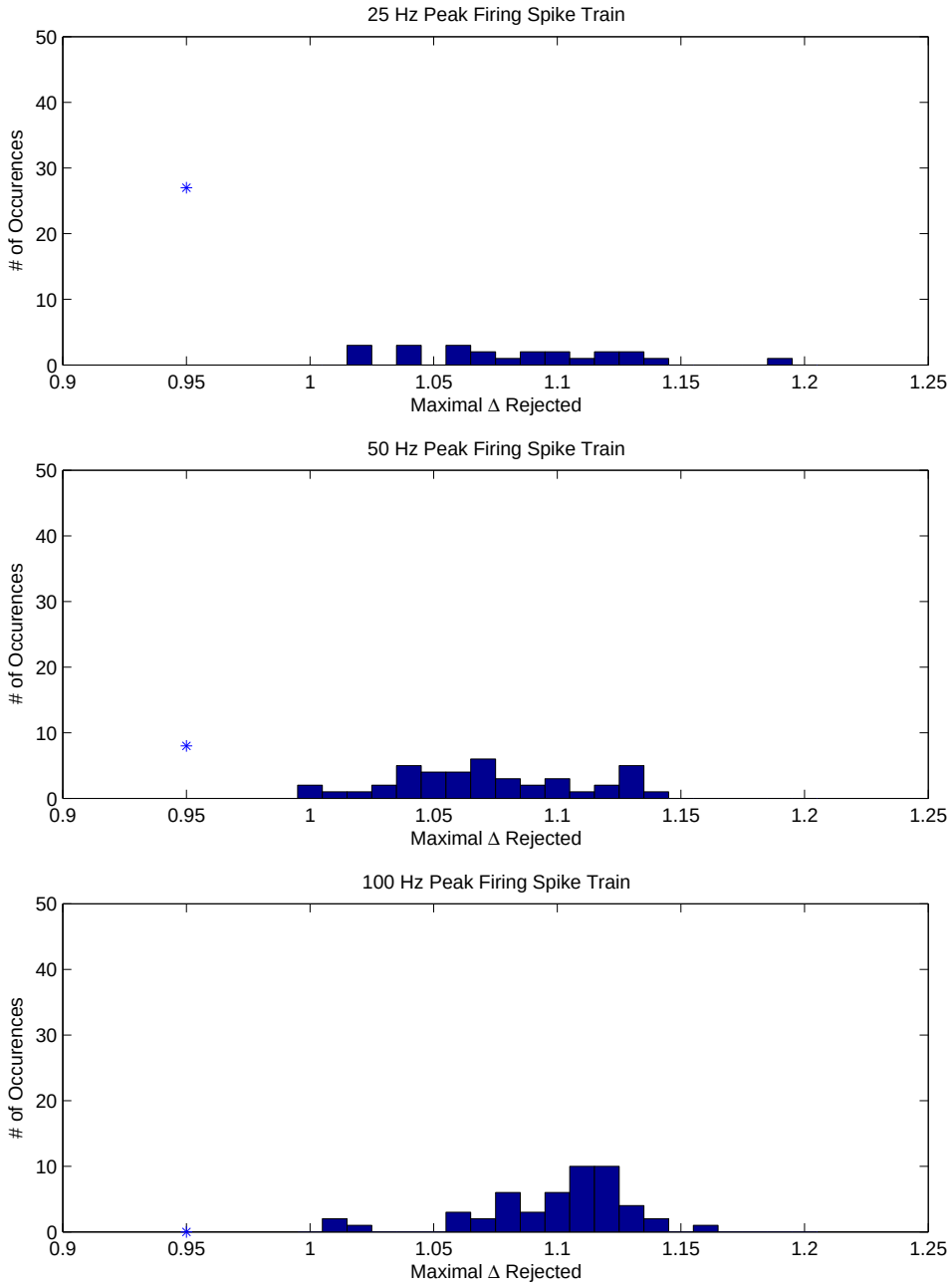

Figure 3.7: Symmetric pattern jitter results for simulations of Bernoulli processes. Bernoulli processes which maximized the number of synchronous spikes were simulated with differing peak firing rates (see text), and the pattern jitter method to determine the maximal Δ that could be rejected at 95% significance. The distribution of the maximal Δ that was rejected at 95% significance for 50 runs is plotted separately for peak firing rates of 25 Hz, 50 Hz, and 100 Hz. The starred point at 0.95 represents the number of samples that were not rejected at any Δ.

and note that $\Omega_a \cap \Omega'_a = \emptyset$ if $a \neq a'$, and $\bigcup_{a\in\{0,1\}^{(N+1)}} \Omega_a = \Omega$. So Ω_a forms a partition of the sample space according to where ties occur among the order statistics. Thus it suffices to show

$$P\left(X_0 > X_{(k)} | \Omega_a\right) \leq \frac{N-k}{N+1} \quad \forall\, a \in \{0,1\}^{(N+1)}.$$

Fixing an $a \in \{0,1\}^{(N+1)}$, take $\pi = (\pi_0, \pi_1, ..., \pi_N) \in \Pi_a$ where Π_a is the space of permutations of $\{0, 1..., N\}$ which satisfy the restriction that $\pi_i > \pi_{i+1}$ if $a_i = 1$. Then define the events

$$\begin{aligned}\Omega_{\pi,a} := \{\omega :& X_{\pi_i} < X_{\pi_{i+1}} \text{ if } a_i = 0 \\ & X_{\pi_i} = X_{\pi_{i+1}} \text{ if } a_i = 1\}\end{aligned}$$

Now define $f_a : \Pi_a \to \Pi_a$ (implicitly) via $f(\pi) = \pi^*$, where

- α satisfies $\pi_\alpha = 0$.
- $\beta = \max\{j < \alpha : a_{j-1} = 0 \text{ or } \pi_{j-1} > \pi_n\} \cup \{0\}$.
- $\pi^*_n = 0$.
- $\pi^*_k = \pi_{k-1}$, for $\beta < k \leq \alpha$.
- $\pi^*_k = \pi_k$, for *hitherto undefined* k.

It is not hard to conclude that

$$\Omega_{\pi,a} \sim_a \Omega_{\pi',a} \text{ if } f(\pi) = f(\pi')$$

defines an *equivalence relation* over (and hence a partition of) $\Omega_a = \bigcup_{\pi\in\Pi_a} \Omega_{\pi,a}$. Thus (and again), fixing $a \in \{0,1\}^{(N+1)}$, and in addition an equivalence class of Ω_a via $\pi^* = f(\pi)$ for some $\pi \in \Pi_a$, it suffices to show

$$P\left(X_0 > X_{(k)} \Big| \bigcup_{\pi: f(\pi)=\pi^*} \Omega_{\pi,a}\right) \leq \frac{N-k}{N+1}.$$

Noting that symmetry alone implies that

$$P(\Omega_{\pi,a}) = P(\Omega_{\pi',a}) \quad \forall \pi, \pi' \in \Pi_a, \tag{3.42}$$

we have

$$\begin{aligned}
& P\left(X_0 > X_{(k)} \Big| \bigcup_{\pi: f(\pi)=\pi^*} \Omega_{\pi,a}\right) \\
&= \frac{P(X_0 > X_{(k)}, \bigcup_{\pi: f(\pi)=\pi^*} \Omega_{\pi,a})}{P(\bigcup_{\pi: f(\pi)=\pi^*} \Omega_{\pi,a})} \\
&= \frac{|\{\pi \in \Pi_a : f(\pi) = \pi^*, \alpha(\pi) > k, \exists j \text{ s.t. } k \leq j \leq \alpha(\pi) \text{ and } a_j = 0\}| \cdot P(\Omega_{\pi,a})}{|\{\pi \in \Pi_a : f(\pi) = \pi^*\}| \cdot P(\Omega_{\pi,a})} \\
&\leq \frac{|\{\pi \in \Pi_a : f(\pi) = \pi^*, \alpha(\pi) > k)\}|}{|\{\pi \in \Pi_a : f(\pi) = \pi^*\}|} \\
&= \frac{N-k}{N+1}.
\end{aligned}$$

Definition. (Temporally-Bounded Inhomogeneous Bernoulli Process Models). *We model a pair of spike trains* $(\mathrm{X_t}, \mathrm{Y_t})$ *by pairs of independent inhomogeneous Bernoulli*

processes. Given a pair of rate functions, $\mathrm{r(t)}$*, and* $\mathrm{s(t)}$*, let us denote their associated inhomogeneous Bernoulli processes by the measure* $\mathrm{P_{r(t),s(t)}}$*. Explicitly, under* $\mathrm{P_{r(t),s(t)}}$*, the probability that the neuron* X *spikes at time t,* $\mathrm{P(X_t = 1) = r(t)}$*, and the probability of* X *spiking at one time is independent of its firing at other times. Similarly,* $\mathrm{P(Y_t = 1) = s(t)}$*, and the probability that neuron* Y *spikes at one time is independent of its firing at other times. Further, the processes* $\mathrm{X_t}$ *and* $\mathrm{Y_t}$ *are independent. Then we define the class of such processes:*

$$\mathcal{P}_T(\gamma, r^*) := \Big\{ P_{(r(t),s(t))} : \frac{1}{\gamma} \le \frac{r(t)}{r(t+1)} \le \gamma \ \forall 1 \le t \le T-1, \quad \sup_t r(t) \le r^* \\ \frac{1}{\gamma} \le \frac{s(t)}{s(t+1)} \le \gamma \ \forall 1 \le t \le T-1, \quad \sup_t s(t) \le r^* \Big\}, \tag{3.43}$$

i.e., $\mathcal{P}_T(\gamma, r^*)$ *is the class of Bernoulli processes whose rate functions cannot change faster than* γ*, and are bounded above by* r^**.*

Theorem 3.7.2 *(Pattern Jitter Consistency) Under the symmetric pattern jitter method,*

$$\lim_{r^* \downarrow 0} \lim_{T \to \infty} \sup_{P \in \mathcal{P}_T(\gamma, r^*)} P(\text{reject} H_0 \text{ at } \Delta) = \begin{cases} 0 & \text{if } \gamma < \Delta, \\ 1 & \text{if } \gamma > \Delta. \end{cases} \tag{3.44}$$

Proof. (We presuppose here some basic tools from probability theory [8, 14]). The basic idea is simple. For a γ-bounded Bernoulli process, some algebraic manipulation produces (Lemma 3.8.4, below):

$$\frac{1}{\gamma^{|A_i|}} \left(\frac{1 - r^*}{1 - r^*/\gamma} \right)^{|A_i|} \le \frac{P(\ell_i = j | A_i)}{P(\ell_i = j+1 | A_i)} \le \gamma^{|A_i|} \left(\frac{1 - r^*/\gamma}{1 - r^*} \right)^{|A_i|}, \tag{3.45}$$

where $|A_i|$ is the number of spikes in window i. Intuitively, in the *low rate* limit of small r^*, the cases $A_i = 1$ (one spike) and $A_i = 0$ (no spikes) will dominate. Windows with no spikes do not affect rejection, and so the relevant ratio of conditional probabilities, $\frac{P(\ell_i = j|A_i)}{P(\ell_i = j+1|A_i)}$, will typically be bounded by $\frac{1}{\gamma} - o(r^*)$ and $\gamma + \mathrm{o}(r^*)$. This connects the Δ of the null hypothesis to the γ of the Bernoulli process.

We start with the first case: $\gamma < \Delta$. The outline of the proof proceeds as follows. As in Corollary 3.7.1, define

$$g_i := \mathbf{1}_{\{(\ell_i, \ell_{N+i}) \in \mathrm{S(A_i, A_{N+i})}\}} \tag{3.46}$$

the indicator of synchrony between windows i and $N + i$.

Then we can write the rejection event as:

$$\begin{aligned} \{\text{reject } H_0 \text{ at } \Delta\} &= \left\{ \sum_{i=1}^N g_i \ge \inf \left\{ t : \mu \left(\sum_{i=1}^N Y_i \ge t \right) \le \alpha \right\} \right\} \\ &= \left\{ \sum_{i=1}^N g_i \ge N \cdot \inf \left\{ t : \mu \left(\frac{1}{N} \sum_{i=1}^N Y_i \ge t \right) \le \alpha \right\} \right\} \\ &= \left\{ \frac{1}{N} \sum_{i=1}^N g_i \ge \inf \left\{ t : \mu \left(\frac{1}{N} \sum_{i=1}^N Y_i \ge t \right) \le \alpha \right\} \right\}, \end{aligned} \tag{3.47}$$

where μ distributes the $Y_i's$ independently with $Y_i \sim \mathrm{Be}(p_i^*(A_i, A_{N+i}))$. (This form makes the connection to the law of large numbers more apparent.) Using the inequality (3.45), we deduce the existence of $R > 0$ and constants c_1 and c_2 such that

$$E_P[p^*(A_i, A_{N+i})] > c_1 > c_2 > E_P[g_i], \quad \forall i, \forall P \in \mathcal{P}_T(\gamma, R) \tag{3.48}$$

Variations on the law of large numbers then imply

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}g_i \le \liminf_{N\to\infty}\ \inf\left\{t:\mu\left(\frac{1}{N}\sum_{i=1}^{N}Y_i\ge t\right)\le\alpha\right\}\ \text{w.p.1}, P\in\mathcal{P}_T(\gamma,R). \quad (3.49)$$

which means (through (3.47))

$$\lim_{N\to\infty}P(\text{reject}H_0\text{ at }\Delta)=0.$$

To establish (3.48), define the event

$$\mathcal{A}_i := \{A_i\ge 1, A_{N+i}\ge 2\}\cup\{A_i\ge 2, A_{N+i}\ge 1\}, \quad (3.50)$$

and write

$$\begin{aligned}
&\frac{E[g_i]}{E[p^*(A_i,A_{N+i}]}\\
&=\frac{P(g_i=1)}{E[p^*(A_i,A_{N+i})]}\\
&=\frac{P(g_i=1|A_i=A_{N+i}=1)P(A_i=A_{N+i}=1)+P(g_i=1|\mathcal{A}_i)P(\mathcal{A}_i)}{p^*(1,1)P(A_i=A_{N+i}=1)+\sum_{(j_1,j_2)\in\{j_1\ge1,j_2\ge2\}\cup\{j_1\ge2,j_2\ge1\}}p^*(j_1,j_2)P(A_i=j_1,A_{N+i}=j_2)}\\
&=\frac{P(g_i=1|A_i=A_{N+i}=1)\frac{P(A_i=A_{N+i}=1)}{P(\mathcal{A}_i)}+P(g_i=1|\mathcal{A}_i)}{p^*(1,1)\frac{P(A_i=A_{N+i}=1)}{P(\mathcal{A}_i)}+\sum_{(j_1,j_2)\in\{j_1\ge1,j_2\ge2\}\cup\{j_1\ge2,j_2\ge1\}}p^*(j_1,j_2)\frac{P(A_i=j_1,A_{N+i}=j_2)}{P(\mathcal{A}_i)}}\\
&\le\frac{P(g_i=1|A_i=A_{N+i}=1)\frac{P(A_i=A_{N+i}=1)}{P(\mathcal{A}_i)}+P(g_i=1|\mathcal{A}_i)}{p^*(1,1)\frac{P(A_i=A_{N+i}=1)}{P(\mathcal{A}_i)}}\\
&\le\frac{P(g_i=1|A_i=A_{N+i}=1)\frac{P(A_i=A_{N+i}=1)}{P(\mathcal{A}_i)}+1}{p^*(1,1)\frac{P(A_i=A_{N+i}=1)}{P(\mathcal{A}_i)}},
\end{aligned} \quad (3.51)$$

where we have used the fact that $\{A_i=0\}\cup\{A_{N+i}=0\}\implies\{g_i=0\}$ and $p^*(0,0)=p^*(1,0)=p^*(0,1)=0,\forall\Delta$. By Lemma 3.8.2 below we have

$$\lim_{\substack{r^*\downarrow0\\r^*>0}}\sup_{P\in\mathcal{P}_T(\gamma,r^*)}\sup_i\frac{P(A_i=A_{N+i}=1)}{P(\{A_i\ge1,A_{N+i}\ge2\}\cup\{A_i\ge2,A_{N+i}\ge1\})}=\infty, \quad (3.52)$$

which in tandem with (3.51) means that it is sufficient for (3.48) to establish that

$$\limsup_{\substack{r^*\downarrow0\\r^*>0}}\sup_{P\in\mathcal{P}_T(\gamma,r^*)}\frac{P(g_i=1|A_i=A_{N+i}=1)}{p^*(1,1)}<1, \quad (3.53)$$

which is the content of Lemma 3.8.1. Thus we have (3.48):

$$E_P[p^*(A_i,A_{N+i})]>c_1>c_2>E_P[g_i].\quad\forall i,\forall P\in\mathcal{P}_T(\gamma,R).$$

Now we would like to show (3.49). First, since the g_i's are bounded ([14])

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} g_i - E[g_i] = 0, \text{w.p.1}, \tag{3.54}$$

and therefore

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} g_i = \limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E[g_i] \le c_2 \quad \text{w.p.1.} \tag{3.55}$$

Secondly, we show that

$$E_P[p^*(A_i, A_{N+i})] > c_1 \implies \liminf_{N\to\infty} \inf\left\{t : \mu\left(\frac{1}{N}\sum_{i=1}^{N} Y_i \ge t\right) \le \alpha\right\} \ge c_1 \quad \text{w.p.1} \tag{3.56}$$

Let

$$M := \left\{\omega : \liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p^*(A_i, A_{N+i}) \ge c_1\right\}. \tag{3.57}$$

Since $E_P[p^*(A_i, A_{N+i})] > c_1, P(M) = 1$. Fixing $\omega \in M$, Lemma 3.8.3 then implies

$$\mu\left(\frac{1}{N}\sum_{i=1}^{N} Y_i \ge t\right) \to 1 \quad \forall t < c_1, \forall \omega \in M. \tag{3.58}$$

Thus

$$\liminf_{N\to\infty} \inf\left\{t : \mu\left(\frac{1}{N}\sum_{i=1}^{N} Y_i \ge t\right) \le \alpha\right\} \ge c_1 \quad \forall \omega \in M. \tag{3.59}$$

But since $P(M) = 1$, this gives (3.56). (3.55) and (3.56) give (3.49):

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} g_i \le \liminf_{N\to\infty} \inf\left\{t : \mu\left(\frac{1}{N}\sum_{i=1}^{N} Y_i \ge t\right) \le \alpha\right\} \text{ w.p.1}, P \in \mathcal{P}_T(\gamma, R).$$

Because of the identity of {reject H_0}, (3.47), (3.49) implies

$$\lim_{N\to\infty} P(\text{reject} H_0 \text{ at } \Delta) = 0.$$

Now we return to the other case: $\gamma > \Delta$. The reasoning is similar. Fixing $\gamma > \Delta$, here it suffices to exhibit inhomogeneous Bernoulli processes which satisfy the γ requirement and asymptotically yield a rejection probability of 1. In fact, we will demonstrate the slightly stronger statement: $\exists\, q > 0$ such that

$$\lim_{N\to\infty} \sup_{P\in\mathcal{P}_T(\gamma, r^*)} P(\text{reject} H_0 \text{ at } \Delta) = 1 \quad \forall r^* \le q. \tag{3.60}$$

We will define a rate function $r(1), r(2), ..., r(n)$ on a window of fixed length n by

$$\begin{aligned} r(1) &= r^* \\ r(i+1) &= \frac{r(i)}{\gamma}, \quad \forall 1 \le i \le n-1. \end{aligned} \tag{3.61}$$

Taking two independent Bernoulli process samples from two windows with this rate function, as before we will denote A_1 and A_{N+1}, the pattern of spikes in the respective windows, ℓ_1

and ℓ_{N+1} their respective locations, and g_1 the event of synchrony between them. We will demonstrate that

$$\liminf_{r^* \downarrow 0} \frac{P(g_1 = 1 | A_1 = A_{N+1} = 1)}{p^*_\Delta(1,1)} > 1. \tag{3.62}$$

To accomplish this, define

$$m_i := P(\ell_1 = i | A_1 = 1) = P(\ell_{N+1} = i | A_{N+1} = 1) \tag{3.63}$$

Then by the algebra of Lemma 3.8.4, and the independence of the two windows

$$\frac{m_i}{m_{i+1}} = \frac{P(\ell_1 = i | A_1 = A_{N+1} = 1)}{P(\ell_1 = i+1 | A_1 = A_{N+1} = 1)} = \frac{r_i}{r_{i+1}} \frac{1 - r_{i+1}}{1 - r_i}. \tag{3.64}$$

Fix $\epsilon > 0$ such that $\epsilon < \gamma - \Delta$. Since $\frac{1-r_{i+1}}{1-r_i} \downarrow 1$ uniformly in i as $r^* \downarrow 0$, $\exists r^* > 0$ such that

$$\gamma \leq \frac{m_i}{m_{i+1}} \leq \gamma + \epsilon \quad \forall i. \tag{3.65}$$

Since in addition $\sum_{i=1}^n m_i = 1$, we have

$$\begin{aligned} P(g_i = 1 | A_1 = A_{N+1} = 1) &= \sum_{i=1}^n m_i^2 \\ &\leq p^*_{\gamma+\epsilon}(1,1) \\ &< p^*_\Delta(1,1), \end{aligned} \tag{3.66}$$

by the definition of $p^*_x(1,1)$, and its strict monotonicity in x (Lemma 3.8.5). This establishes (3.62). This in turn implies that

$$\liminf_{r^* \downarrow 0} \frac{E[g_1]}{E[p^*_\Delta(A_1, A_{N+1})]} > 1, \tag{3.67}$$

by the same argument which leads to (3.53) (i.e., Lemma 3.8.2). Thus we have exhibited a rate function for a pair of windows such that

$$E[g_1] > E[p^*_\Delta(A_1, A_{N+1})], \tag{3.68}$$

which satisfies the γ constraints for any $\gamma > \Delta$. Replicating these windows infinitely across time will give

$$\lim_{N \to \infty} P(\text{reject} H_0 \text{ at } \Delta) = 0, \tag{3.69}$$

by the same arguments as Case I (i.e., laws of large numbers). This completes the proof.

Lemma 3.8.1.

$$\limsup_{\substack{r^* \downarrow 0 \\ r^* > 0}} \sup_{P \in \mathcal{P}_T(\gamma, r^*)} \frac{P(g_i = 1 | A_i = A_{N+i} = 1)}{p^*(1,1)} < 1. \tag{3.70}$$

Proof. Here we will explicitly note p^*'s dependence on the value of Δ by writing p^*_Δ. Define

$$x_j := P(\ell_i = j | A_i) \quad y_j := P(\ell_{N+i} = j | A_{N+i}). \tag{3.71}$$

Then by the inequality of Lemma 3.8.4, there exists $\epsilon > 0$ such that

$$\frac{1}{\gamma + \epsilon} \leq \frac{x_j}{x_{j+1}} \leq \gamma + \epsilon \quad \text{and} \quad \frac{1}{\gamma + \epsilon} \leq \frac{y_j}{y_{j+1}} \leq \gamma + \epsilon \quad \forall j, \tag{3.72}$$

for sufficiently small $r^* > 0$ with $\gamma + \epsilon < \Delta$. Furthermore, x_j and y_j satisfy $\sum_j x_j = 1, \sum_j y_j = 1, x_j \geq 0 \forall j$, and $y_j \geq 0 \forall j$. As a consequence, and using in addition independence properties of the Bernoulli processes, taking $r^* > 0$ sufficiently small we have

$$\begin{aligned} P(g_i = 1 | A_i = A_{N+i} = 1) &= \sum_j P(l_i = j, l_{N+i} = j | A_i = A_{N+i} = 1) \\ &= \sum_j P(l_i = j | A_i) P(l_{N+i} = j | A_{N+i}) \\ &= \sum_j x_j y_j \\ &\leq p^*_{\gamma+\epsilon}(1,1) \\ &\overset{(a)}{<} p^*_{\Delta}(1,1). \end{aligned} \tag{3.73}$$

since, by the definition of p^*, (3.32), x_j and y_j satisfy the conditions which $p^*_{\gamma+\epsilon}$ optimizes over, and (a) follows from the fact that $p^*_{\Delta}(1,1)$ is strictly monotonic increasing in Δ (see Lemma 3.8.5).

Lemma 3.8.2. *Under the conditions of Theorem 3.7.2,*

$$\lim_{\substack{r^* \downarrow 0 \\ r^* > 0}} \sup_{P \in \mathcal{P}_T(\gamma, r^*)} \sup_i \frac{P(A_i = A_{N+i} = 1)}{P(\{A_i \geq 1, A_{N+i} \geq 2\} \cup \{A_i \geq 2, A_{N+i} \geq 1\})} = 0. \tag{3.74}$$

Proof. First, we establish the simpler limit

$$\lim_{\substack{r^* \downarrow 0 \\ r^* > 0}} \sup_{P \in \mathcal{P}_T(\gamma, r^*)} \sup_i \frac{P(|A_i| = 1)}{P(|A_i| = j)} = \infty \tag{3.75}$$

if $j > 1$. Fix the sequence of probability vectors $r^{(k)}$, where $r^{(k)} = (r_1^{(k)}, r_2^{(k)}, ..., r_n^{(k)})$, corresponding to the Bernoulli probabilities for window A_i of length n, such that $\max_i r_i^{(k)} > 0 \, \forall k$, and $\lim_{k=1}^{\infty} \max_i r_i^{(k)} = 0$. Define

$$i_*^{(k)} := \arg \max_i r_i^{(k)}. \tag{3.76}$$

Then

$$\begin{aligned}
\lim_{k\to\infty}\frac{P(|A_i|=1)}{P(|A_i|=j)} &= \lim_{k\to\infty}\frac{\sum_{i=1}^{n} r_i^{(k)}\prod_{m\neq i}(1-r_m^{(k)})}{\sum\limits_{\substack{x_1,\ldots,x_n\\ x_i\in\{0,1\}\\ \sum_i x_i=j}}\prod_{i=1}^{n}(r_i^{(k)})^{x_i}(1-r_i^{(k)})^{(1-x_i)}} \\
&\geq \lim_{k\to\infty}\frac{r_{i_*^{(k)}}^{(k)}\prod_{m\neq i_*}(1-r_m^{(k)})}{\sum\limits_{\substack{x_1,\ldots,x_n\\ x_i\in\{0,1\}\\ \sum_i x_i=j}}\prod_{i=1}^{n}(r_i^{(k)})^{x_i}(1-r_i^{(k)})^{(1-x_i)}} \\
&\geq \lim_{k\to\infty}\frac{r_{i_*^{(k)}}^{(k)}\prod_{m\neq i_*}(1-r_m^{(k)})}{\sum\limits_{\substack{x_1,\ldots,x_n\\ x_i\in\{0,1\}\\ \sum_i x_i=j}}(r_{i_*^{(k)}}^{(k)})^{j}\prod_{i=1}^{n}(1-r_i^{(k)})^{(1-x_i)}} \\
&= \lim_{k\to\infty}\frac{\prod_{m\neq i_*}(1-r_m^{(k)})}{(r_{i_*^{(k)}}^{(k)})^{j-1}\sum\limits_{\substack{x_1,\ldots,x_n\\ x_i\in\{0,1\}\\ \sum_i x_i=j}}\prod_{i=1}^{n}(1-r_i^{(k)})^{(1-x_i)}} \\
&= \infty \ \text{(if } j>1\text{)},
\end{aligned} \tag{3.77}$$

which gives us (3.75), which in turn, applied to the decomposition

$$\begin{aligned}
&\frac{P(\{A_i\geq 1, A_{N+i}\geq 2\}\cup\{A_i\geq 2, A_{N+i}\geq 1\})}{P(A_i=A_{N+i}=1)} \\
&= \frac{\sum\limits_{(j_1,j_2)\in\{j_1\geq 1,j_2\geq 2\}\cup\{j_1\geq 2,j_2\geq 1\}} P(A_i=j_1,A_{N+i}=j_2)}{P(A_i=A_{N+i}=1)} \\
&= \sum_{(j_1,j_2)\in\{j_1\geq 1,j_2\geq 2\}\cup\{j_1\geq 2,j_2\geq 1\}}\frac{P(A_i=j_1,A_{N+i}=j_2)}{P(A_i=A_{N+i}=1)} \\
&= \sum_{(j_1,j_2)\in\{j_1\geq 1,j_2\geq 2\}\cup\{j_1\geq 2,j_2\geq 1\}}\frac{P(A_i=j_1)}{P(A_i=1)}\cdot\frac{P(A_{N+i}=j_2)}{P(A_{N+i}=1)}
\end{aligned} \tag{3.78}$$

establishes the lemma.

Lemma 3.8.3. *If $Y_1, Y_2, ..., Y_N$ are i.i.d. Be(p_i) where $p_1, p_2, ...$ satisfy*

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} p_i \geq c, \tag{3.79}$$

then

$$\lim_{N\to\infty} P\left(\frac{1}{N}\sum_{i=1}^{N} X_i \geq c-\epsilon\right) = 1 \quad \forall\epsilon>0 \tag{3.80}$$

Proof. Since the Y_i's are bounded (see [14]),

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}(Y_i - p_i) = 0 \quad \text{w.p.1} \tag{3.81}$$

Hence

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} Y_i = \liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p_i \geq c \quad \text{w.p.1} \tag{3.82}$$

Convergence in probability is then just a weaker implication of the above (almost sure) convergence:

$$\lim_{N\to\infty} P\left(\frac{1}{N}\sum_{i=1}^{N} X_i \geq c - \epsilon\right) = 1 \quad \forall \epsilon > 0. \tag{3.83}$$

Lemma 3.8.4. *Consider a window of inhomogenous Bernoulli events of length $2l+1$, $X_{-l}, X_{-l+1}, ..., X_{l-1}, X_l$, where $X_i \sim Be(r(i))$, and the X_i's are mutually independent. Suppose in addition that $r(t)$ satisfies*

$$\frac{1}{\gamma} \leq \frac{r(t)}{r(t+1)} \leq \gamma \; \forall 1 \leq t \leq T-1, \quad \sup_t r(t) \leq r^* \tag{3.84}$$

As usual, we define

$$\ell := \min_{-l\leq i\leq l}\{i : X_i = 1\},$$

the location of the first spike, and

$$A := \text{ the pattern of spikes in the window .}$$

Then for all $-l \leq i \leq l-1$,

$$\frac{1}{\gamma^{|A|}}\left(\frac{1-r^*}{1-r^*/\gamma}\right)^{|A|} \leq \frac{P(\ell = i|A)}{P(\ell = i+1|A)} \leq \gamma^{|A|}\left(\frac{1-r^*/\gamma}{1-r^*}\right)^{|A|}, \tag{3.85}$$

where $|A|$ is the number of spikes in the window.

Proof. We will follow the convention that $i \in A$ refers to those positions in the pattern of spikes A which contain spikes, and $i \in A^c$ refers to those positions in the pattern of spikes A which do not contain spikes (for example if $A = 1011$, then $\{i : i \in A\} = \{1,3,4\}$, and $\{i : i \in A^c\} = \{2\}$). Writing down definitions, a little algebra leads to

$$\begin{aligned}
\frac{P(\ell = i|A)}{P(\ell = i+1|A)} &= \frac{1-r(n+1)}{1-r(1)}\prod_{i\in A}\frac{r(i)}{r(i+1)}\prod_{i\in A^c}\frac{1-r(i)}{1-r(i+1)} \\
&= \prod_{j=-l}^{l-1}\left(\frac{1-r(j)}{1-r(j+1)}\right)^{-1}\prod_{i\in A}\frac{r(i)}{r(i+1)}\prod_{i\in A^c}\frac{1-r(i)}{1-r(i+1)} \\
&= \prod_{i\in A}\frac{1-r(i+1)}{1-r(i)}\prod_{i\in A}\frac{r(i)}{r(i+1)} \\
&= \prod_{i\in A}\frac{r(i)}{r(i+1)}\frac{1-r(i+1)}{1-r(i)}.
\end{aligned} \tag{3.86}$$

An elementary optimization argument then gives

$$\max_{\{s,t:\frac{1}{\gamma}\leq\frac{s}{t}\leq\gamma,\, 0\leq s,t\leq r^*\}}\frac{s}{t}\frac{1-s}{1-t} = \gamma\left(\frac{1-r^*/\gamma}{1-r^*}\right), \tag{3.87}$$

and analogously

$$\min_{\{s,t:\frac{1}{\gamma}\leq\frac{s}{t}\leq\gamma,0\leq s,t\leq r^*\}}\frac{s}{t}\frac{1-s}{1-t}=\frac{1}{\gamma}\left(\frac{1-r^*}{1-r^*/\gamma}\right). \tag{3.88}$$

Plugging (3.87) into (3.86) produces

$$\begin{aligned}\frac{P(\ell=i|A)}{P(\ell=i+1|A)}&\leq\max\prod_{i\in A}\frac{r(i)}{r(i+1)}\frac{1-r(i+1)}{1-r(i)}\leq\prod_{i\in A}\max\frac{r(i)}{r(i+1)}\frac{1-r(i+1)}{1-r(i)}\\&=\gamma^{|A|}\left(\frac{1-r^*/\gamma}{1-r^*}\right)^{|A|},\end{aligned} \tag{3.89}$$

and a similar application of (3.88), the other side of the inequality:

$$\frac{P(\ell=i|A)}{P(\ell=i+1|A)}\geq\frac{1}{\gamma^{|A|}}\left(\frac{1-r^*}{1-r^*/\gamma}\right)^{|A|}. \tag{3.90}$$

Lemma 3.8.5.

$$p^*_\Delta(1,1)=\left(\frac{1}{1+\frac{1}{\Delta}+\frac{1}{\Delta^2}+\cdots+\frac{1}{\Delta^{n-1}}}\right)^2\sum_{i=1}^{n}\left(\frac{1}{\Delta^{i-1}}\right)^2, \tag{3.91}$$

and $p^*_\Delta(1,1)$ is strictly monotonically increasing in Δ, where n is the number of bins in a window.

Proof. For simplicity, write the constraint set $\mathcal{C}^{n,\Delta}$

$$\mathcal{C}^{n,\Delta}:=\left\{x\in\mathbb{R}^n:\frac{1}{\Delta}\leq\frac{x_i}{x_{i+1}}\leq\Delta,\sum_{i=1}^{n}x_i=1\right\}, \tag{3.92}$$

which we will write as $\mathcal{C}$ when we assume that n and Δ are implicitly fixed. Following the definition (3.32)

$$p^*(1,1)=\max_{x\in\mathcal{C},y\in\mathcal{C}}\sum_{i=1}^{n}x_iy_i. \tag{3.93}$$

The Cauchy-Schwarz inequality (C-S) implies that the maximum is attained when $x=y$:

$$\begin{aligned}\max_{x\in\mathcal{C}}\sum_{i=1}^{n}x_i^2&=\max_{\substack{x\in\mathcal{C}\\y=x}}\sum_{i=1}^{n}x_iy_i\leq\max_{\substack{x\in\mathcal{C}\\y\in\mathcal{C}}}\sum_{i=1}^{n}x_iy_i\\&\overset{\text{C-S}}{\leq}\max_{\substack{x\in\mathcal{C}\\y\in\mathcal{C}}}\sqrt{\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}y_i^2}=\sqrt{\max_{x\in\mathcal{C}}\sum_{i=1}^{n}x_i^2\max_{y\in\mathcal{C}}\sum_{i=1}^{n}y_i^2}\\&=\max_{x\in\mathcal{C}}\sum_{i=1}^{n}x_i^2.\end{aligned} \tag{3.94}$$

The maximum $\max_{x\in\mathcal{C}}\sum_{i=1}^{n}x_i^2$ is attained because $\mathcal{C}$ is a convex set. Furthermore, any $x\in\mathcal{C}$ has a version $x'\in\mathcal{C}$ such that x' is monotonically decreasing and preserves the value of the objective function: $\sum_{i=1}^{n}(x'_i)^2=\sum_{i=1}^{n}x_i^2$, since all permutations of $(x_1,x_2,...,x_n)$ preserve the objective function $\sum x_i^2$.

So it is sufficient to identify the maximum among monotonically decreasing elements of $\mathcal{C}$. We can show by contradiction that the optimal x^* in this class has the extremal form $x^*_i/x^*_{i+1}=\Delta,\forall i$, (which uniquely characterizes the solution). Assume not: $x^*=$

$(x_1^*, x_2^*, ..., x_n^*)$ is optimal, with $x_1^* \geq x_2^* \geq \cdots \geq x_n^*$, and $x^* \in \mathcal{C}$, but not of extremal form. Then let

$$\begin{aligned} j &:= \inf\left\{ i : \frac{x_i^*}{x_{i+1}^*} < \Delta \right\} \\ \ell &:= \max\left\{ \{ i \geq j : \frac{x_k^*}{x_{k+1}^*} = 1 \, \forall j < k < i \} \cup \{ j+1 \} \right\} \end{aligned} \tag{3.95}$$

Then defining

$$r_m^{(\epsilon)} = \begin{cases} x_j^* + \epsilon & \text{if } m = j, \\ x_k^* - \epsilon & \text{if } m = k, \\ x_m^* & \text{otherwise,} \end{cases} \tag{3.96}$$

$\exists \epsilon > 0$ such that $r_m^{(\epsilon)} \in \mathcal{C}$ and is monotonic decreasing. Further,

$$\sum_{i=1}^{n} \left(r_i^{(\epsilon)} \right)^2 \geq \sum_{i=1}^{n} (x_i^*)^2 , \tag{3.97}$$

since $[(x_j^* + \epsilon)^2 + (x_k^* - \epsilon)^2] - [(x_j^*)^2 + (x_k^*)^2] \geq 2\epsilon^2$, by monotonicity. But this contradicts the optimality of x^*, so x^* must be of extremal form. Now, using the relation $x_i^*/x_{i+1}^* = \Delta$ and the constraint $\sum_{i=1}^n x_i^* = 1$, we obtain

$$x_i^* = \frac{\dfrac{1}{\Delta^{i-1}}}{\displaystyle\sum_{i=1}^{n} \frac{1}{\Delta^{i-1}}}, \tag{3.98}$$

which upon substitution into $\sum_{i=1}^n (x_i^*)^2$ establishes (3.91).

To show that p_Δ^* is strictly monotonically increasing in Δ, fix arbitrary $\Delta_1 < \Delta_2$, with $x^* \in \arg\max_{x \in \mathcal{C}^{n,\Delta_1}}$ the optimal solution for Δ_1 given by (3.98). Then consider a perturbation y^ϵ:

$$y_m^\epsilon = \begin{cases} x_1^* + \epsilon & m = 1 \\ x_2^* - \epsilon & m = 2 \\ x_m^* & \text{otherwise.} \end{cases} \tag{3.99}$$

Clearly, $\exists \epsilon > 0$ such that $y^\epsilon \in \mathcal{C}^{n,\Delta_2}$ because $\Delta_1 > \Delta_2$. Further,

$$\sum_{i=1}^{n} (y_i^\epsilon)^2 - (x_i^*)^2 \geq 2\epsilon^2, (\forall \epsilon > 0) \tag{3.100}$$

by monotonicity, again. So $p_{\Delta_1}^*(1,1) < p_{\Delta_2}^*(1,1)$.

Chapter 4

Final Thoughts

The statistical methods which have been developed in this dissertation have been crafted in the interest of questions which fundamentally belong to the domain of neuroscience: questions involving, for example, the quantitative nature of the neural code, and the identification of regularities in neural responses. In that light, we have tried to apply them to data sets drawn with the help of the modern tools developed by neuroscientists, such as *in vivo* multi-electrode recordings in awake, behaving subjects. These tools offer the significant hope of broadly expanding the types of questions which are practically amenable to empirical investigation in neuroscience. We have tried, as well, to be careful to incorporate into these methods phenomena unique to the fundamental problems of inference in neuroscience, such as credibly-motivated models of variability, and biophysical knowledge such as the existence of bursts and refractory periods, and to coherently identify limits to the inferences which are available to us. These questions lie deep, however, and the light our application of these methods can cast on them are certainly of a very preliminary nature. The central problem, for example, of identifying *functional* causes from the statistical conclusions we can draw from these experiments, will probably require a great deal of work, and the concerted efforts of scientists drawn from widely-varying domains of expertise, ranging from biology to physics and mathematics. Many of these questions might be more fruitfully explored if methods such as these are incorporated more directly and *at the level of design* in neuroscientific experiments, rather that *ex post facto*, as was done here, to pre-existing data sets which were perhaps collected with other purposes in mind. Certainly, at least, these methods were developed with this idea in mind.

Bibliography

[1] M Abeles. *Corticonics: Neural Circuits of the Cerebral Cortex.* Cambridge University Press, Cambridge, England; New York., 1991.

[2] M Abeles, H Bergman, E Margalit, and E Vaadia. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *Journal of Neurophysiology*, 70(4):1629–1638, 1993.

[3] M Abeles and G L Gerstein. Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *Journal of Neurophysiology*, 60(3):909–924, 1988.

[4] E D Adrian. *The Basis of Sensation: the Action of the Sense Organs.* W. W. Norton, New York, New York, 1928.

[5] A Aertsen, G L Gerstein, M K Habib, and G Palm. Dynamics of neuronal firing correlation: modulation of 'effective connectivity'. *Journal of Neurophysiology*, 61(5):900–917, 1989.

[6] R Bellman. *Dynamic Programming.* Princeton University Press, Princeton, N.J., 1957.

[7] M J Berry and M Meister. Refractoriness and neural precision. *Journal of Neuroscience*, 18(6):200–2211, 1998.

[8] L Breiman. *Probability.* Society for Industrial & Applied Mathematics, 1992.

[9] C D Brody. Slow variations in neuronal resting potentials can lead to artefactually fast cross-correlations in the spike trains. *Journal of Neurophysiology*, 80:3345–3351, 1998.

[10] C D Brody. Correlations without synchrony. *Neural Computation*, 11:1537–1551, 1999.

[11] D R Cox and V Isham. *Point Processes.* Chapman & Hall, London & New York, 1980.

[12] A Date, E Bienenstock, and S Geman. On the temporal resolution of neural activity. Technical report, Brown University Division of Applied Mathematics, 1998.

[13] P Dayan and L F Abbott. *Theoretical Neuroscience.* MIT Press, Cambridge, Massachusetts, 2001.

[14] R Durrett. *Probability: Theory and Examples.* Wadsworth Publishing Company, 10 Davis Drive, Belmont CA, USA, 1996.

[15] B Efron and R Tibshirani. *An Introduction to the Bootstrap.* Chapman & Hall, New York, 1993.

[16] J Fodor and J Pylyshyn. Connectionism and cognitive architecture. *Cognition*, 28:3–71, 1988.

[17] W Freiberger and U Grenander. *A Short Course in Computational Probability and Statistics*. Springer-Verlag, New York, 1971.

[18] R D Frostig, Z Frostig, and R M Harper. Recurring discharge patterns in multiple spike trains: I. detection. *Biological Cybernetics*, 62:487–493, 1990.

[19] R D Frostig, R C Frysinger, and R M Harper. Recurring discharge patterns in multiple spike trains: Ii. application in forebrain areas related to cardiac and respiratory control during different sleep-waking states. *Biological Cybernetics*, 62:495–502, 1990.

[20] S Geman and K Kochanek. Dynamic programming and graphical representation of error-correcting codes. *IEEE Transactions in Information Theory*, 47:549–568, 2001.

[21] C M Gray, P Konig, A K Engel, and W Singer. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338:334–337, 1989.

[22] S Grün, M Diesmann, and A Aertsen. Unitary events in multiple single-neuron spiking activity: I. detection & significance. *Neural Computation*, 14(1):43–80, 2002.

[23] S Grün, M Diesmann, and A Aertsen. Unitary events in multiple single-neuron spiking activity: Ii. nonstationary data. *Neural Computation*, 14(1):81–119, 2002.

[24] M Gur, A Beylin, and D M Snodderly. Response variability in primary visual cortex (v1) of alert monkey. *Journal of Neuroscience*, 17(8):2914–2920, 1997.

[25] R Gütig, A Aertsen, and S Rotter. Statistical significance of coincident spikes: Count-based versus rate-based statistics. *Neural Computation*, 14(1):121–153, 2002.

[26] J M Hammersley and D C Handscomb. *Monte Carlo Methods*. Methuen & Co, Ltd, London, England, 1964.

[27] N Hatsopoulos, S Geman, A Amarasingham, and E Bienenstock. At what time scale does the nervous system operate? *Neurocomputing*, 52:25–29, 2003.

[28] R E Kass and V Ventura. A spike-train probability model. *Neural Computation*, 13:1713–1720, 2001.

[29] C Koch. *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press, 198 Madison Avenue, New York, New York, 1999.

[30] G Laurent. Dynamical representation of odors by oscillating and evolving neural assemblies. *Trends in Neuroscience*, 19(11):489–495, 1996.

[31] E L Lehmann. *Testing Statistical Hypotheses*. Springer-Verlag, New York, New York, 1986.

[32] N K Logothetis and D L Sheinberg. Visual object recognition. *Annual Review of Neuroscience*, 1996.

[33] B Lucena. *Dynamic Programming, Tree-Width, and Computation on Graphical Models*. PhD thesis, Brown University Division of Applied Mathematics, 2002.

[34] D G Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, Reading, Massachusetts, 1984.

[35] L Martignon, G Deco, K Laskey, M Diamond, W Freiwald, and E Vaadia. Neural coding: Higher-order temporal patterns in the neurostatistics of cell assemblies. *Neural Computation*, 12:2621–2653, 2000.

[36] E M Maynard, N G Hatsopoulos, C L Ojakangas, B D Acuna, J N Sanes, R A Normann, and J P Donoghue. Neuronal interactions improve cortical population coding of movement direction. *Journal of Neuroscience*, 19:8083–8093, 1999.

[37] J R Muller, A B Metha, J Krauskopf, and P Lennie. Information conveyed by onset transients in responses of striate cortical neurons. *Journal of Neuroscience*, 21(17):6978–6990, 2001.

[38] M W Oram, M C Wiener, R Lestienne, and B J Richmond. Stochastic nature of precisely timed spike patterns in visual system neural responses. *Journal of Neurophysiology*, 81:3021–3033, 1999.

[39] G Palm, A M H J Aertsen, and G L Gerstein. On the significance of correlations among neuronal spike trains. *Biological Cybernetics*, 59:1–11, 1988.

[40] Q Pauluis and S N Baker. An accurate measure of the instantaneous discharge probability, with application to unitary joint-event analysis. *Neural Computation*, 12(3):647–669, 200.

[41] D H Perkel, G L Gerstein, and G P Moore. Neuronal spike trains and stochastic point processes ii. simultaneous spike trains. *Biophysical Journal*, 7:419–440, 1967.

[42] D. I. Perrett, E. T. Rolls, and W. Caan. Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res*, 47(3):329–42, 1982.

[43] Y Prut, E Vaadia, H Bergman, I Haalman, H Slovin, and M Abeles. Spatiotemporal structure of cortical activity: properties and behavioral relevance. *Journal of Neurophysiology*, 79:2857–2874, 1998.

[44] F Rieke, D Warland, R van Steveninck, and W Bialek. *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, Massachusetts, 1997.

[45] M N Shadlen and W T Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10):3870–3896, 1998.

[46] D L Sheinberg and N K Logothetis. The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences*, 94:3408–3413, 1997.

[47] D L Sheinberg and N K Logothetis. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *Journal of Neuroscience*, 21:1340–1350, 2001.

[48] D L Sheinberg, J J Peissig, K Kawasaki, and R E B Mruczek. Eye movements during visual recognition. *J Neurophysiology*, submitted.

[49] S D Silvey. *Statistical Inference*. Chapman & Hall, 2-6 Boundary Row, London SE1 8HN, 1975.

[50] W Singer. Time as coding space in neocortical processing: a hypothesis. In G Buzsaki and Y Christen, editors, *Temporal Coding in the Brain*, pages –. Springer-Verlag, 1994.

[51] W Singer. Neural synchrony: a versatile code for the definition of relations? *Neuron*, 24:49–65, 1999.

[52] W R Softky and C Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *Journal of Neuroscience*, 13:334–350, 1993.

[53] M Stopfer, S Bhagavan, B H Smith, and G Laurent. Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature*, 390:70–74, 1997.

[54] D J Tolhurst, J A Movshon, and A F Dean. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, 23:775–785, 1983.

[55] R De R v Steveninck and W Bialek. Real-time performance of a movement-sensitive neuron in the blowfly visual system – coding and information-transfer in short spike sequences. *Proceedings of the Royal Society B*, 234:379–414, 1988.

[56] A E P Villa, I V Tetko, B Hyland, and A Najem. Spatiotemporal activity patterns of rat cortical neurons predict responses in a conditioned task. *Proc Natl Acad Sci USA*, 96:1106–1111, 1999.

[57] R. Vogels. Categorization of complex visual images by rhesus monkeys. part 1: behavioural study. *Eur J Neurosci*, 11(4):1223–38, 1999.

[58] R. Vogels. Categorization of complex visual images by rhesus monkeys. part 2: single-cell study. *Eur J Neurosci*, 11(4):1239–55, 1999.

[59] C von der Malsburg. The correlation theory of brain function. Technical Report Internal Report 81-2, Max Planck Institut Biophys. Chem., 1981.

[60] C von der Malsburg. Binding in models of perception and brain function. *Current Opinions in Neurobiology*, 5:520–526, 1995.