LETTER Communicated by John Hertz

Probability of Repeating Patterns in Simultaneous Neural Data

Anne C. Smith
annesmith@ucdavis.edu
Department of Anesthesiology and Pain Medicine, University of California, Davis, Davis, CA 95616, U.S.A.

Vinh Kha Nguyen
hacbao@gmail.com
Department of Mathematics, University of California, Davis, Davis, CA 95616, U.S.A.

Mattias P. Karlsson
karlssonm@janelia.hhmi.org
Loren M. Frank
loren@phy.ucsf.edu
Department of Physiology, University of California, San Francisco, San Francisco, CA 94143, U.S.A.

Peter Smith
p.smith@maths.keele.ac.uk
School of Computing and Mathematics, University of Keele, Keele, Staffordshire, ST5 5BG, U.K.

Large data sets arising from neurophysiological experiments are frequently observed with repeating temporal patterns. Our ability to decode these patterns is dependent on the development of methods to assess whether the patterns are significant or occurring by chance. Given a hypothesized sequence within these data, we derive probability formulas to allow assessment of the likelihood of recurrence occurring by chance. We illustrate our approach using data from hippocampal neurons from awake, behaving rats.

1 Introduction

Many experiments in neuroscience generate large quantities of neural data in the form of simultaneous binary time series. Visual inspection of the data in the form of a raster (with time on the x-axis and observation index on the y-axis) often yields apparent repeating patterns. Being able to assign

Neural Computation **22, 2522–2536** (2010)

significance to these patterns is important to distinguish true network effects from chance observations. In this letter, we introduce some results that build on theory designed to help identify significant temporal sequences in the data. Our formalism was originally developed with the specific problem of place cell replay during sleep in mind (Pavlides & Winson, 1989; Wilson & McNaughton, 1994; Lee & Wilson, 2002). However, the approach may also be helpful when applied to any data set with temporal sequence information such as the sequence of brain region activations observed in fMRI experiments (Bullmore & Sporns, 2009).

Previous statistical methods designed to analyze simultaneously recorded neural spike data include sliding window methods (Abeles & Gerstein, 1988; Louie & Wilson, 2001; Nadasdy, Hirase, Czurko, Csicsvari, & Buzsaki, 1999) and correlation methods applied pairwise between cells (Skaggs & McNaughton, 1996), across multiple cells but neglecting temporal order (Kudrimoti, Barnes, & McNaughton, 1999), and across cells and time (Foster & Wilson, 2006). Other efforts have focused on computing the likelihood of individual longer sequences occurring either computationally using shuffling techniques (Nadasdy et al., 1999) or theoretically (Lee & Wilson, 2002, 2004; Smith & Smith, 2006). More recent approaches involve using either decoding to determine whether the place cell sequence represents a meaningful trajectory (Karlsson & Frank, 2009; Davidson, Kloosterman, & Wilson, 2009) or use of high-throughput temporal data mining algorithms (Laxman, Sastry, & Unnikrishnan, 2007). An advantage of these later methods is that they include higher-order temporal information not available using methods reliant on pairwise correlations.

Our approach in this letter is to compute the likelihood of certain temporal patterns occurring by chance. In our previous work (Smith & Smith, 2006), we computed bounds for the probability of a strictly increasing sequence of length j or more occurring in word of length n. Here, we extend this work by computing the probability of an increasing sequence when some elements of the sequence occur more frequently than others, exact solutions for intermediate-length sequences, asymptotic approximations for large N, and the probability of a strictly increasing sequence with a single interruption. In current form, the formulas we have derived provide quick estimates of the significance of observations in situations where other methods might be extremely computationally intensive. We illustrate our approach applied to place cell replay of rat hippocampal neurons during sharp-wave ripple events recording during awake periods when the animal was moving slowly (less than 2 cm/sec) (Karlsson & Frank, 2009).

2 Methods

2.1 Probability of Increasing Runs in Sequential Data. Consider a stream of data that is recorded sequentially. When we use an urn model, the elements in the stream are numbers obtained by selecting them randomly, with replacement, from an urn containing numbered counters. We examine the data stream from the beginning, and our intention is to search for the appearance of runs of increasing numbers of given minimum length.

There are several mathematical problems of practical interest in identifying temporal patterns in what can appear as random output. In one problem considered by Smith and Smith (2006), the urn contained N counters labeled $1, 2, 3, \ldots, N$. A formula for the probability of j strictly increasing elements being selected (with replacement) from a sequence of length n was obtained. In this letter, the research is extended to include cases of runs with repeated counters, exact solutions of intermediate-length sequences and asymptotic approximations, and interrupted sequences.

2.2 Increasing Runs with Varied Frequencies. We extend our analysis to include the possibility that some counters occur in the urn with greater frequency than other counters. In the example of place cell replay, this would correspond to the observation that different neurons have different intrinsic firing rates.

Suppose that the urn contains m_1 counters labeled 1, m_2 labeled 2, and so on, as far as m_N counters labeled N. The urn contains $N_T = \sum_{i=1}^{N} m_i$ counters. Let the vector $\mathbf{m} = (m_1, m_2, \ldots, m_N)$. Choose $j (j \leq N)$ counters from the urn, with replacement, in order. The probability that, from all possible sequences, a sequence or run is increasing is given by

$$P_1(j, N, \mathbf{m}) = \frac{1}{N_T^j} \sum_{i_j=1}^{N+1-j} \sum_{i_{j-1}=i_j+1}^{N+2-j} \cdots \sum_{i_1=i_2+1}^{N} (m_{i_j} m_{i_{j-1}}, \ldots, m_{i_1}). \tag{2.1}$$

We call this probability $p_{j,N,\mathbf{m}}$.

If the multiplicities are all 1, then

$$\begin{aligned} P_1(j, N, \mathbf{1}) &= \frac{1}{N^j} \sum_{i_j=1}^{N+1-j} \sum_{i_{j-1}=i_j+1}^{N+2-j} \cdots \sum_{i_1=i_2+1}^{N} 1 = \frac{1}{N^j} \binom{N}{j} \\ &= \frac{N!}{N^j j!(N-j)!}, \end{aligned} \tag{2.2}$$

a result of which can be found in Smith and Smith (2006). It is helpful to observe that the number of increasing runs is combinatorially the same as the number of subsets of the N elements that have j different elements.

For the case of a strictly increasing sequence in which one counter has a multiplicity greater than 1 (say, $m_1 = m > 1$ and the rest are 1), it follows that

$$P_1(j, N, \mathbf{m_s}) = \frac{1}{N_T^j} \left[\binom{N-1}{j-1} m + \binom{N-1}{j} \right], \tag{2.3}$$

where $N_T = m + N - 1$ and $\mathbf{m}_s = (m, 1, 1, \ldots, 1)$.

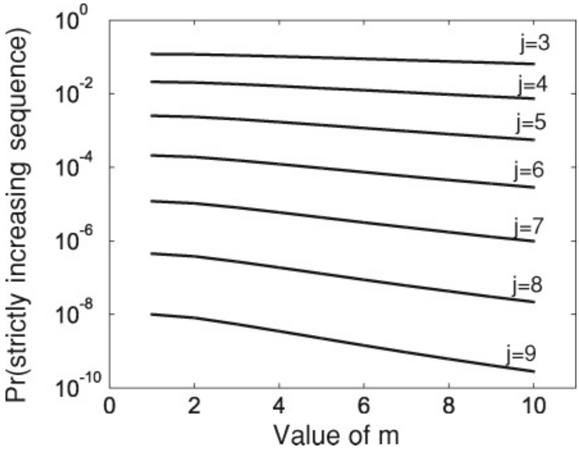

Figure 1: Probability computed using equation 2.3 of observing a strictly increasing sequence of length j as the multiplicity of one counter is varied from 1 to 10 with $N = 10$.

Equation 2.1 can be used to compute the upper and lower bounds of the probability of an increasing sequence of length j within a longer sequence of length n. For completeness, we provide the formulas in appendix A, though the derivation follows step 2 in Smith and Smith (2006).

In Figure 1 we illustrate the probability of observing a strictly increasing sequence when the multiplicity of one counter is varied from $m = 1$ to 10 from an urn with counters numbered from 1 to 10 ($N = 10$) for values of sequence length from $j = 3$ (top curve) to 9 (lowest curve). For larger sequence values (e.g., $j = 9$), a tenfold increase in the frequency of one counter results in a tenfold decrease in the probability of observing such a sequence.

2.3 Probability of an Increasing Run in a Longer Sequence. In this section we compute the exact probability of observing a strictly increasing sequence of at least length j in a stream of data of length less than three times j. We start by considering the probability that an increasing run starts from a particular position in the sequence. The urn now contains no multiple counters but unique counters labeled $1, 2, \ldots, N$. The numbers in the sequence are denoted in the order in which they appear by $c_1, c_2, c_3, \ldots$ and are randomly chosen with replacement. We are interested in increasing runs of length n at least $j < N$.

There are two cases to consider: the run starts at c_1, and the run starts at $c_r (n - j + 1 \geq r > 1)$. For the former, the probability is given by equation 2.2. Here we will denote it by

$$Q_1(j, N) = \frac{N!}{N^j\, j!(N-j)!} = \frac{1}{N^j}\binom{N}{j}. \tag{2.4}$$

For the other cases, consider a run that starts at $c_r (r > 1)$ with the prior element $c_{r-1} \geq c_r$. The number of runs is (see appendix B for outline derivations)

$$j\binom{N+1}{j+1}.$$

The probability that an increasing run starts at c_r is

$$Q(j, N) = \frac{N^{r-1}}{N^{r+j}}\, j\binom{N+1}{j+1} = \frac{j}{N^{j+1}}\binom{N+1}{j+1}. \tag{2.5}$$

As expected, this probability is independent of r.

If $n \geq 2j$, we need to be aware that two or more distinct increasing runs as defined can be present in a sequence of length n. If $n < 2j$, only one sequence is possible. In this case, the probability that an increasing run occurs in sequence of length n is

$$Q_n(j, N) = Q_1(j, N) + (n-j)Q(j, N), \tag{2.6}$$

which simplifies to

$$Q_n(j, N) = \frac{N![(N+1)j(n-j) + N(j+1)]}{N^{j+1}(j+1)!(N-j)!}. \tag{2.7}$$

On the other hand when $n \geq 2j$, equation 2.6 represents the probability that at least one increasing run occurs and the probabilty of only one increasing run becomes more complicated. In particular, if $n = 2j$, the probability that there is one increasing run of length at least j is

$$Q_{2j}(j, N) = Q_1(j, N) + jQ(j, N) - U_1(j, N), \tag{2.8}$$

where the final term, which is given by equation B.2 in appendix B, deducts the probability of sequences that include two distinct runs of length j.

For $n = 2j + 1$,

$$\begin{aligned} Q_{2j+1}(j, N) = {} & Q_1(j, N) + (j+1)Q(j, N) - [Q_1(j, N)Q(j, N) \\ & + U_1(j, N) + U_2(j, N)], \end{aligned} \tag{2.9}$$

where $U_2(j, N)$ is given by equation B.3 in appendix B. For $n > 2j + 1$, the probabilities become more complicated with combinations of combinations.

However, for $2j < n < 3j$, by careful counting, the formula for $Q_n(j, N)$ can be expressed as

$$Q_n(j,N) = Q_1(j,N) + (n-j)Q(j,N) - \left[U_1(j,N) + (n-2j)\{U_2(j,N) + Q_1(j,N)Q(j,N)\} + \left\{\sum_{k=1}^{n-2j}(n-2j-k)\right\}Q(j,N)^2\right], \tag{2.10}$$

or, after summing the series,

$$Q_n(j,N) = Q_1(j,N) + (n-j)Q(j,N) - \Big[U_1(j,N) + (n-2j)U_2(j,N) + (n-2j)Q_1(j,N)Q(j,N) + \frac{1}{2}(n-2j)(n-2j-1)Q(j,N)^2\Big]. \tag{2.11}$$

It is also possible to obtain asymptotic formulas for $Q_n(j, N)$ for large N. Stirling's formula for the factorial function has the leading asymptotic expansion,

$$N! = \sqrt{2\pi}N^{N+\frac{1}{2}}e^{-N}\left[1 + \frac{1}{12N} + O(N^{-2})\right], \tag{2.12}$$

as $N \to \infty$, or, for fixed j and small j/N,

$$(N-j)! = \sqrt{2\pi}N^{N-j+\frac{1}{2}}e^{-N}\left[1 + \frac{6j^2 - 6j + 1}{12N} + O(N^{-2})\right]. \tag{2.13}$$

In the case of $n < 2j$ (see equation 2.6), the asymptotic series for $Q_n(j, N)$ is given by

$$Q_n(j,N) = \frac{1}{(j+1)!}\left[(n+1) + \frac{(j+1)(2n - jn - j)}{2N} + O(N^{-2})\right] \tag{2.14}$$

for fixed j and n, as $N \to \infty$. Hence

$$\lim_{N\to\infty} Q_n(j,N) = \frac{n+1}{(j+1)!}.$$

If n takes its maximum value $(2j-1)$, then the limit is $2j/(j+1)!$ which gives a measure of the probability in terms of the word length j.

2.4 Increasing Runs with Interruptions. Often a data stream is imperfect in the search for increasing runs. There may be runs with perhaps one or more false or unrecorded readings interrupting the strictly increasing sequence. Here we consider the case in which one interruption occurs. The urn contains N counters as before.

We define an acceptable increasing sequence as one in which j numbers except 1 form a strictly increasing sequence. Suppose the sequence starts at c_1. It is described as an interrupted run of length if, for some s where $2 \le s \le j-2$, $c_1, c_2, \ldots, c_{s-1}$ and $c_{s+1} \ldots, c_j$ are strictly increasing. The element c_s could make the sequence increasing but may take a spurious value. Hence we require that $c_{s+1} \ge c_{s-1} + 1$. If $j = 7$ and $N = 12$, possible examples of interrupted runs are

$$1, 2, \underline{2}, 4, 6, 7, 12\ (s = 3), \quad 1, 2, 4, \underline{10}, 6, 8, 9\ (s = 4),$$
$$2, 3, 4, 6, 7, \underline{4}, 10\ (s = 6).$$

For a run starting at c_1, the probability that an interrupted sequence of length j or greater occurs, with a false output at c_s, is (see appendix B for a derivation of the result)

$$R_1(j, N) = \frac{1}{N^j}\left[N\binom{N-1}{N-j} - \binom{N}{j}\right] = \frac{(j-1)}{N^j}\binom{N}{j}. \tag{2.15}$$

The formula is independent of s. From equations 2.4 and 2.15, we can compare the effect of a misreading on the probabilities. Thus,

$$\frac{R_1(j, N)}{Q_1(j, N)} = j - 1,$$

which increases the probability by a factor of $(j-1)$ and is independent of N. The probability that an interrupted run occurs, including all possible single interruptions within the run, is

$$W_1(j, N) = (j-2)R_1(j, N) = \frac{(j-1)(j-2)}{N^j}\binom{N}{j}. \tag{2.16}$$

Interrupted runs can also start at any position c_r where $c_{r-1} \ge c_r (r > 1)$. When equation 2.4 is used, the probability that an interrupted run of length at least j with an interruption somewhere within the run is

$$W(j, N) = \frac{(j-2)}{N^{j+1}}\left[N(j-1)\binom{N}{j} - j\binom{N}{j+1}\right]. \tag{2.17}$$

With the restriction $n < 2j$ (to avoid duplicated runs), the probability that a strictly increasing interrupted run occurs in the time series of length n is

$$W_n(j, N) = W_1(j, N) + (n - j)W(j, N), \tag{2.18}$$

where $W_1(j, N)$ and $W(j, N)$ are given by equations 2.16 and 2.17. This formula acts as an upper bound for a time series of length $n \geq 2j$.

2.5 Application to Place Cell Replay. During exploratory behavior, neurons in the rat hippocampus have been identified that fire only in a specific location in space (O'Keefe & Dostrovsky, 1971). In periods of quiet awake and sleep, ordered sequences of neural spikes corresponding to an individual rat's traversal of the environment have been observed to replay in both forward and reverse order (Foster & Wilson, 2006; Lee & Wilson, 2002; Nadasdy et al., 1999; Karlsson & Frank, 2009). This replay is observed during sharp wave-ripple (SWR) events. These are short (100–200 msec) periods characterized by high-frequency (150–250 Hz) oscillations on the EEG. The probability techniques outlined here enable assessment of the likelihood of various sequences repeating in these events.

We illustrate the application of the method to place cell data from a 15-minute epoch as a rat performs in a W-track learning experiment (Karlsson & Frank, 2009). Eighteen CA1 and CA3 place cells were recorded and numbered sequentially based on their peak rates during running, such that locations corresponding to the place cells were encountered in order. (All neurons were bidirectional.) We then looked at the neural firing patterns during ripple events during the same period when the rat is relatively still, identified based on amplitude of the EEG and rat movement speed being less than 2 cm/sec. During this period, 70 sharp-wave ripple events where four or more cells fired were observed. This data set provides a useful test for our algorithm because decoding methods clearly demonstrated the presence and prevalence of replay in the data (see Karlsson & Frank, 2009, for more details).

Figures 2A and 2B illustrate two SWR events with 10 and 12 of 18 possible cells firing, respectively. Gray vertical lines represent the boundaries of the ripple events as assessed from the EEG signal. Rather than consider all spikes of each burst (see the short, vertical lines in Figure 2) we consider for our analysis the median spike firing time (see Figure 2A, triangles) for each burst of spikes. Any two spikes are considered to be part of the same burst if their interspike interval is less than 50 msec.

The temporal order of median spike firing for the first example (see Figure 2A) is $\{5\ 9\ 7\ 2\ 8\ 10\ 11\ 13\ 14\ 16\}$, an $n = 10$ length word with a sequence of $j = 7$ strictly increasing numbers. Assuming that the firing rates of all cells are equal (i.e., employing equation 2.2) and using equation

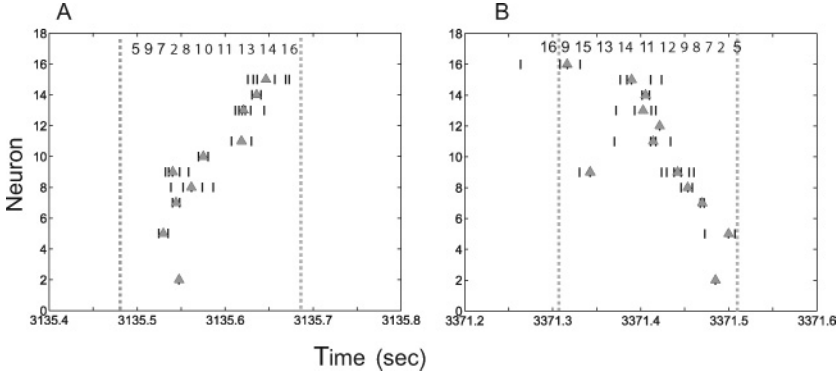

Figure 2: Close-up of hippocampal spike rasters from two sharp wave-ripple events recorded when the rat is still during a behavioral epoch. Short vertical lines indicate spike times, and triangles denote the median spike time of a burst. The numbers above each event indicate the neuron numbers listed in temporal order of occurrence. Gray vertical lines indicate the beginning and end of the SWR event as assessed from the EEG.

2.7 with $N = 18$ and $n = 10$, the probability of this event occurring is 0.00019601. We can also consider the data stream composed of all 70 SWR events concaternated together (ignoring firing during non-SWR events). By doing this, we avoid testing each SWR event independently and then having to correct results for multiple comparisons. Across the entire 15-minute epoch, there are 370 burst-averaged spikes during SWR events, so we might think of this as a $j = 7$ strictly increasing sequence lying within a longer word of length 370. If we assume that $n = 370$, this probability increases and lies within the interval [0.017307, 0.017334] using the bounds in appendix A. Thus, both scenarios result in a significant probability value at the $p < 0.05$ level.

We now consider the probability of this SWR event sequence occurring assuming unequal spike firing rates. In the data set considered, the relative frequencies of firing for neurons numbered 1 through 18 are $\mathbf{m} = \{2\ 17\ 3\ 20\ 50\ 9\ 38\ 36\ 61\ 26\ 33\ 36\ 29\ 36\ 37\ 16\ 4\ 6\}$, respectively. (These frequencies are based on the median spikes per burst rather than the actual numbers of spikes firing. That is, a burst counts as one value.) Using equation 2.1 and the bounds in appendix A, the probability of a sequence of length 7 is 0.000110 when $n = 10$. If instead we take $n = 370$, then the probability of the event lies within the interval [0.009820, 0.009827]. Thus, by taking into account the unequal firing rates, the probabilities of observing the sequences approximately halved in value.

As a second example (see Figure 2B), we consider an SWR event with a longer word length ($n = 12$) containing what can be thought of as either a strictly decreasing sequence of length 7 with one interruption or, alternatively, one strictly decreasing sequence of length 5. The probabilities and bounds for these events are, respectively, 0.008662 and [0.032331, 0.032402]. For the larger value of $n = 370$, these values are 0.5171869 (upper bound only from equation 2.18) and [0.405369, 0.775357]. Thus, a single SWR event might indicate the significance of (reverse) sequence replay, but within the longer data series of 70 SWR events ($n = 370$), the results indicate that this type of sequence will occur by chance. (Since this event lies in the same epoch as the sequence from Figure 2A, the conclusion of replay occurring still holds.) Other decoding methods applied to this data example would most likely indicate significant replay based on the fact that the place cells appear mostly in order. The method outlined here currently allows for only one interruption but may be extended in future.

2.6 Application to Network Identification from Functional Magnetic Resonance Imaging. Behavioral experiments are often performed while a human subject is within an fMRI scanner. These experiments allow identification not only of brain regions involved with various tasks but also allow inferences about correlation (or at least coactivation) between these regions and the temporal order of the activated regions. Estimation of the true network connectivity is difficult because of the magnitude of the data sets involved and noise in the system. Current techniques employed to generate estimates of connectivity often rely on thresholding pairwise correlations between brain regions, resulting in a binary connectivity matrix between regions (Bullmore & Sporns, 2009). In the future, this approach might be extended to include a temporal coactivation component. The resulting raster of binary data indicating activation with time on the x-axis and brain region on the y-axis could be analyzed using the techniques outlined in section 2.

3 Conclusion

In this letter we considered three problems that arise in the analysis of sequences in neural data. First, we computed the probability of observing a strictly increasing (or decreasing) sequence of counters drawn from an urn assuming the counters do not appear with equal probability. Second, for short sequences, we derived an exact formula for the probability of a strictly increasing sequence of numbers picked randomly and an asymptotic approximation for large N. Third, we computed the probability of observing an increasing sequence assuming a single interruption in the sequence.

Application of our methods to neural spike data from SWR events recorded from the rat hippocampus indicates that the approach is able to identify significant sequence replay in neural data. In the data set we considered, the replay effects have already been established using Bayesian decoding. We believe our methodology might prove useful in analyzing other neural spike data with a less-well-understood structure and less obvious covariate (place). A disadvantage of our approach is that it computes the probability of at least one unlikely longer sequence. The probability of, for example, six $j = 5$ strictly increasing sequences will be smaller and is not computed in this letter. Computing an upper bound for this event should, however, be mathematically tractible. A second disadvantage of this approach applied to the place field example is that it assumes that the order of place fields based on peak firing rate is the correct order. This works well if every cell has only one clearly defined gaussian place field, but this is not always the case in practice.

Our methods might be most helpful for use as a control when comparing with either nonparametric methods such as temporal data mining or with parametric model-based approaches. Parametric model-based approaches, while providing a principled framework for inference, may be slow to estimate on large data sets. That is, the number of higher-order coefficients that need to be estimated will increase considerably as the length of sequences studied is increased. For example, to compute pairwise connections between N neurons (assuming only one delay), one needs to compute only N^2 coefficients. However, the number of coefficients to be computed for a longer sequence would be likely prohibitively many (e.g., a j-length sequence requires N^j coefficient estimates).

Recent papers (Mokeichev et al., 2007; Baker & Lemon, 2000; Oram, Lestienne, & Richmond, 1999) have pointed out sensitivity of results in the field of neural spike train analysis to statistical assumptions. The results presented in this letter also address very specific problems but may be extended to other situations using a similar framework.

Appendix A: Bounds for the Probability of Selecting a Strictly Increasing Sequence in a Longer Word from an Urn with Varied Counter Frequencies

For completeness, we include here upper and lower bounds for computing the probability of a strictly increasing sequence of length j in a word of length n from an urn with varied counter frequencies. These results are derived in Smith and Smith (2006), but we now replace $p_{j,N}$ with $p_{j,N,\mathbf{m}}$. Define $Y_r^*(j, N, \mathbf{m})$ as the probability of an increasing sequence of length j that starts at position r with elements chosen from N different counters with multiplicities m. The * superscript indicates that the probability does not include the events where an increasing sequence of length j or more occurs at a positions before r. By doing this, the events do not overlap and are therefore independent.

The upper bound is a recursive sum of

$$\sum_{r=1}^{n-j-1} Y_r^*(j, N, \mathbf{m}), \tag{A.1}$$

where

$$\begin{aligned} Y_1^*(j, N, \mathbf{m}) &= p_{j,N,\mathbf{m}} \\ Y_r^*(j, N, \mathbf{m}) &= p_{j,N,\mathbf{m}} - p_{j+1,N,\mathbf{m}} \quad \text{for } 1 < r \le j+1 \end{aligned}$$

and

$$Y_r^*(j, N, \mathbf{m}) \le (p_{j,N,\mathbf{m}} - p_{j+1,N,\mathbf{m}}) \left(1 - \sum_{s=1}^{r-j-1} Y_s^*(j, N, \mathbf{m})\right)$$
$$\text{for } j+1 < r < n-j+1. \tag{A.2}$$

The lower bound is also computed recursively using equation A.1 but replacing equation A.2 with

$$\begin{aligned} Y_r^*(j, N, \mathbf{m}) &\ge (p_{j,N,\mathbf{m}} - p_{j+1,N,\mathbf{m}}) - (r-j)p_{j,N,\mathbf{m}}^2 \\ &\qquad \text{for } j+1 \le r < n-j+1 \\ &\quad \text{if positive} \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Appendix B: Additional Details for Results in Sections 2.3 and 2.4

This section outlines how to obtain the probabilities in sections 2.3 and 2.4. The methods employed include a mixture of binomial identities (see Prudnikov, Brychkov, & Marichev, 1986, section 4.2 and Knuth, 1998) and symbolic computation using Mathematica.

The result for $Q_1(j, N)$ (see equation 2.4) follows since the sequence is equivalent to the number of ways in which j objects can be chosen from N, which is then divided by the total number of possible sequences.

Equation 2.5 for $Q(j, N)$ can be verified by observing that the number of runs of length j starting with 1 is $\binom{N-1}{j-1}$, starting with 2 is $\binom{N-2}{j-1}$, as far as $\binom{j-1}{j-1}$ starting with $N-j+1$. Hence, the probability that a run starts at $c_r (r \ne 1)$ such that $c_{r-1} \ge c_r$ is

$$Q(j, N) = \frac{1}{N^{j+1}} \sum_{p=1}^{N-j+1} (N-p+1)\binom{N-p}{j-1} = \frac{j}{N^{j+1}}\binom{N+1}{j+1}. \tag{B.1}$$

For equation 2.8, in which $n = 2j$, we simply find the probability that the first j and the second consecutive j are increasing, and then subtract the probability of an increasing run of length $2j$. Therefore, the probability that two distinct increasing runs occur is given by

$$U_1(j, N) = \frac{1}{N^{2j}}\left[\binom{N}{j}^2 - \binom{N}{2j}\right]. \tag{B.2}$$

For $n = 2j + 1$ there will also occur consecutive runs of length j but with the first run starting at c_2. For equation 2.9, we apply the same method as above but use equation B.2. Thus,

$$U_2(j, N) = \frac{1}{N^{2j+1}}\left[j\binom{N+1}{j+1}\binom{N}{j} - 2j\binom{N+1}{2j+1}\right]. \tag{B.3}$$

In section 2.4 on interrupted runs, equation 2.15 can be derived as follows. The element c_{s-1} can take any integer such that $s - 1 \le c_s \le (s - 1) + (N - j)$, while, correspondingly, $s + 1 \le c_{s+1} \le (s + 1) + (N - j)$. Hence, the sum of the products of the pairwise probabilities is

$$\sum_{k=0}^{N-j}\binom{s-2+k}{s-2}\binom{N-s-k}{j-s} = \binom{N+1}{N-j},$$

which is independent of s. Taking account that c_s can take countervalues, including ones that make the j runs increasing, which must be included, it follows that

$$R_1(j, N) = \frac{1}{N^j}\left[N\binom{N-1}{N-j} - \binom{N}{j}\right] = \frac{(j-1)}{N^j}\binom{N}{j}$$

after simplification.

Equation 2.17 can be constructed in a similar manner using $R_1(j, N)$. Thus,

$$\begin{aligned} W(j, N) &= (j-2)[NR_1(j, N) - R_1(j+1, N)] \\ &= \frac{(j-2)}{N^{j+1}}\left[N(j-1)\binom{N}{j} - j\binom{N}{j+1}\right]. \end{aligned}$$

References

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

Baker, S. N., & Lemon, R. N. (2000). Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *J. Neurophysiol.*, *84*, 1770–1780.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198.

Davidson, T. J., Kloosterman, F., & Wilson, M. A. (2009). Hippocampal replay of extended experience. *Neuron*, *63*(4), 497–507.

Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, *440*(7084), 680–683.

Karlsson, M. P., & Frank, L. M. (2009). Awake replay of remote experiences in the hippocampus. *Nature Neuroscience*, *12*(7), 913–918.

Knuth, D. E. (1998). *The art of computer programming*. Reading, MA: Addison-Wesley.

Kudrimoti, H. S., Barnes, C. A., & McNaughton, B. L. (1999). Reactivation of hippocampal cell assemblies: Effects of behavioral state, experience, and EEG dynamics. *Journal of Neuroscience*, *19*(10), 4090–4101.

Laxman, S., Sastry, P. S., & Unnikrishnan, K. P. (2007). Discovering frequent generalized episodes when events persist for different durations. *IEEE Transactions on Knowledge and Data Engineering*, *19*(9), 1188–1201.

Lee, A. K., & Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, *36*(6), 1183–1194.

Lee, A. K., & Wilson, M. A. (2004). A combinatorial method for analyzing sequential firing patterns involving an arbitrary number of neurons based on relative time order. *Journal of Neurophysiology*, *92*(4), 2555–2573.

Louie, K., & Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, *29*(1), 145–156.

Mokeichev, A., Okun, M., Barak, O., Katz, Y., Ben-Shahar, O, & Lampl, I. (2007). Stochastic emergence of repeating cortical motifs in spontaneous membrane potential fluctuations in vivo. *Neuron*, *53*(3), 413–425.

Nadasdy, Z., Hirase, H., Czurko, A., Csicsvari, J., & Buzsaki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience*, *19*(21), 9497–9507.

O'Keefe, J., & Dostrovsky, J. (1971). Hippocampus as a spatial map: Preliminary evidence from unit activity in freely-moving rat. *Brain Research*, *34*(1), 171–175.

Oram, M. W., Lestienne, R., & Richmond, B. J. (1999). Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *Journal of Neurophysiology*, *81*, 3021–3033.

Pavlides, C., & Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *J. Neurosci.*, *9*, 2907–2918.

Prudnikov, A. P., Brychkov, I. U. A., & Marichev, O. I. (1986). *Integrals and series*. New York: Gordon and Breach Science Publishers.

Skaggs, W. E., & McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, *271*, 1870–1873.

Smith, A. C., & Smith, P. (2006). A set probability technique for detecting relative time order across multiple neurons. *Neural Computation*, *18*(5), 1197–1214.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, *265*(5172), 676–679.

Received December 4, 2009; accepted March 31, 2010.