Published in final edited form as:
J Neurosci Methods. 2008 August 15; 173(1): 129–139. doi:10.1016/j.jneumeth.2008.05.013.

Probability distributions of the logarithm of inter-spike intervals yield accurate entropy estimates from small datasets

Alan D. Dorval[1]

1*Department of Biomedical Engineering, Duke University, Durham NC*

Abstract

The maximal information that the spike train of any neuron can pass on to subsequent neurons can be quantified as the neuronal firing pattern entropy. Difficulties associated with estimating entropy from small datasets have proven an obstacle to the widespread reporting of firing pattern entropies and more generally, the use of information theory within the neuroscience community. In the most accessible class of entropy estimation techniques, spike trains are partitioned linearly in time and entropy is estimated from the probability distribution of firing patterns within a partition. Ample previous work has focused on various techniques to minimize the finite dataset bias and standard deviation of entropy estimates from under-sampled probability distributions on spike timing events partitioned linearly in time. In this manuscript we present evidence that all distribution-based techniques would benefit from inter-spike intervals being partitioned in logarithmic time. We show that with logarithmic partitioning, firing rate changes become independent of firing pattern entropy. We delineate the entire entropy estimation process with two example neuronal models, demonstrating the robust improvements in bias and standard deviation that the logarithmic time method yields over two widely used linearly partitioned time approaches.

Keywords

entropy estimation; finite dataset bias; firing pattern; inter-spike interval; information theory

Introduction

In response to changing environmental conditions, living objects alter their behavior to better succeed in the new environment. In simple organisms, behavioral responses are typically reflexive responses driven primarily by chemical gradients interacting with organic molecules at the organism-environment interface. With the advent of the nervous system however, higher animals abstracted some behavioral responses, enabling animals to learn appropriate responses (e.g., by trial and error) over the course of a lifetime, rather than relying on the laborious trudge of evolution. To understand how the passage of electrical signals between neurons enables sensation, commands movements and in humans at least gives rise to self-awareness, we would like to understand the symbols and ultimately the language that neurons use to communicate (for review, see Rieke et al., 1997).

Correspondence: Alan D. Dorval, Address Box 90281, Hudson 139, Duke University, Durham NC, 27708, Email alan.dorval@duke.edu, Phone (919) 660-5250, Fax (919) 684-4488.

While some sensory neurons may code for environmental cues in an analog domain (e.g. the smoothly graded membrane potentials of photoreceptors in retinal responses to photons) and local concentrations of some other molecules do play a role, the majority of signals in the nervous system are passed via action potential transmission. The arrival times of presynaptic action potential "spikes" carry information upon which a neuron must operate. As experimentalists we may not know what information a train of presynaptic spikes conveys but we can quantify the amount of information it carries, typically measured in bits. One bit of information is equivalent to the answer to a single true-or-false question. Information however, must be *about* something. Without knowing the neuronal input, we do not know what information is about, and information measures are impossible.

Instead of information, we quantify the variability of neuronal output with entropy, also measured in bits. Neuronal firing pattern entropy bounds the maximum information a neuron could transmit downstream, given some noise-free, ideally responsive, downstream neuron (Shannon & Weaver, 1949; MacKay & McCulloch, 1952). As electrophysiologists, we are often stuck recording action potentials from neurons without access to their inputs. In such cases, pattern entropy is an optimal measure of information transmission in the neurons, quantifying the variability present in their patterns of activity. Because neuronal spike rate and spike variability may change independently, we would like firing pattern entropy to be orthogonal to average spike rate, enabling entropy to remain constant in response to mere frequency changes.

Although firing pattern entropy is an optimal measure of variability, estimating entropy is perceived as a difficult venture. Many generally accepted algorithms require large amounts of data to settle on robust entropy estimates (Strong et al., 1998). Comparisons of entropy estimation techniques often focus on which techniques minimize bias or standard deviation for given probability distributions (e.g., Paninski, 2003). However, the effects of the probability distribution construction method has received comparably little attention in the literature.

Since construction of a probability distribution requires binning or rounding-off, some information inherent in each spike train is always discarded. In evaluating how best to bin, we consider which characteristics of a spike train are likely to carry reliable information, and which are likely to be noise that it would be reasonable to round out. Information estimation techniques that routinely rely on sub-microsecond changes in spike times are likely to require too many spikes to achieve experimentally. Indeed, they are unlikely to matter at all, given the roughly 100 μs time scales of the fastest electrophysiologically relevant reactions. Likewise, information carried by one spike is unlikely to depend on the arrival time of an action potential that arrived 24 hours in the past. Intuitively, time scales of milliseconds to maybe tens of seconds are the ones likely to matter to a neuron. But, do they all matter equally? In particular, is there a difference between inter-spike intervals (ISIs) of 2 and 3 ms? Is there an equally significant difference between ISIs of 102 and 103 ms? We propose that time differences of informational interest should scale with duration, or more precisely, that the measure that varies most linearly with biological importance is the logarithm of the ISI.

We begin by exploring intrinsic differences between the distributions of ISIs and the distributions of the logarithm of ISIs. Assuming initially that all ISIs are independent, we set a fixed number of probability distribution bins at the outset, and compare entropy calculated from the traditional linear ISI method (Rieke et al., 1993; Dayan and Abbott, 2001) with entropy calculated from a method that builds upon logarithmic binning (Sigworth & Sine, 1987; Newman, 2005), as recently applied by others to neuronal ISI distributions (Selinger et al., 2007). Employing the direct entropy estimation technique with a fixed number of bins, we present analytical results indicating that ISI distributions binned logarithmically yield consistently higher entropy estimates with less relative bias and less standard deviation than

ISI distributions binned linearly. We show that firing rate is orthogonal to the entropy of logarithmically binned ISI distributions, but highly correlated to the entropy of linearly binned ISI distributions. We then use computational simulations of two different neuron models to show that logarithmically binned ISIs reduce the bias and standard deviation of entropy estimates from both linearly binned ISIs and from an alternate linear binning approach (Strong et al., 1998), regardless of the entropy estimation technique employed. Final, we show with simulations that these improvements in bias and standard deviation hold even when subsequent ISIs are not assumed to be independent.

Methods

Computational analyses and theoretic entropy calculations were performed with Octave, an open source, numerical computation environment (http://octave.org). Neuronal models were implemented and simulated within the differential equation solving framework. Entropy estimation from the simulated neuronal activity was performed with the Spike Timing Analysis Toolkit (http://neuroanalysis.org/toolkit) which was adapted and compiled to run in Octave.

Entropy Estimation

For the class of entropy estimation techniques utilized, spike times must be converted into a train of events whose probabilities can be estimated from the data. Various techniques can be used subsequently to estimate the entropy from the probability distribution of events. In this work, we converted the spike times into three different event classes and compared the results.

For the first event classification method, the “Linear ISI” method (Rieke et al., 1993; Dayan and Abbott, 2001), spike times were converted into inter-spike intervals (ISIs). Those ISIs were binned in linear time into K_{Lin} bins of equal width (t_{bin}) and labeled by their bin identity: 1 to K_{Lin}. For the second event classification method, the “Logarithmic ISI” method, spike times were also converted into ISIs. They were segmented into bins of constant logarithmic time, where the right-hand side of the k^{th} bin was defined as: $ISI_k = ISI_0 * 10^{k/\kappa}$, where k ranged from 1 to K_{Log}. The zeroth time ISI_0 was set below the shortest observed ISI, and K_{Log} was set such that ISI_{KLog} was larger than the longest ISI. The particular choice of ISI_0 did not alter significantly subsequent calculations. The parameter κ, the number of discrete time bins per ISI decade, is conceptually similar to the reciprocal of the time bin resolution in linear ISI method. For the first section of the manuscript K_{Lin} and K_{Log} were set to each other to provide a fair analytical comparison, and t_{bin} and κ were set such that the linear and logarithmic distributions spanned the same temporal domain. For analysis of the simulations in the second and third sections, t_{bin} and κ were set to reasonable values for the data (as they would be in the real world experimental case) that would yield roughly equivalent entropy estimates (a constraint that is only relevant for the forced comparisons considered here). Thus for simulations, K_{Lin} and K_{Log} were set such that (t_{bin} K_{Lin}) and (ISI_0 $10^{K_{Log}/\kappa}$) exceeded the longest ISI.

For the third and most widely publicized event classification method, henceforth referred to as the “Spike Count” method, time is segmented into tiny bins labeled with the number of spikes they contain. If the bin width is shorter than the briefest ISI, each bin is thus label '0' or '1' for no spikes or one spike, respectively. To keep comparisons as straight-forward as possible, the bin size of the spike count method was always equal to that of the linear ISI method. Naive entropy estimation from the distribution of spike vs no-spike events is independent of spike pattern, merely reflecting the average firing rate and user-selected bin width. Therefore, trains of M consecutive spike counts were constructed in which the probability of a spike in any bin is not assumed to be independent of previous or subsequent bins (Strong et al., 1998). For the work presented through the second section of Results, we consider trains of M = 6 consecutive bins to constitute an event. With two possible bin states (0 and 1) and six bins, there are 2^6 or

64 possible events or 'words' (i.e., 000000, 000001, 000010, … 111111). Smaller and larger bin counts are explored in the final section of Results.

At this point for each method, spike trains had been converted into series of events labeled from $1–K_{Lin}$, $1–K_{Log}$ or $1–2^{M}$ for the linear ISI, logarithmic ISI or spike count methods, respectively. With the three classifications completed, subsequent entropy estimations were essentially identical for the three methods. A one dimensional probability distribution was constructed over the set of events where the probability of each, $P(ISI_k)$ or $P(word_m)$, was found as the number of times each event occurred divided by the total number of events in the series (e.g., Fig. 5). From these distributions, entropy was calculated via seven different estimation techniques that vary in degree of bias, standard deviation and computational complexity: the classical direct technique (Shannon & Weaver, 1949), Ma lower bound (Ma, 1981), best upper bound (Paninski, 2003), Treves-Panzeri-Miller-Carlton (Treves and Panzeri, 1995;Miller, 1955;Carlton, 1969), Jackknife (Efron & Tibshirani, 1993), Wolpert-Wolf (Wolpert & Wolf, 1994;Wolpert & Wolf, 1995) and Chao-Shen (Chao & Shen, 2003). As an example, with the classic estimation technique we found the direct entropy estimate H_{Dir} as:

$$H_{\mathrm{Dir}}^{1\mathrm{D}} = -\sum_{k=1}^{K} P(ISI_k)\log_2 P(ISI_k)$$

$$H_{\mathrm{Dir}}^{M} = \frac{-1}{f}\sum_{m=1}^{2^M} P(\mathrm{word}_m)\log_2 P(\mathrm{word}_m)$$

where f is a rate correction factor, in our case the mean number of spikes per word so that estimates from the three methods could be compared directly in units of bits per spike. For analysis of the computational simulations, the probability distributions and corresponding entropy estimates were found for each data trial, for a number of subset sizes ranging from 1 second to 3 minutes in duration, yielding means with standard deviation for each estimate as a function of dataset size. For the 3 methods by 7 estimation techniques, the average difference between the entropy estimates from each data subset and the full dataset was found as the bias for that method, technique and subset size. Standard deviations and biases were compared for the 21 pairings of method and technique.

The entropy estimation described above assumes that successive events, be they 6-bin words or ISIs, are independent. To avoid this assumption, the same estimation techniques can be performed on the probabilities of strings of events, yielding higher dimensional probability distributions. The spike count method extends to higher dimensions simply by increasing the number of bins per word M, and adjusting f (Strong et al., 1998). Higher dimensional probability distributions for the ISI methods were found by measuring the probabilities of all sets of consecutive ISIs. As examples from the two dimensional or paired ISI case: $P(ISI_b \mid ISI_a)$ and $P(ISI_b,ISI_a)$ were found as the number of times ISI_b followed ISI_a divided by the occurrences of ISI_a or the total number of ISI pairs, respectively (Fig. 8). Entropy for these higher dimensional distribution was estimated via the same seven methods. Continuing our example, the higher dimensional direct entropies are found as (Rieke et al., 1993):

$$H_{\mathrm{Dir}}^{2\mathrm{D}} = \frac{-1}{2}\sum_{a=1}^{K}\sum_{b=1}^{K} P(ISI_a,ISI_b)\log_2 P(ISI_b \mid ISI_a)$$

$$H_{\mathrm{Dir}}^{3\mathrm{D}} = \frac{-1}{3}\sum_{a=1}^{K}\sum_{b=1}^{K}\sum_{c=1}^{K} P(ISI_a,ISI_b,ISI_c)\log_2 P(ISI_c \mid ISI_a,ISI_b)$$

and so on, where the fractional coefficients (i.e. ½, ⅓, etc.) scale the estimates from units of bits per word to units of bits per spike.

Entropy estimates for data subsets of each dimension were plotted versus the reciprocal of their respective dimensions (Strong et al., 1998). These monotonic plots approach zero at very high dimensions due to dataset size limitations. A least-squares linear fit was made to data points that were not overly contaminated by the finite dataset size limitation. The zero crossing for this fit, corresponding to an estimate of the entropy per spike for infinitely long spike trains, was taken as the best estimate of the true entropy that each technique could yield. For the spike count method, the number of data points used in the linear fit was varied to return the most reasonable estimate of the true entropy for the various data subset sizes. While such adjustments could have been made for the ISI methods, we minimized user-introduced bias by fitting only the first three dimensional estimates in all cases and for all data subsets.

The above procedure was repeated for all three classification methods with each of the seven entropy estimation techniques, for data subset sizes ranging from 1 second to 3 minutes. Bias and standard deviations were found and compared for each combination of subset size, estimation technique and classification method.

Neuronal Models

Two computational model neurons were implemented in Octave. The first regular spiking (RS) neuronal model consisted of only persistent sodium and fast potassium conductances (Izhikevich, 2007). The persistent sodium conductance responded as an instantaneous function, leaving only two state variables in this model: membrane potential V_m and a regular K^+ activation variable n_r. The second intrinsic bursting (IB) neuronal model resembled the first but with both fast and slow K^+ conductances (Izhikevich, 2007). The bursting model has three state variables: membrane potential V_m, fast K^+ activation n_f, and slow K^+ activation n_s. The equations for both models are:

$$\frac{dV_m}{dt} = \left(\tilde{g}_{Na} m_\infty (V_{Na} - V_m) + (\tilde{g}_{Kr} n_r + \tilde{g}_{Kf} n_f + \tilde{g}_{Ks} n_s)(V_K - V_m) + g_L (V_L - V_m) + I_{\text{app}}\right) / c_m$$

$$\frac{dn_x}{dt} = (n_\infty - n_x)/\tau_{nx} \quad m_\infty = \left(1 + e^{\frac{20+V_m}{-15}}\right)^{-1} \quad n_\infty = \left(1 + e^{\frac{25+V_m}{-5}}\right)^{-1}$$

where the dn_x/dt equation applies for n_r in RS (i.e., $g_{Kf} = g_{Ks} = 0$), and both n_f and n_s in IB (i.e., $g_{Kr} = 0$). The K^+ activation maximal conductances and time constants were set at 10, 9, and 5 mS/cm^2, and 1.00, 0.15 and 15.00 ms for the regular, fast and slow K^+ conductances, respectively. The applied current I_{app} changed depending on the experiment described. All other other parameters were the same in both models: $\{g_{Na}, g_L\}=\{20, 8\} mS/cm^2$, $\{V_{Na}, V_K, V_L\} = \{60, -90, -80\} mV$, and $c_m = 1.0\ \mu F/cm^2$.

Applied Current Waveforms

Various constant currents in the perithreshold regime were applied to each model neuron to determine the minimum currents required to elicit repetitive firing, the rheobase I_{rheo}. Applied constant currents to each model were then ranged from a minimum of $I_{rheo}+1.0\ pA/cm^2$ to a maximum of $I_{rheo}+100 \mu A/cm^2$, by which point both models had entered depolarization block. The spike times of each model responding to all currents were transformed into trains of ISIs.

Subsequent to mapping responses to constant input, noisy current inputs were constructed and presented to each model. These current inputs were nominally 1/f noise, band-limited from 1Hz, to keep signals stationary across seconds, to 100kHz. The noise signal was generated from the inverse fast Fourier transform of the ideal frequency spectrum with pseudo-random, uniformly distributed phases. To avoid ultra-high frequency current changes that would destroy the performance of the differential equation solver, a cubic spline was fit to the 10μs spaced current values. The exact current at each time point addressed by the differential equation solver

was computed from the spline vector. Two current inputs were presented to each model: weak and strong noise. The weak noise consisted of a weakly supra-threshold constant current (5 $\mu A/cm^2$) superimposed with 1/f noise with total 2.5 $\mu A^2/cm^4$ of power. The strong noise consisted of zero mean 1/f noise with 25 $\mu A^2/cm^4$ of power.

Simulation Analyses

Both models were presented with weak and strong noise inputs. Simulations were run via the stiff backward differentiation solver included in Octave, with analytically computed Jacobian functions and a maximum time step of 10μs. The model outputs were transformed into trains of spike times identified when the membrane voltage V_m crossed −25 mV with positive slope.

Bin widths for the linear ISI and spike count methods, were set to 3.0 and 0.5 ms for the RS and IB models respectively. Thus for the RS model, ISIs were assigned to bins with edges: 0, 3, 6 … and so on up to $3K_{Lin}$, the right edge of the bin belonging to the longest observed ISI. For the logarithmic ISI method, κ was set equal to 10 for both models. Thus, ISIs from 1 to 10ms were assigned to ten bins with edges: 1.00, 1.26, 1.58, 2.00, 2.51, 3.16, 3.98, 5.01, 6.31, 7.94 and 10.00ms.

Results

These results are divided into three section. In the first, we provide some theoretical considerations for the differences between linear and logarithmically binned inter-spike interval (ISI) distributions, and explore how the binning affects subsequent direct entropy estimations. In the second section we bin the output spike times of two computational neuronal models according to both the logarithmic and linear ISI methods, and a third, alternate linear method. We show that under the assumption that neuronal activity is independent of its history, firing pattern entropy estimates are most reliable for logarithmically binned ISIs, regardless of the estimation technique employed. In the final section we incorporate history effects into the entropy estimates and find that logarithmically binned ISIs yield the most accurate estimates of the three binning methods for all entropy estimation techniques.

Theoretical Considerations

We begin by assuming that neuronal ISIs are drawn from an arbitrary continuous cumulative distribution function: $F_{ISI}(t) = P(ISI \leq t)$. Introducing the logarithm of time variable $\tau \equiv \log(t)$, we substitute $t = 10^{\tau}$ to express the cumulative distribution as $F_{ISI}(10^{\tau})$. Taking the derivatives of F_{ISI} with respect to t and τ separately, yield:

$$\frac{d}{dt}(F(t)) = \frac{dF}{dt} \quad \text{and} \quad \frac{d}{d\tau}(F(10^{\tau})) = \frac{d}{d\tau}(10^{\tau})\frac{d}{dt}(F(t)) = ln(10)10^{\tau}\frac{dF}{dt} = ln(10)\left(t\frac{dF}{dt}\right)$$

which constitute the linear and logarithmic probability density functions (PDFs), respectively. Note the logarithmic PDF is simply the linear PDF times t, scaled by fixed gain: ln(10).

To illustrate this relationship with some examples, we show the PDFs (Fig. 1a) of four functions that have been used to approximate ISI distributions: power law, exponential, log normal, and gamma distributions (Table 1). Parameters were chosen to yield a spread of average ISIs across the four distributions (Table 2). Note the differences between the linear PDFs plotted on a logarithmic abscissa as presented typically in the literature (Fig. 1a, middle), and the truly logarithmic PDFs (Fig. 1a, right). By approximating these PDFs with discrete probability mass functions (Fig. 1b), we see that unlike the linear case, the logarithmic probability functions have equal bin widths in the logarithmic space, making their visual representation easier to interpret.

Because the four example distributions have substantial positive skewness – i.e., a thick tail extending toward long ISIs (Fig. 1A, left) – their logarithmic PDFs are more uniformly distributed than their linear PDFs. This increased uniformity follows from the above equation: the logarithmic PDF is simply a scaled version of the linear PDF times t, increasing the probability density in the long tail of large ISIs. The relationship is easy to validate when the PDFs are integrated into an equal number of bins, K=40, such that p_k is the integral of the PDF over the k^{th} bin (Fig. 1b). If a distribution were uniform, the probability of each event would be the reciprocal of the number of bins: $p_k = K^{-1}$, for all k. The closer a distribution is to uniform, the higher its entropy (see Supp. 1). The logarithmic representations of our example functions have fewer high probability ($p_k \gg K^{-1}$) and low probability ($p_k \ll K^{-1}$) events, are thus closer to uniform and contain more entropy (Fig. 2a).

Regardless of the method used to construct ISI probability distributions, entropy estimates are victim to two types of error: bias and standard deviation. Bias is the difference between the expected entropy estimate and the true entropy: $E[H_{est}] - H_{true}$. Standard deviation is the square root of the difference from the expected squared estimate and the squared expected estimate: $(E[H_{est}^2] - E^2[H_{est}])^{1/2}$. Entropy estimation techniques differ in the amount of bias and standard deviation they introduce. Techniques that reduce bias will increase standard deviation, and vice versa (Paninski, 2003).

Negative bias hampers most entropy estimation techniques. In particular, the expected value of the direct entropy estimate H_{Dir} can be expressed as the true entropy H_{True} plus a simple bias term (Miller, 1955; Carlton, 1969):

$$E[H_{\mathrm{Dir}}] = H_{\mathrm{True}} - \frac{K-1}{N\,2\ln(2)} + O(N^{-2})$$

where N is the number of samples in the dataset and K the number of bins in the distribution. Since the bias depends on only the samples N and bins K, the absolute biases inherent in the direct estimates are roughly equivalent for linearly and logarithmically binned probability distributions to $O(N^{-1})$.† The relative bias will be generally less for the distribution with the greater true entropy (see Supp. 2). From our example functions, the linear and logarithmic distributions yield roughly equivalent relative biases except for in the power law case (Fig. 2b).

The standard deviation of the direct estimate H_{Dir} can be expressed as (Harris, 1975):

$$STD[H_{\mathrm{Dir}}] = \frac{\left(\sum_{k=1}^{K} p_k \log_2^2 p_k - H_{\mathrm{true}}^2\right)^{1/2}}{N^{1/2}(ln(2))^{1/2}} + \frac{(K-1)^{1/2}}{N(2ln(2))^{1/2}} + O(N^{-3/2})$$

where again, the $O(N^{-1})$ term is independent of the distribution. Thus we focus on distributions that reduce the $O(N^{-½})$ term. This term is smaller for higher entropy distributions, reaching zero for the uniform distribution. When comparing two distributions with equal N and K, the

†A straight forward expansion of the bias to higher order: $E[H_{\mathrm{Dir}}] = H_{\mathrm{True}} - \frac{K-1}{N\,2\ln(2)} + \frac{1-\sum_{k=1}^{K} p_k^{-1}}{N^2\,12\ln(2)} + O(N^{-3})$ (Harris, 1975) reveals that the $O(N^{-2})$ term does depend on the distribution of p_k, the probability associated with the k^{th} bin. The $O(N^{-2})$ term is always negative, and minimal in the absolute sense for the uniform distribution: $O(N^{-2})_{uniform} = (1-K^2)/(N^2\,12\,ln(2))$. As the distribution diverges from uniformity, the term becomes more negative, reaching a singularity as any p_k approaches zero! Because the logarithmic probability densities will have fewer low probability bins ($p_k \ll K^{-1}$) than the linear, this term reinforces our claim that logarithmically binned ISIs yield superior entropy estimates. However, the actual bias is better behaved than this $O(N^{-2})$ term suggests (Paninski, 2003), misleading our analytic exploration of the higher order term. See the supplemental material for a more thorough discussion.

one with more entropy will have less standard deviation (see Supp. 3). Because logarithmic ISI distributions have more true entropy, their estimates have less standard deviation (Fig. 2b).

Entropy is an ideal quantification of disorder, or variability. As such, we would like our estimates of entropy, measured in units of bits per spike, to be invariant to changes in neuronal firing rate that do not affect neuronal firing pattern variability. We ranged the rate parameters of the four example functions to yield ISIs from 3 to 300 ms (Table 3), while holding the variability constant. Entropy estimates from the linear distributions increased with increasing ISI (Fig. 3, grey): entropy increases were proportional roughly to the logarithm of the ISI. However, entropy estimates from the logarithmic distributions were independent of average ISI (Fig. 3, black).

Simulations, assuming event independence

We explored the behavior of two computational models under simple conditions. The rheobase, the minimum applied current required to elicit continuous spiking, was found to be I_{rheo} = 4.512867 or 4.601353 μA/cm^2 for RS or IB respectively. A range of input currents (1 pA/cm^2 to 100 μA/cm^2) was added to the rheobase, and the resulting ISIs plotted versus supra-rheobase input strength (Fig. 4). The RS neuron exhibited nearly precise power law behavior (i.e., the logarithm of the input current is inversely proportional to the logarithm of the ISI) over 6 – 8 orders of magnitude (Fig. 4d, top).

The IB neuron exhibited two qualitatively distinct behaviors (Fig. 4b, bottom). For weak inputs of less than ~10μA/cm^2 over rheobase, the neuron fired bursts of high frequency action potentials separated by comparatively long inter-burst intervals. The inter-burst intervals followed power law behavior for 5 – 7 orders of magnitude during which the average intraburst ISI remained constant (Fig. 4d, bottom), although ISIs at the start of each burst were shorter than ISIs at the end (Fig. 4c). For stronger inputs, IB fired regular trains of action potentials with ISIs again following power law behavior (Fig. 4d, bottom) until the model ceased to spike as it entered depolarization block.

Two noise current waveforms that yielded roughly equivalent average firing rates were applied to both cells (Fig. 5): weak noise ($\{<I_{app}>, \sigma_{Iapp}\} \equiv \{5, \sqrt{2.5}\}$μA/cm^2) and strong noise ($\{<I_{app}>, \sigma_{Iapp}\} \equiv \{0, 5\}$μA/cm^2). The noise waveforms were identical in relative frequency amplitudes and phases, such that in the time domain, one waveform was a scaled and mean-shifted version of the other. The spike times for RS and IB models were converted to events of three types as described in Methods: words of six consecutive spike counts in time bins of duration 3.0 or 0.5 ms; ISIs partitioned into linearly spaced bins of width 3.0 or 0.5 ms; and ISIs partitioned into logarithmically spaced bins of width 0.0769 or 0.10 ($\boldsymbol{\kappa} \equiv 13$ or 10). Within a reasonable range, subject to the amount of data in hand, the values of these parameters are unimportant qualitatively. Specific values used here were chosen so that the three entropy methods yielded similar entropy estimates, enabling a fair comparison of bias and variability.

The probability distribution of six-bin words show that RS neurons rarely fired more than two spikes per (6 * 3.0 ms =) 18 ms word (Fig. 5c, left). In contrast, IB neurons were nearly equally likely to fire 1, 2, 3 or 4 spikes per (6 * 0.5ms =) 3 ms word, and even 5 spikes per word was common (Fig. 5c, right). Within each model, behavioral changes in response to the different inputs are visible, but difficult to interpret. Linear ISI distributions were visually independent of stimulus type (Fig. 5d), save for some differences in the distribution tails at high magnification (Fig. 5d, insets). In contrast, the distributions of logarithmic ISIs distinguished clearly between input type. For the RS model, noisier input yielded both shorter and longer ISIs, exhibited by the strong-noise input inducing a left-shift of the diminished peak probability, and a large increase in the tail density with respect to the weak-noise input. A similar gain and shift is seen in the responses of the IB neuron, only that primary peak corresponds to the

intraburst spikes. Thus, we see that the strong-noise input induces faster and less regular intraburst ISIs. In addition, the second peak in the IB distributions, corresponding to the inter-burst intervals, is diminished in peak amplitude and spread markedly to the left, reflecting increasingly irregular inter-burst intervals, including many roughly 10 ms intervals that were never seen in response to the weak-noise input.

The above probability distributions were constructed directly from spike train subsets of durations ranging from 1 second to 1 minute. Firing pattern entropy was estimated via seven different techniques from each probability distribution. The estimates for three of those techniques – direct estimation, Ma lower bound and best upper bound – were plotted against dataset duration (Fig. 6). Estimates constructed from the spike count method are highly dependent upon dataset size (Fig. 6, top). Estimates from the linear ISIs are substantially improved, although the Ma lower bound estimates retain substantial dependence upon dataset duration, and even with 60s worth of data the three techniques yield visibly distinct entropy values (Fig. 6, middle). Estimates from the logarithmic ISI method approach their final value with only a few seconds of data, and all three techniques yield estimates within 1% of each other with only 15s worth of data for both cell models.

The entropy estimate for each of the 21 pairings of classification method and estimation technique from a 60s trial was taken as the full entropy for that pairing. While all methods yielded a downward bias for small datasets, estimates from the direct technique were least dependent upon dataset size for the logarithmic ISI method, and most dependent for the spike count method (Fig. 7a). The biases and standard deviations of each method-technique pairing at 1 and 10 seconds were calculated (Fig. 7b). Bias and standard deviation were greatest for the spike count method and least for logarithmic ISI method. The standard deviation of the full entropy estimates across all techniques was also an order of magnitude smaller for logarithmic ISI than for either of the other methods.

Simulations, incorporating history effects

Estimates made in the previous sections assumed that each ISI, or each six-bin word in the spike count case, was independent of prior events. To obviate that assumption, we accounted for higher order relationships between consecutive events by calculating higher dimensional probability distributions of words consisting of up to 24 bins for the spike count method and of multiple ISI for the other methods. For example, the 2D distribution $P(ISI_1,ISI_2)$ depicts the probabilities of an ISI pair consisting of ISI_1 followed by ISI_2 (Fig. 8). While the linear ISI distribution adequately represents RS behavior, the IB behavior is difficult to interpret. The majority of ISI pairs consisted of ISIs below 1.5 ms. Tails of low probability stretch along both axes complicating any compact visualization of ISI pair behavior. In contrast, the 2D distribution of logarithmically spaced ISIs is easily read and interpreted for both cell types. The RS model fired most often with pairs of ~10ms ISIs, although they ranged from 4–500 ms in a fairly smooth and symmetric fashion. The IB model fired most often with pairs of ~0.8ms ISIs, although the intraburst intervals ranged from 0.4–3 ms. The large lobe of high probability to the right of the purely intraburst mode represents the model entering a burst with a long ISI of 10–300 ms followed by an intraburst ISI peaking from 0.5–1.0 ms. The high probability lobe above the purely intraburst mode represents the model leaving a burst with a final intraburst ISI peaking from 1.0–1.5 ms followed by a post-burst ISI of 10–300 ms. The differences between these lobes depict the increasing ISIs within bursts. Finally, a diffuse smattering of low probability long ISI pairs depict the rare isolated spikes the IB neuron can produce.

Entropies were estimated with all seven techniques for data subsets, for the first four dimensions (i.e. words consisting of up to four consecutive ISIs) in the ISI methods and 24 bins (i.e., words consisting of up to 24 time bins) in the spike count method. Linear least-

squares fits to the estimates versus the reciprocal of their dimension, or the words-per-spike scaled equivalent in the spike count case, were calculated. The zero crossing of each linear fit was taken as the estimate of the true entropy, for that method-technique pairing. The direct technique showed the greatest dependence upon dataset size for the spike count method (Fig. 9). The number of bins included for spike count method extrapolation was varied in an ad hoc manner to yield monotonicity of the bias with dataset duration. While not explored in detail, this required peculiarity makes systematic estimation of large numbers of cells and/or inputs impractical with the spike count method. In contrast, both ISI methods exhibited a consistent bias that dropped off predictably with increasing dataset size (Fig. 10a).

The same high dimensional estimates and their extrapolations were performed for all 21 method-technique pairings for several dataset durations (Fig. 10b). Calculated from 30s of data, the logarithmic ISI method had the least bias for each techniques in the 1D, 2D, 3D and extrapolated cases. In the worst case of 3D distributions, there was only a 5% bias in estimates made from the logarithmic ISI method, while the linear ISI and spike count methods yielded biases of 8% and 35%, respectively.

Discussion

The firing pattern entropy of a neuron bounds the amount of information that neuron can possibly transmit. Only a few years after Shannon's hallmark information papers (Shannon & Weaver, 1949), physiologists had adapted the theory to analyze neuronal activity (MacKay & McCulloch, 1952). However, despite major advances in entropy estimation techniques (e.g., Treves and Panzeri, 1995; Strong et al., 1998; Paninski, 2003) and substantial insights information theory has helped illuminate (e.g., Reinagel et al., 1999; Koch et al., 2006), its application to neuroscience remains a niche field. While there may be many justifiable reasons for neuroscientists to avoid information theory, we believe the primary reason that entropy measures are not more widely used is a perception that information theoretic results are less intuitive and more difficult to interpret than results from simpler variability statistics, including coefficients of variation (CV) or coherences, and ad hoc measures such as burst indexes. While these other measure have their utility, entropy exhibits some preferable characteristics in many situations. For example, ISI CV is used frequently to assess ISI variability (e.g., Dorval & White, 2006), but the results can be misleading particularly in the presence of strong neural rhythms. A neuron that generally fires in phase with a rhythm but occasionally skips cycles could have a very high CV even though its firing pattern is extremely predictable. Not fooled by cycle skipping periodicity however, comparably low entropy estimates would correctly signify a high degree of predictability. Indeed, information theory was founded upon the proof that entropy is an optimal measure of signal variability (Shannon & Weaver, 1949).

We have presented this work to illustrate the bias and standard deviation improvements provided to all estimation techniques by using the probability distribution of the logarithm of ISIs. Intricate comparisons aside, previous work by others suggested that logarithmic ISI distributions are easily interpreted for neurons with disparate spiking patterns (Selinger et al., 2007). In fact, even in response to constant suprathreshold inputs, resulting ISIs are more intuitively described by their logarithm due to the power law relationship between input current and firing rate (Fig. 4). We showed that this ease of interpretation extends to the 2D distribution of ISI pairs (Fig. 8). The relatively small bias and standard deviation of firing pattern entropy estimates from the logarithmic distributions lend credence to the intuition that the true nature of the firing pattern can be garnered more readily from logarithmically, as opposed to linearly, spaced probability distributions.

Of note, while all seven estimation techniques would eventually converge to the same entropy value for enough data, each classification method will yield different entropy estimates. The

binning of spike times is necessarily a lossy compression. After classification, whether we take the spike counts, linear ISIs or logarithmic ISIs, we could not invert them to return to the exact spike times. One question of interest is what information is thrown out by each classification method? For the spike count and linear ISI methods, knowledge of the relative spacing of very short ISIs is lost. For our RS model with 3.0 ms bins for example, ISIs of 3.1 ms are classified into the same bin as ISIs of 5.9 ms. For the IB model during a fairly typical burst, all intraburst ISIs are between 0.5 and 1.0 ms, which, with spike count classification and our 0.5 ms bin size, yields strings in the vicinity of a burst of mostly ones with a few zeros whose placement is more dependent upon the relative phase of the bins with respect to the spikes than on the ever increasing intraburst ISIs. While both of these problems could be overcome by decreasing the bin size, doing so drastically increases the amount of data required to achieve a good estimate, and further diminishes the impact of long ISIs. Logarithmic ISI classification is also non-invertible, but a very different sort of information is removed: the differences between long intervals of similar relative durations. For example, in our IB model with 10 bins per ISI decade, inter-burst intervals of 101 ms are classified in the same bin as intervals of 125 ms. Is that difference important to a bursting neuron, or is it more important to have a robust estimate of how often such 101–125 ms ISI occur?

While that question remains to be answered, if the example distributions we used to illustrate the linear versus logarithmic representations (Fig. 1) are in fact reasonable models of neuronal activity, the nervous system would be more robust to noise and more energy efficient if it computed information via logarithmically binned ISIs. There is a fundamental difference in the underlying assumptions of the linear and logarithmic ISI methods. The linear ISI method assumes that absolute time differences of equal duration are all of equivalent importance: that the 1ms difference between inter-burst intervals of 100 and 101 ms is as significant biologically as the difference between intraburst intervals of 2 and 3 ms. In contrast, the logarithmic ISI method assumes that relative time differences of equal proportion are of equivalent importance: that the 50% difference between inter-burst intervals of 100 and 150 ms is as significant biologically as the difference between intraburst intervals of 2 and 3 ms.

In support of the relative time difference approach, conductance gates recover according to a variety of exponential mechanisms often combing to exhibit power law behavior over a wide range of input strengths (Gilboa et al., 2005). Power law responses are optimally sensitive to relative differences in input strength. Indeed, entropy estimates from the power law distribution were most improved of all by the logarithmically versus linearly binned ISIs (Fig. 2, top). Furthermore, logarithmic binning of ISI distributions enable entropy estimates to be independent of firing rate (Fig. 3). This satisfying result may enable neurons to encode and decode the same information presented at different rates, so called time-warp invariance. A number of studies have verified time-warp invariance, particularly in the context of auditory processing, at different neurological scales: from human processing of variable rate speech (Miller et al., 1984; Newman & Sawusch, 1996) to courtship song detection in the grasshopper (von Helversen, 1972; Gollisch, 2008).

We have shown that, at least for the neuronal models used here, entropy estimates from the distribution of logarithmic ISIs are less biased with smaller standard deviations for limited data sets than either previously published classification method. The 1D and 2D probability distributions of the logarithm of ISIs are easy to generate and intuitive to interpret. Summarizing the disorder, and therefore the information transmitting capacity, the firing pattern entropy is preferable to other statistics and straight-forwardly calculated from these distributions via the direct entropy estimation technique. Preferable estimation techniques, e.g. best upper bound (Paninski, 2003) or Chao-Shen (Chao & Shen, 2003), are provided with the free and open source Spike Train Analysis Toolkit (http://neuroanalysis.org/toolkit), a

compilable plug-in for Matlab™ or Octave. The computational models and all other analysis code used in this work are available from the author upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

First, we thank William Bialek for helpful discussions that precipitated this work. We thank David Goldberg and Daniel Gardner for their writing and maintaining the Spike Train Analysis Toolkit, and especially for helping us configure the toolkit on our computers. We thank John W. Eaton, David Grohmann and Paul Kienzle for their ongoing work in the development of Octave, and in particular for the debugging required to get the Spike Train Analysis Toolkit to compile under Octave. Finally, we thank Warren Grill for directed advice concerning this work, and guidance during its development. This work was supported by funding from the National Institutes of Health, K25-NS0535444 (A.D.D.).

References

Rieke, F.; Warland, D.; de Ruyter van Steveninck, R.; Bialek, W. Spikes. Cambridge, MA: MIT Press; 1997.

Shannon, CE.; Weaver, W. The Mathematical Theory of Communication. Urbanan, IL: University of Illinois Press; 1949.

MacKay DM, McCulloch WS. The limiting information capacity of a neuronal link. Bull Math Biophys 1952;14:127–135.

Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W. Entropy and information in neural spike trains. Phys Rev Lett 1998;80:197–200.

Paninski L. Estimation of entropy and mutual information. Neural Comput 2003;15:1191–1253.

Rieke F, Warland D, Bialek W. Coding efficiency and information rates in sensory neurons. Europhys Lett 1993;22:151–156.

Dayan, P.; Abbott, LF. Theoretical Neuroscience. Cambridge, MA: MIT Press; 2001.

Sigworth F, Sine SM. Data transformations for improved display and fitting of single-channel dwell time histograms. Biophys J 1987;52:1047–1054. [PubMed: 2447968]

Newman MEJ. Power laws, Pareto distributions and Zipf's law. Contemp Phys 2005;46:323–351.

Selinger JV, Kulagina NV, O'Shaughnessy TJ, Ma W, Pancrazio JJ. Methods for characterizing interspike intervals and identifying bursts in neuronal activity. J Neurosci Methods 2007;162:64–71. [PubMed: 17258322]

Ma S. Calculation of entropy from data motion. J Stat Phys 1981;26:221–240.

Treves A, Panzeri S. The upward bias in measures of information derived from limited data samples. Neural Comput 1995;7:399–407.

Miller, G. Note on the bias of information estimates. In: Quastler, H., editor. Information Theory in Psychology II-B. Glencoe, IL: Free Press; 1955.

Carlton AG. On the bias of information estimates. Psychol Bull 1969;71:108–109.

Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. London: Chapman & Hall; 1993.

Wolpert DH, Wolf DR. Estimating functions of probability distributions from a finite set of samples, Part 1: Bayes estimators and the Shannon entropy. arXiv. 1994arXiv>comp-gas:9403001.

Wolpert DH, Wolf DR. Estimating functions of probability distributions from a finite set of samples. Phys Rev E 1995;52:6841–6854.Erratum in Phys Rev E 54:6973.

Chao A, Shen T. Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. Environ Ecol Stat 2003;10:429–443.

Izhikevich, EM. Dynamical Systems in Neuroscience. Cambridge, MA: MIT Press; 2007.

Harris, B. The statistical estimation of entropy in the non-parametric case. Wisconsin University Madison, Mathematics Research Center; 1975. p. 318

Reinagel P, Godwin D, Sherman SM, Koch C. Encoding of visual information by LGN bursts. J Neurophysiol 1999;81:2558–2569. [PubMed: 10322089]

Koch K, McLean J, Segev R, Freed MA, Berry MJ, Balasubramanian V, Sterling P. How much the eye tells the brain. Curr Biol 2006;16:1428–1434. [PubMed: 16860742]

Dorval AD, White JA. Synaptic input statistics tune the variability and reproducibility of neuronal responses. Chaos 2006;16:26105.

Gilboa G, Chen R, Brenner N. History-dependent multiple-time-scale dynamics in a single-neuron model. J Neurosci 2005;25:6479–6489. [PubMed: 16014709]

Miller J, Grosjean F, Lomanto C. Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. Phonetica 1984;41:215–225. [PubMed: 6535162]

Newman RS, Sawusch JR. Perceptual normalization for speaking rate: Effects of temporal distance. Percept Psychophys 1996;58:540–560. [PubMed: 8934686]

von Helversen D. Gesang es mannchens und lautschema des weibchens bei der feldheuschrecke Chorthippus biguttulus (Orthoptera, Acrididae). J Comp Physiol A Neuroethol Sens Neural Behav Physiol 1972;81:381–422.

Gollisch T. Time-warp invariant pattern detection with bursting neurons. New J Phys 2008;10:15012.

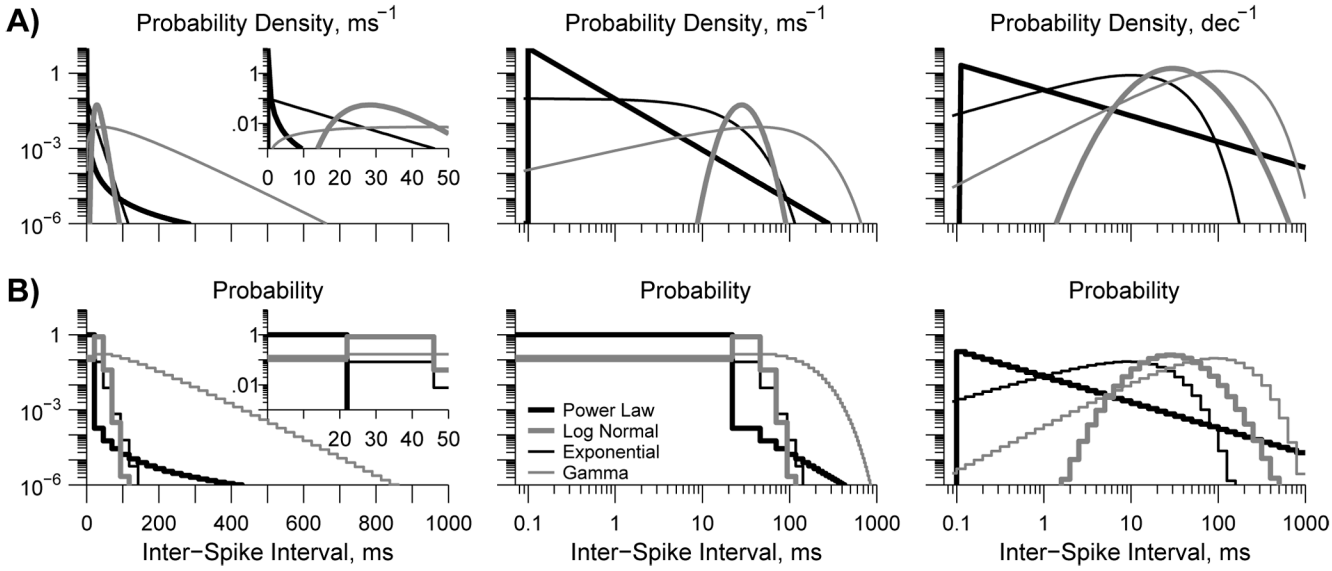

Figure 1. ISI probability distributions of 4 renewal models

a) Three representations of the probability densities of the four distributions listed in Table 1, with parameters from Table 2. **b**) Discrete, probability mass function versions of the probability densities in the same columns. *Left*) Traditional probability density and mass functions, found by taking the derivative of the cumulative distributions with respect to time. *Center*) Traditional probability density and mass functions, but plotted on a logarithmic abscissa. *Right*) Logarithmic probability density and mass functions, found by taking the derivative of the cumulative distributions with respect to the logarithm of time.

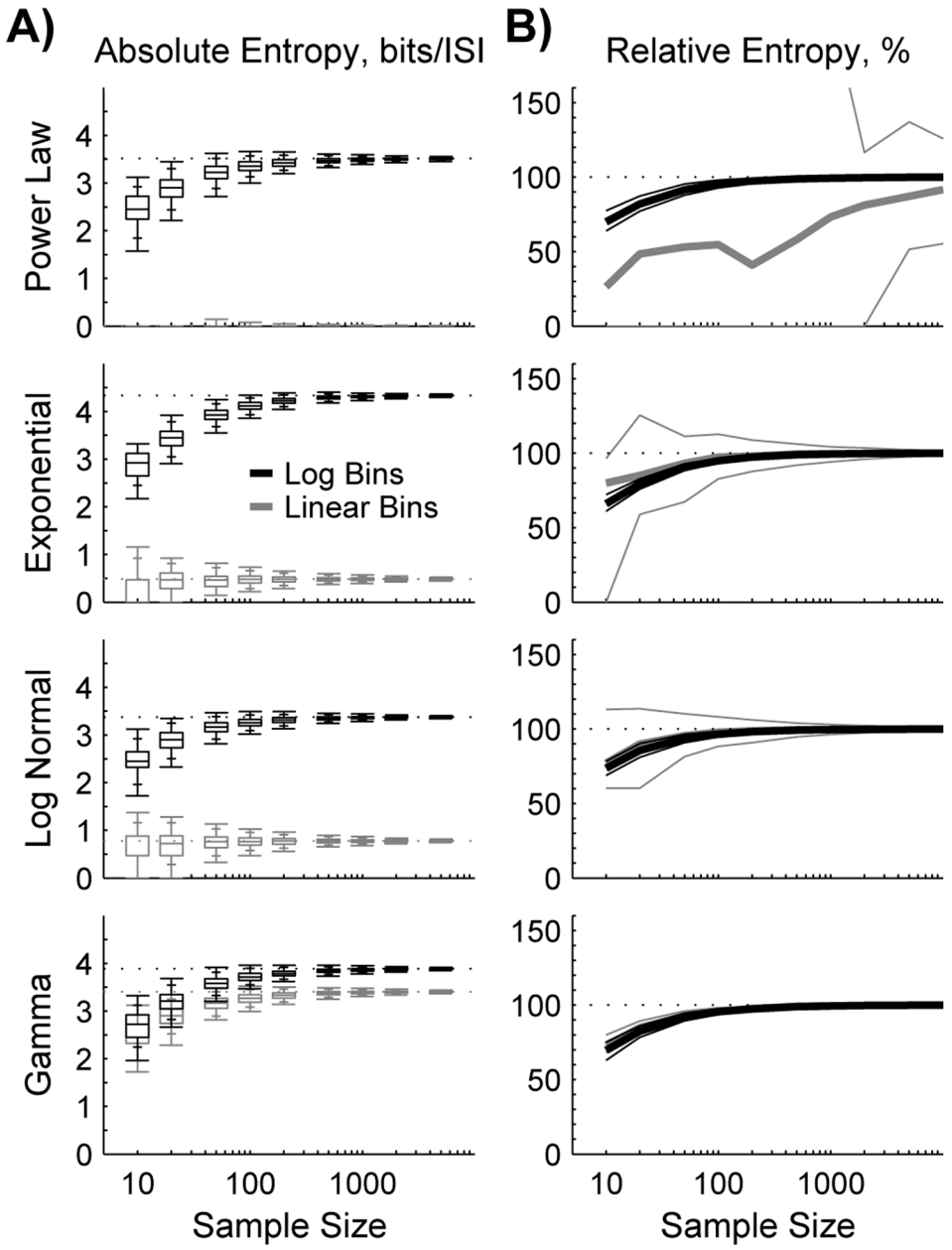

Figure 2. Entropy estimates of 4 renewal models as a function of sample size

a) Entropy estimates from data drawn from the four functions displayed in figure 1, after partitioning the data into 40 bins linearly from 0 to 1000 ms (*grey*) or logarithmically from 0.1 to 1000 ms (*black*). Values are plotted as the entropy in bits per ISI as a function of the data set size, i.e., the number of random draws from the ideal distribution. Results are from 10,000 trials at each condition, and presented as bars depicting the median plus 25 – 75% confidence intervals, with hashes at the 5 – 95% and 1 – 99% confidence intervals. Dashed lines show the true entropy of each distribution. In all cases the logarithmic distributions had more entropy. **b**) Relative entropy estimates, from the same data as in a, reported as means (*thick*) and means ± standard deviations (*thin*) across the 10,000 trials. For the power law case, the only in which the relative biases were not roughly equivalent, the linear distribution had the larger negative

bias (*top*). In all cases, the linear distributions yielded larger relative standard deviations, or coefficients of variation, independent of sample size.

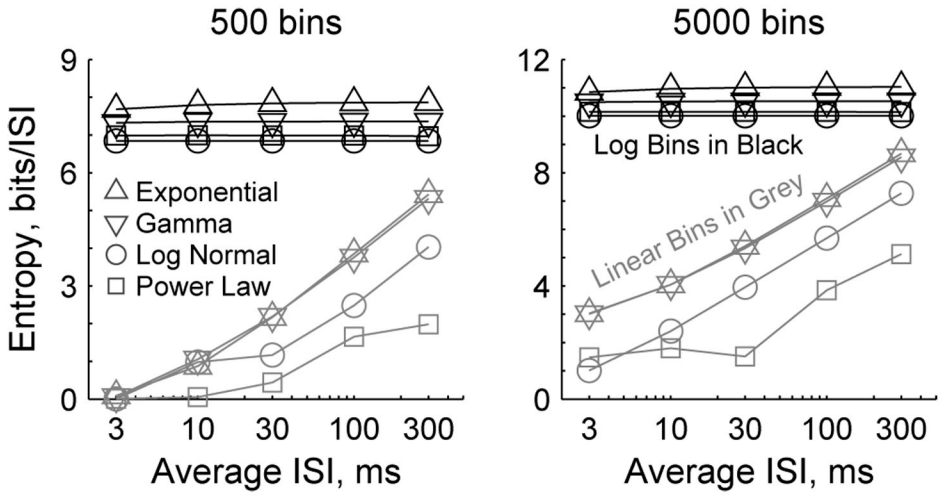

Figure 3. Entropy measures of 4 renewal models as a function of rate

a) Entropies from distributions of the four functions displayed in figure 1, plotted as a function of the average ISIs that resulted from varying the rate parameters as listed in Table 3. Entropies were calculated analytically from the ideal distributions and the number of bins, either 500 (*left*) or 5000 (*right*), partitioned linearly from 0 to 10,000 ms (*grey*) or logarithmically from 0.1 to 10,000 ms (*black*). Entropies from the linear bins increased in rough proportion to the logarithm of the average ISI. Entropies from the logarithmic bins were independent of ISI.

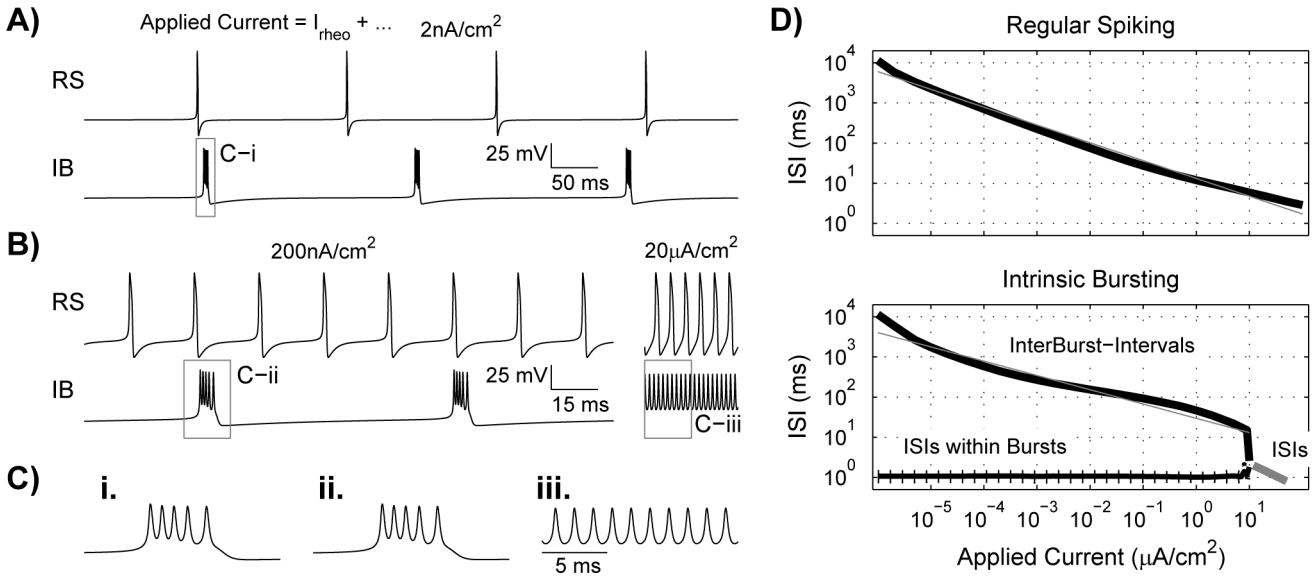

Figure 4. Two model neurons exhibit power law responses to applied currents

a,b) Membrane potential traces of RS (*top*) and IB (*bottom*) model neurons in response to applied currents of 0.002, 0.2 and 20 μA/cm^2 over rheobase. Note the time scale bar is 50 ms in *a* but only 15 ms in *b*. **c**) Magnified views of the grey boxes in *a* and *b* depict IB burst shape (*i & ii*) and short ISIs (*iii*). **d**) Time between spikes for RS (*top*) and IB (*bottom*) in response to supra-rheobase input current. For RS (*thick black line*), the logarithm of ISI was inversely proportional to the logarithm of current (*thin grey line depicts best fit power law*). For IB, the logarithm of the inter-burst interval (*thick black line*) was inversely proportional to the logarithm of current (*thin grey line depicts best fit power law*) until, at ~10 μA/cm^2, IB begins to fire regularly spaced action potentials (*thick grey line*). In the bursting region, the average intraburst ISI was constant (*thin black line*) although individual ISIs varied (*standard deviation bars*).

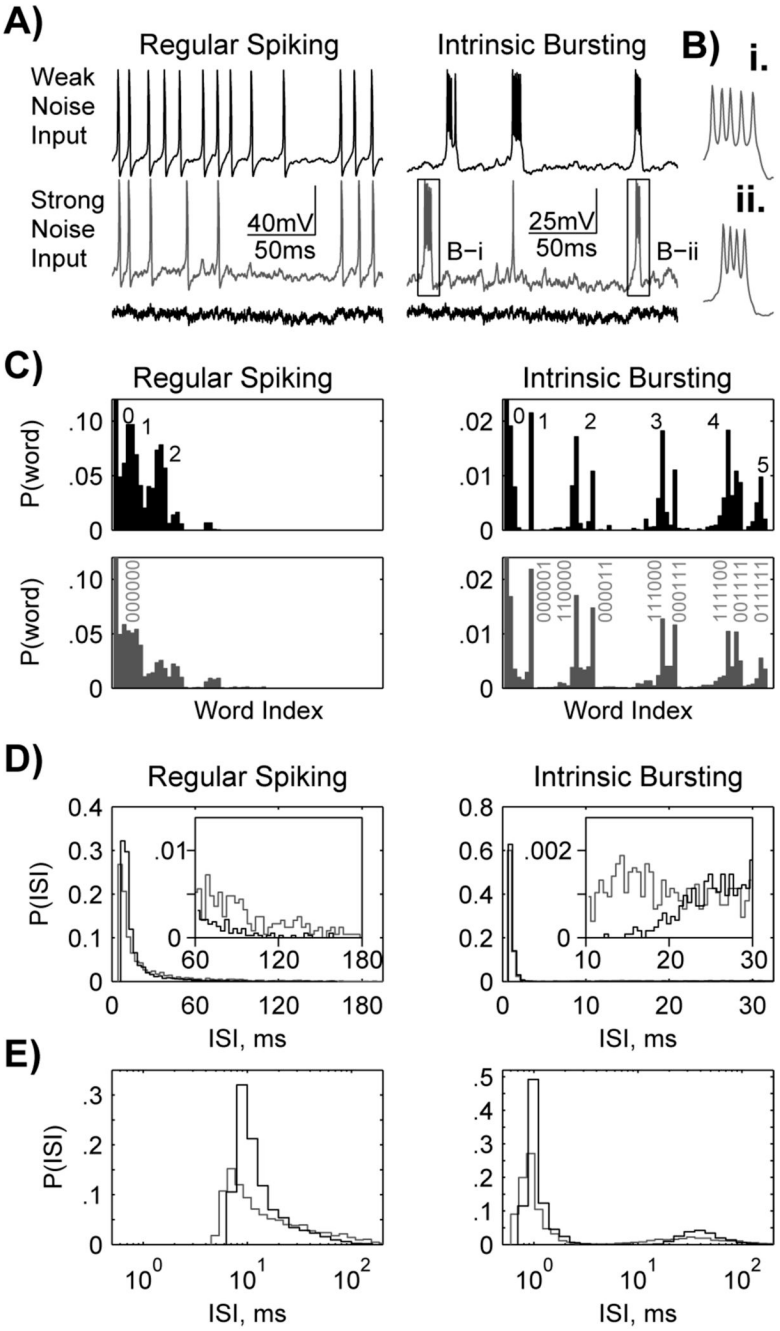

Figure 5. Model neurons respond to weakly and strongly noisy inputs

a) Membrane potential traces of RS (*left*) and IB (*right*) model neurons in response to weakly (*top, black*) and strongly (*middle, grey*) noisy currents. Each current (*bottom*) had the same shape, but with different amplitudes and offsets: $\{<I_{app}>, \sigma_{Iapp}\} \equiv \{5, \sqrt{2.5}\}$ and {0, 5} μA/cm^2, for weak and strong, respectively. **b**) Magnification of the grey boxes in *a* illustrate that IB intraburst intervals and the spikes-per-burst vary between bursts. **c**) Probability distributions of words constructed from the spike counts in six consecutive bins of duration 3.0 or 0.5 ms for RS and IB respectively, in response to weakly (*top, black*) and strongly (*bottom, grey*) noisy inputs. Single digit numbers (*top, black*) denote the number of spikes in each word in the distribution modes beneath them. Six digit numbers (*bottom, grey*) denote the spike count in

each of the six consecutive bins of the large peaks to which the numbers correspond, e.g., the final peak in the IB response to strong noise labeled '011111' represents a no spike bin, followed by five consecutive spike bins. Probability of the first word (000000) is off-scale for ease of viewing. **d**) Probability distributions of ISI as a function of linear time, partitioned into bins of 3.0 (*left, RS*) or 0.5 ms (*right, IB*) width. The RS primary peaks overlap substantially and the IB intraburst interval peaks are essentially identical, and much larger than any other bin, making discrimination between input types difficult. *insets*) Magnified views of the distribution tails show slight differences in the linear ISI distributions. **e**) Probability distributions of ISI as a function of logarithmic time, partitioned into 13 (*left, RS*) or 10 (*right, IB*) bins per ISI decade. The RS model fires both shorter and longer ISIs in response to the strongly noisy input (*grey*) compared with the weakly noisy input (*black*). The IB model fires shorter intraburst intervals with previously unseen short duration (~10ms) inter-burst intervals in response to strongly noisy input (*grey*).

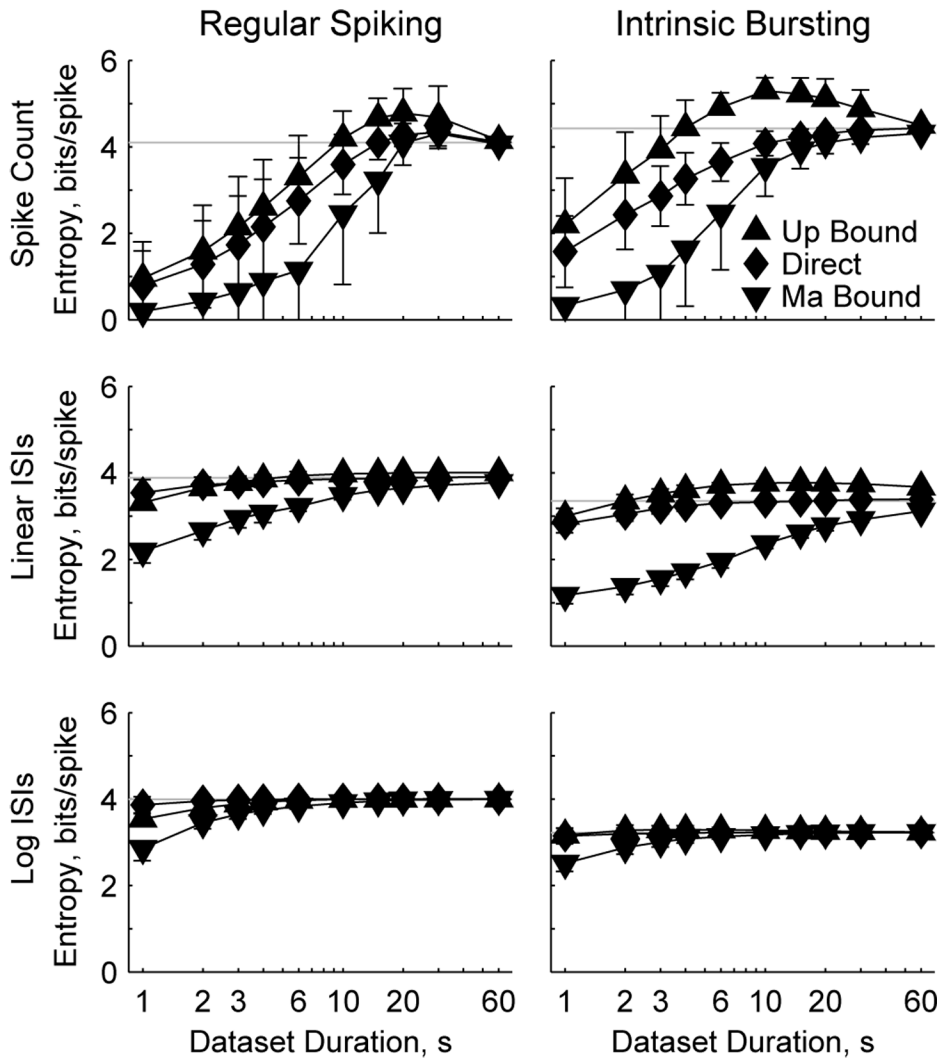

Figure 6. Examples of finite dataset estimation bias depending upon spike train classification method

Entropy was estimated from the probability distributions described in Fig. 5, for data subsets of both neuronal models responding to highly noisy applied current. Repeated estimates were made for all subsets of fixed duration, and combined to yield a mean and standard deviation for each data subset duration. The mean value of each estimation technique with standard deviation is plotted versus data subset duration for the spike count (*top*), linear ISI (*middle*) and logarithmic ISI (*bottom*) methods, for RS (*left*) and IB (*right*). Grey lines mark the average of the three estimation techniques at 60 s.

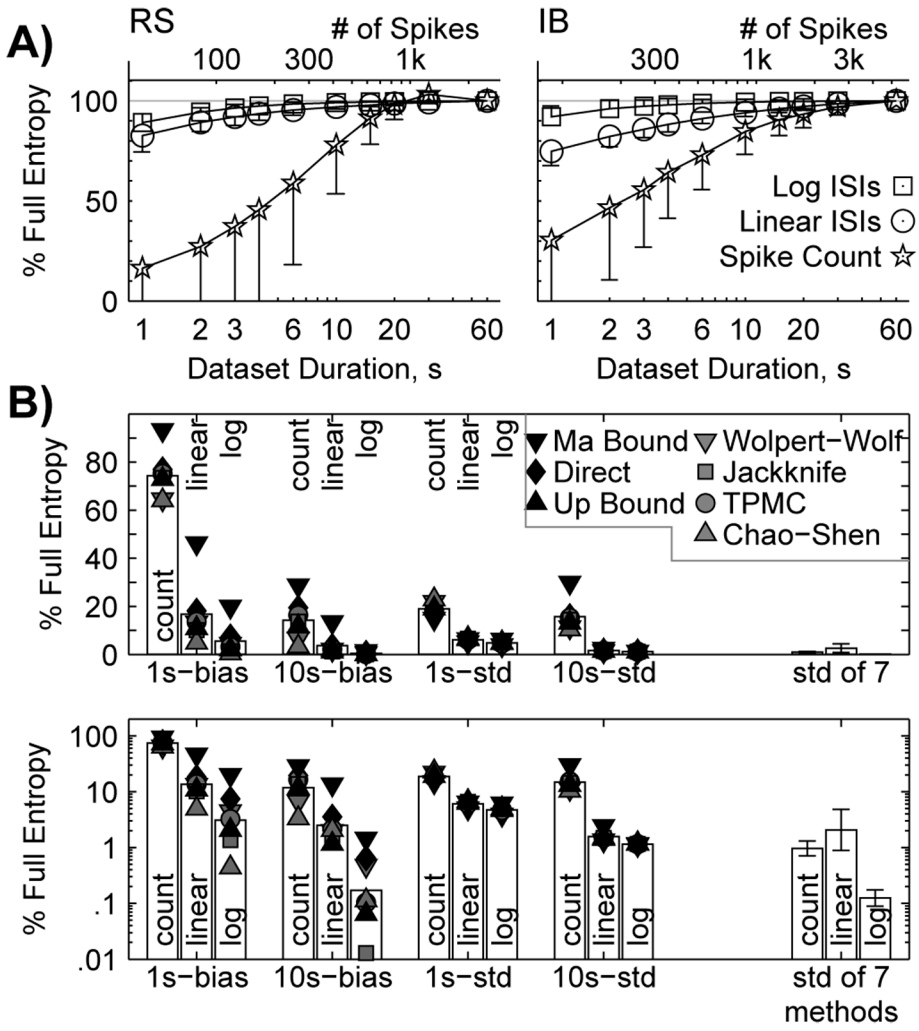

Figure 7. Finite dataset entropy estimation errors depend on spike train classification method

Each entropy estimate from 60 s of data for the 21 pairings of classification method and estimation technique was taken as the full entropy for that pairing. Except for the final far right three bars in *b*, all summary statistics are average values from the four simulations run: two cell models responding to two noisy inputs. **a**) Percent of the full entropy as a function of data subset duration (*bottom axis*) and number of spikes (*top axis*) for each of the three classification methods paired with the direct estimation technique. All three show a downward bias for small data set size, but the effect is greatest for spike counts and least for logarithmic ISIs. **b**) Average bias and standard deviation as a percent of the full entropy for each of the 21 pairings for 1 and 10 s duration datasets with linear (*top*) and logarithmic (*bottom*) ordinates. Note the spike count bias is 2 to 10 times the linear ISI bias which is 2 to 10 times the logarithmic ISI bias for all estimation techniques. Both ISI methods show a marked decrease in standard deviation from the spike count method. The three bars at far right depict the average standard deviation of the estimates from all techniques with 60 s of data. The error bars depict the standard deviation across the four datasets. Standard deviations from the logarithmic ISI method are an order of magnitude smaller than from the other methods.

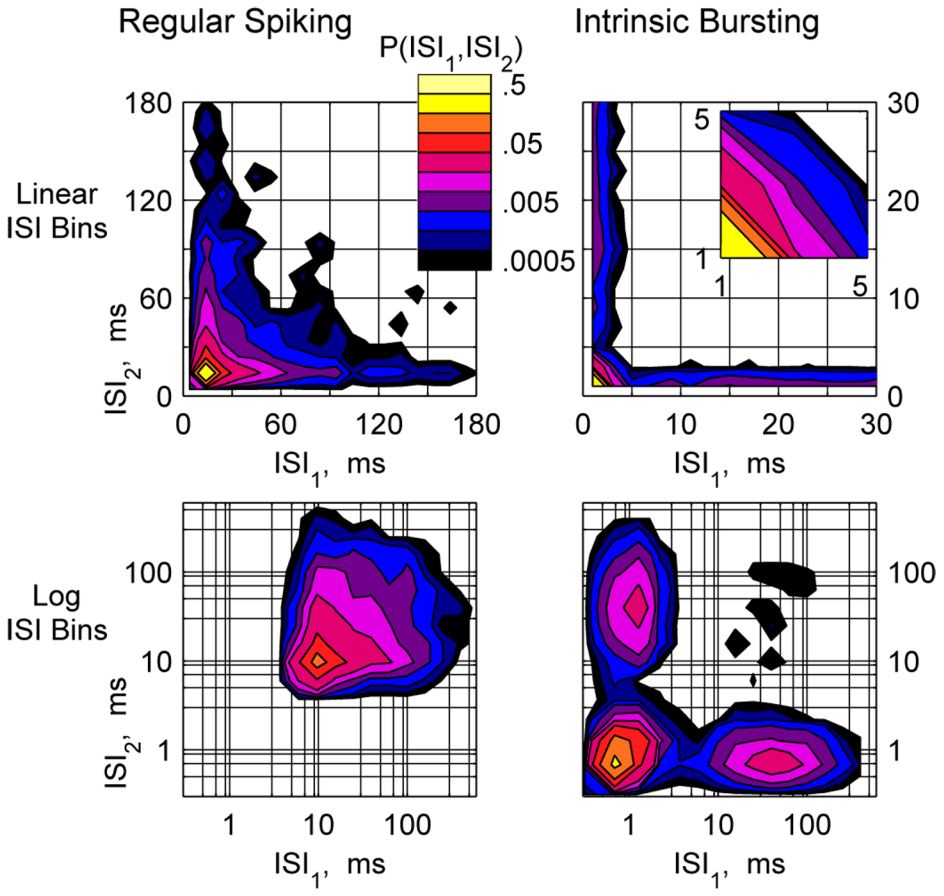

Figure 8. Probability distributions of ISI pairs

The probability of ISI pairs, ISI_2 following ISI_1, for the RS (*left*) and IB (*right*) models in response to the strongly noisy input. Note the logarithmic color bar applies to all four plots. With linear ISI bins (*top*) the IB behavior is difficult to capture pictorially, because the vast majority of ISI pairs are in the sub-1.5 ms regime (*inset*). With logarithmic ISI bins (*bottom*), a concise representation of these disparate cell behaviors can be cleanly presented on axes of the same scale. The RS model fires most often with pairs of ~10ms ISIs, varying smoothly from 4–500 ms. The IB model fires most often with intraburst ISI pairs of ~0.8 ms, varying from 0.4–3 ms. The mode to the right represents burst initiation with an ISI of 10–300 ms and an intraburst ISI which peaks from 0.5–1.0 ms. The mode above the burst mode represents burst completion with a final intraburst ISI peaking from 1.0–1.5 ms and a post-burst ISI of 10–300 ms.

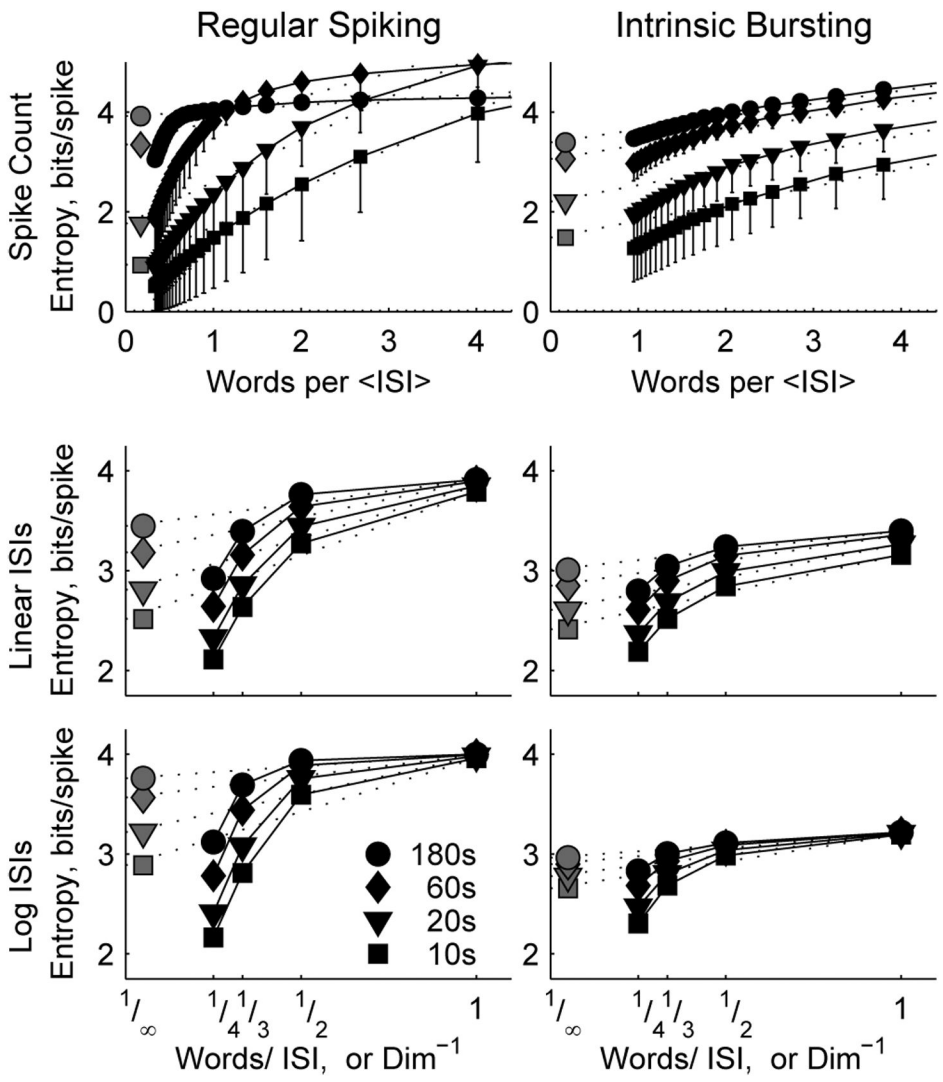

Figure 9. Examples of finite dataset estimation bias after incorporating history effects

The direct entropy estimation technique was applied to the probability distributions of words of up to 4 ISIs or up to 24 time bins – analogous to those described in figure 5 without history effects for only 1 ISI or 6 time bins – for data subsets of the RS (*left*) and IB (*right*) neuronal models responding to highly noisy applied current. Repeated estimates were made for all subsets of fixed duration, and combined to yield a mean and standard deviation for each data subset duration. The mean value of each estimation technique with standard deviation is plotted versus the reciprocal of the dimension (i.e., one divided by the number of ISIs) for the logarithmic (*bottom*) and linear (*middle*) ISI methods, or the firing rate adjusted, reciprocal dimension equivalent for the spike count method (*top*). Least squares fits were extrapolated to zero for all cases and considered the best estimate of the true entropy (*grey symbols*).

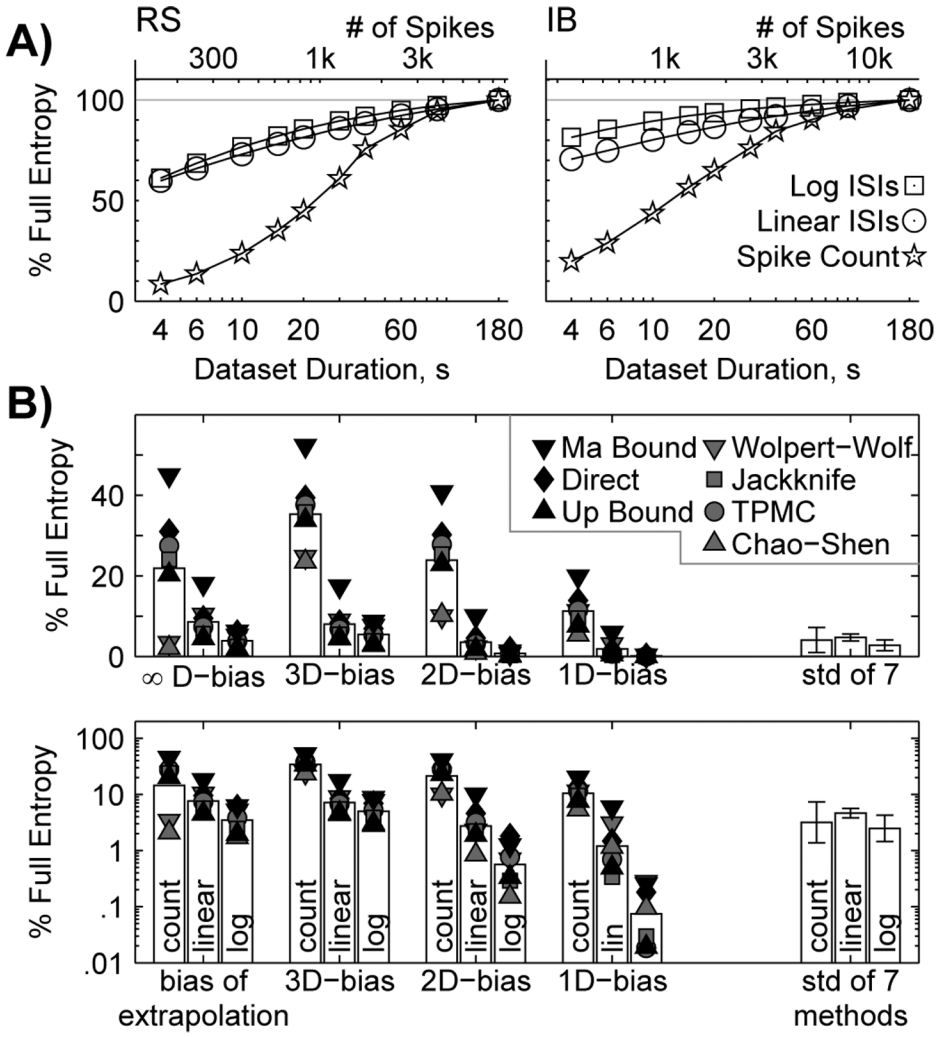

Figure 10. Finite dataset estimation errors depend on spike train classification method even after incorporating history effects

The entropy, accounting for history effects, estimated from 3 minutes of data for each of the 21 pairings of classification method and estimation technique was taken as the full entropy for that pairing. Except for the far right three bars in *b*, all summary statistics are average values from the four simulations run: two cell models responding to two noisy inputs. **a**) Percent of the full extrapolated entropy as a function of data subset duration (*bottom axis*) or number of spikes (*top axis*) for each of the three classification methods paired with the direct estimation technique. All three showed a downward bias for small data set size, but the effect was most evident for the spike count method. **b**) The bias at 30 s as a percentage of the full entropy values for each of the 21 pairings for the extrapolated, 3D, 2D and 1D entropy estimates on linear (*top*) and logarithmic (*bottom*) ordinates. The three bars at far right depict the average standard deviation of the extrapolated estimates from all techniques with 180s of data. The error bars depict the standard deviations across the four datasets.

Table 1

Equations for example distributions

The four example distributions shown in Fig. 1–Fig. 3 listed as cumulative distributions functions (***F*** $_{ISI}$) and the linear (***dF/dt***) and logarithmic (***dF/dτ***) probability density functions. Function abbreviations: *ln()* is the natural logarithm, *log()* is the base 10 logarithm, *erf()* is the error function, Γ*()* is the regular gamma function, and γ*()* is the incomplete gamma function.

Name		$F_{ISI}(t)$	dF/dt	dF/dτ
Power Law	$t \le t_0$	0	0	0
	$t > t_0$	$1-(t/t_0)^{-\alpha+1}$	$(\alpha-1)t^{-\alpha}\big/t_0^{-\alpha+1}$	$ln(10)\,(\alpha-1)(t/t_0)^{-\alpha+1}$
Exponential	$t \le 0$	0	0	0
	$t > 0$	$1-e^{-\lambda t}$	$\lambda e^{-\lambda t}$	$ln(10)\,\lambda t e^{-\lambda t}$
Log Normal	$t \le 0$	0	0	0
	$t > 0$	$\frac{1}{2}\left(1+erf\left(\frac{log(t)-log(\mu)}{\log(\sigma)\sqrt{2}}\right)\right)$	$\frac{t^{-1}e^{\frac{(log(t)-log(\mu))^2}{-2log^2(\sigma)}}}{ln(10)\,log(\sigma)\sqrt{2\pi}}$	$\frac{e^{\frac{(log(t)-log(\mu))^2}{-2log^2(\sigma^2)}}}{log(\sigma)\sqrt{2\pi}}$
Gamma	$t \le 0$	0	0	0
	$t > 0$	$\frac{\gamma(\xi,\ t/\theta)}{\Gamma(\xi)}$	$\frac{t^{\xi-1}e^{-t/\theta}}{\theta^{\xi}\Gamma(\xi)}$	$ln(10)\,\frac{t^{\xi}e^{-t/\theta}}{\theta^{\xi}\Gamma(\xi)}$

Table 2

Parameters for example distributions

Parameter used to calculate the results in Fig. 1 & Fig. 2. Parameters were set to yield the listed different average ISIs with qualitatively similar variabilities.

Name	Rate Parameter	Variability Parameter	Average ISI
Power Law	$t_0 = 0.1$	$\alpha = 59/29$	3 ms
Exponential	$\lambda = 0.1$	**NA**	10 ms
Log Normal	$\mu = 30$	$\sigma = \sqrt[4]{10}$	30 ms
Gamma	$\theta = 75$	$\xi = 4/3$	100 ms

Table 3

Ranging rate parameters

Rate parameters used to calculate the results in figure 3. Variability parameters were the same as in Table 2. Fractions in the gamma distribution case are left not reduced, to simplify interpretation of their progression.

Name, Rate Parameter	Average ISI				
	3 ms	**10ms**	**30ms**	**100ms**	**300ms**
Power Law, t_0	1/10	1/3	1	3	10
Exponential, λ	1/3	1/10	1/30	1/100	1/300
Log Normal, μ	3	10	30	100	300
Gamma, θ	9/4	30/4	90/4	300/4	900/4