Published in final edited form as:
Neural Comput. 2010 October ; 22(10): 2477–2506. doi:10.1162/NECO_a_00015.

Discrete Time Rescaling Theorem: Determining Goodness of Fit for Discrete Time Statistical Models of Neural Spiking

Robert Haslinger[1,2], **Gordon Pipa**[2,3,4], and **Emery Brown**[2,5]

[1]Martinos Center for Biomedical Imaging, Massachusetts General Hospital, 149 13th Street, Suite 2301, Charlestown, MA 02129, USA

[2]Massachusetts Institute of Technology (MIT), Department of Brain and Cognitive Sciences, 77 Massachusetts Ave., Cambridge, MA 02139, USA

[3]Max-Planck Institute for Brain Research, Department Neurophysiology, Deutschorden-str. 46, 60528 Frankfurt am Main, Germany

[4]Frankfurt Institute for Advanced Studies (FIAS), Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany

[5]Massachusetts General Hospital, Department of Anesthesia and Critical Care, 55 Fruit Street, Boston, MA 02114, USA

Abstract

One approach for understanding the encoding of information by spike trains is to fit statistical models and then test their goodness of fit. The time rescaling theorem provides a goodness of fit test consistent with the point process nature of spike trains. The interspike intervals (ISIs) are rescaled (as a function of the model's spike probability) to be independent and exponentially distributed if the model is accurate. A Kolmogorov Smirnov (KS) test between the rescaled ISIs and the exponential distribution is then used to check goodness of fit. This rescaling relies upon assumptions of continuously defined time and instantaneous events. However spikes have finite width and statistical models of spike trains almost always discretize time into bins. Here we demonstrate that finite temporal resolution of discrete time models prevents their rescaled ISIs from being exponentially distributed. Poor goodness of fit may be erroneously indicated even if the model is exactly correct. We present two adaptations of the time rescaling theorem to discrete time models. In the first we propose that instead of assuming the rescaled times to be exponential, the reference distribution be estimated through direct simulation by the fitted model. In the second, we prove a discrete time version of the time rescaling theorem which analytically corrects for the effects of finite resolution. This allows us to define a rescaled time which *is* exponentially distributed, even at arbitrary temporal discretizations. We demonstrate the efficacy of both techniques by fitting Generalized Linear Models (GLMs) to both simulated spike trains and spike trains recorded experimentally in monkey V1 cortex. Both techniques give nearly identical results, reducing the false positive rate of the KS test and greatly increasing the reliability of model evaluation based upon the time rescaling theorem.

Key words and phrases

Time Rescaling Theorem; Goodness of Fit; Point Processes; Spike Train Analysis; Generalized Linear Model; Conditional Intensity Function; Kolmogorov Smirnov Test

1 Introduction

One strategy for understanding the encoding and maintenance of information by neural activity is to fit statistical models of the temporally varying and spike history dependent spike probability (conditional intensity function) to experimental data. Such models can then be used to deduce the influence of stimuli and other covariates on the spiking. Numerous model types and techniques for fitting them exist, but all require a test of model goodness of fit as it is crucial to determine a model's accuracy before making inferences from it. Any measure of goodness of fit to spike train data must take the binary nature of such data into account. E.g. discretized in time, a spike train is a series of zeros and ones. This makes standard goodness of fit tests, which often rely on assumptions of asymptotic normality, problematic. Further, typical distance measures such as the average sum of squared deviations between recorded data values and estimated values from the model can often not be computed for point process data.

One technique, proposed by Brown and collaborators, for checking the goodness of fit of statistical models of neural spiking makes use of the time rescaling theorem (Brown, 2001). This theorem states that if the conditional intensity function is known, then the interspike intervals (ISIs) of any spike train (or indeed any point process) can be rescaled so that they are Poisson with unit rate, e.g. independent and exponentially distributed. Checking goodness of fit is then easily accomplished by comparing the rescaled ISI distribution to the exponential distribution using a Kolmogorov Smirnov (KS) test. (Press, 2007; Massey, 1951). The beauty of this approach is not only its theoretical rigor, but also its simplicity, as the rescaling requires only the calculation of a single integral. Further, a second transformation takes the exponentially distributed rescaled times to a uniform distribution and the KS test can then be performed graphically using a simple plot of the cumulative density function (CDF) of the rescaled times versus the CDF of the uniform distribution to determine if the rescaled times lie within analytically defined confidence bounds. Due to its many appeals, the time rescaling theorem has been extensively used to test model goodness of fit to spike train data (Frank, 2002; Truccolo, 2005; Czanner, 2008; Song, 2006).

There are however certain neurophysiological situations in which in which the standard time rescaling approach can give misleading results, indicating poor goodness of fit when model fit may in fact be very good. This is a consequence of the practical numerical consideration that when a statistical model is fit to spike data one almost always discretizes time into bins. The time rescaling theorem applies exactly to a continuous time point process, e.g. if we have infinite temporal precision and if the events (spikes) are instantaneous. In a practical neuroscience setting however we usually do not have infinite temporal precision. Firstly, a spike is an event that lasts for a finite (~ 1 msec) period of time and any temporal resolution which is far below this lacks physical relevance.[1] Secondly, from a computational perspective, the fitting of statistical models requires much less computer time when the temporal discretization is coarser. Temporal discretization therefore imposes both physical and practical numerical constraints on the problem.

Often the probability per bin of a spike is always small and the distinction between continuous and discrete time of no concern because the width of a spike is very short compared to the average interspike interval. On the other hand, there are cases for which firing rates can be very high due to strong stimuli and ISIs short due to burst type dynamics and here the the per bin spike probability can be large even at 1 msec resolution or less. Such situations can arise in,

[1]This statement applies if one considers the spike as a *event* as we do here. If one instead is interested in the shape and timing of the spike *waveform*, for example the exact time of the waveform peak, then temporal resolutions of << 1 msec may indeed be physically relevant.

for example, primate visual experiments where neurons can be extremely active (De Valois, 1982; MacEvoy, 2007, and also see the Results section of this paper) exhibiting firing rates of up to 100 Hz or more. In such situations, it is important to ensure that the the rescaled ISIs are still (approximately) exponentially distributed and if not, to determine the correct distribution before performing the KS test.

Our aim in this paper is to develop simple and easily applied goodness of fit tests for the discrete time case. We first restate the standard, continuous time form of the time rescaling theorem for point processes and then demonstrate the discretization problem using a simple homogeneous Bernoulli (discretized homogeneous Poisson) process. We show theoretically that the discrete nature of the Bernoulli process results in first a lower bound upon the smallest possible rescaled ISI and second, because there can be only one spike per bin, a spike probability less than that which would be estimated by a continuous time model. These differences lead to biases in the KS plot, biases which are caused by fundamental differences in the shapes of the geometric and exponential distributions, not by poor spike sampling or poor numerical integration technique. We demonstrate further that these biases persist for more complicated simulated neural data with inhomogeneous firing rates and burst type spike history effects. Interestingly we show that the biases increase when spike history effects are present.

We then propose two computationally tractable modifications to the time rescaling theorem applicable to discrete time data. The first is similar in spirit to a bootstrap and involves direct simulation of confidence bounds on the rescaled ISI distribution using the statistical model being tested. In the second method, by randomly choosing exact spike times within each bin and introducing a correction to the fitted discrete spike probabilities, we define an analytic rescaled time which *is* exponentially distributed at arbitrary temporal discretizations. Use of this analytical method gives results nearly identical to the numerical approach. In this paper we use Generalized Linear Models (GLMs) with logistic link functions (McCullagh, 1989; Wasserman, 2004). However we emphasize that both procedures will apply to any discrete time statistical model of the time varying spike probability, not only GLMs. We demonstrate both approaches using simulated data and also data recorded from real V1 neurons during monkey vision experiments. In all our examples, the KS plot biases are eliminated. Models for which the original KS plots originally lay outside 95% confidence bounds are demonstrated to in fact be very well fit to the data, with the with modified KS plots lying well within the bounds. In addition to providing more accurate statistical tests for discrete time spiking models, our approaches allow for the use of larger time bin sizes and therefore can substantially decrease the computation time required for model fitting.

2 Theory

The time rescaling theorem states that the interspike intervals (ISIs) of a continuous time point process can be transformed, or rescaled, so that the rescaled process is Poisson with unit rate, e.g. the rescaled ISIs are independent and exponentially distributed. This variable transform takes the form

$$\tau_i = \int_{t_{i-1}}^{t_i} \lambda(t|H_t)dt \tag{1}$$

$\{t_i\}$ is the set of spike times and $\lambda(t|H_t)$ is the conditional intensity function: temporally varying and history dependent spike probability. Although we will henceforth drop the H_t in our notation, such conditioning on the previous spiking history is always implied. Intuitively, the ISIs are stretched or shrunk as a function of total spike probability over the ISI interval so that the rescaled ISIs are centered about a mean of 1. Several proofs of this theorem exist, (Brown,

2001). Here we present a simple proof of the exponential distribution of the rescaled ISIs. A proof of their independence is located in Appendix A.

The proof proceeds by discretizing time into bins of width Δ, writing down the probability for each discrete ISI, and then taking the continuous time limit, e.g. $\Delta \rightarrow dt$. The discrete time bins are indexed as k and the bins within which the spikes occur are denoted as k_i. Further we define p_k as the discrete probability mass of a spike in bin k and like $\lambda(t)$ it should be taken as conditionally dependent on the previous spiking history.

The probability of the i-th ISI is the probability that there is a spike in bin k_i given that the preceding spikes were located in bins $k_1, k_2,\ldots,k_{i-1}$.

$$P(ISI_i)=P(k_i|k_1,k_2,\ldots,k_{i-1})=\left[\prod_{l=1}^{L_i-1}(1-p_{k_{i-1}+l})\right]p_{k_{i-1}+L_i} \tag{2}$$

where L_i is defined such that $k_{i-1} + L_i = k_i$. This is simply the product of the probabilities that there are no spikes in bins $k = \{k_{i-1} + 1, k_{i-1} + 2,\ldots k_i - 1\}$ multiplied by the probability that there is a spike in bin $k = k_i$. For simplicity we now drop the i subscripts.

In preparation for taking the small bin size limit we note that when Δ becomes small, so does p e.g. $p << 1$ for all bins. This implies that $1 - p \approx e^{-p}$ allowing the above to be rewritten as

$$P(ISI)=P(k+L)\approx \exp\left[-\sum_{l=1}^{L}p_{k+l}\right]p_{k+L} \tag{3}$$

Note that the upper limit of the sum has been changed from $L - 1$ to L with the justification that we are in a regime where all the p's are very small. We define the lower and upper spiketimes as $t_k = k\Delta$ and $t = t_{k+L} = (k + L)\Delta$, define $\lambda(t_{k+l})$ such that $p_{k+l} = \lambda(t_{k+l})\Delta$ [2] and also define the ISI probability *density* $P(t)$ such that $P(k + L) = P(t)\Delta$. After substituting these into the above equation and converting the sum to an integral we obtain

$$P(t)dt=e^{-\int_{t_k}^{t}\lambda(t')dt'}\lambda(t)dt \tag{4}$$

Consulting equation 1 we note that the integral in the exponent is by definition τ. Further, applying the Fundamental Theorem of Calculus to this integral gives $d\tau = \lambda(t)dt$ [3] Changing variables from t to τ we finally obtain

$$P(\tau)d\tau=e^{-\tau}d\tau \tag{6}$$

[2] $\lambda(t_{k+l}) =< \lambda(t) >_{k+l}$ where the average is taken over the time bin k. This definition holds only when the bin size is very small. We will soon show that for moderately sized bins $p_{k+l} \neq \lambda(t_{k+l})\Delta$ and that this leads to biases in the KS plot.

[3] Specifically,

$$\frac{d\tau}{dt}=\frac{d}{dt}\int_{t_k}^{t}\lambda(t')dt'=\lambda(t) \tag{5}$$

and therefore $d\tau = \lambda(t)dt$.

which is now exponentially distributed and completes the proof.

Although the τ_i can be compared to the exponential distribution, it is useful to note that a second variable transform will make the rescaled ISIs uniformly distributed.

$$z_i = 1 - e^{-\tau_i} \tag{7}$$

General practice is to sort the rescaled ISIs z_i into ascending order and plot them along the y axis versus the the uniform grid of values $b_i = \frac{i - 0.5}{N}$ where N is the number of ISIs and $i = 1, \ldots, N$. If the rescaled ISIs z_i are indeed uniformly distributed then this plot should lie along the 45 degree line. Essentially, the cumulative density function (CDF) of the rescaled ISIs z_i is being plotted against the CDF of the uniform distribution (the b_i's). We show an example of such a plot in Figure 1. Such a plot can be thought of as a visualization of a Kolmogovov Smirnov (KS) test which compares two CDFs and is usually referred to as a "KS plot." Formally we can state the null hypothesis H_0 of this test as follows:

H_0 : Given a model of the conditional intensity function which is statistically adequate, the experimentally recorded ISIs can be rescaled so that they are distributed in the same manner as a Poisson process (exponentially distributed) with unit rate.

Under the null hypothesis the maximum distance between the two CDFs will, in 95% of cases, be less than $\frac{1.36}{\sqrt{N}}$ where N is the number of rescaled ISIs (Brown, 2001; Johnson and Kotz, 1970). Equivalently, the plotted line of rescaled ISIs will lie within the bounds $b_k \pm \frac{1.36}{\sqrt{N}}$ in 95% of cases under the null hypothesis. It should be kept in mind that this is *not* equivalent to saying that the line of rescaled ISIs lying within these bounds implies a 95% chance of the model being correct.

2.1 Temporal Discretization Imposes KS Plot Bias

The time rescaling theorem applies exactly to a point process with instantaneous events (spikes) and infinite temporal precision, e.g. continuous time. As a practical matter when fitting a statistical model one generally discretizes time. For discrete time the integral of equation 1 is naively replaced by a sum

$$\tau_i = \sum_{k=k_{i-1}+1}^{k_i} p_k \tag{8}$$

If $p_k << 1 \; \forall k$, e.g. in situations where either the bin size is very small $\Delta \rightarrow 0$ or the firing rate is very low, the time rescaling theorem will apply approximately even if a discrete time model is used. However, it often happens that p_k is in fact large. For example 50 Hz spiking sampled at 1 msec implies $p \approx 0.05$, and under many conditions the firing rate can be much higher, at least over some subset of the recording, e.g. during bursting. In such cases, the rescaled times τ_i will *not* be exponentially distributed and the KS plot will exhibit significant biases (divergences from the 45% line) even if the discrete time model for p_k is exactly correct. We this demonstrate in Figure 1 where two KS plots generated using the exact same spikes and time varying firing rate are shown, but a temporal discretization was imposed for one of the plots.

These biases originate in two distinct consequences of discretizing a continuous process. First, there is a lower bound upon the smallest possible ISI (one bin) which leads to a lower bound on the smallest possible rescaled time *z*. Second because only a single spike per bin is allowed, using a discrete time model to estimate the firing rate of a continuous time process results in an underestimation of the firing rate. To demonstrate these issues fully we now consider the simple case of a homogeneous Bernoulli process with a constant spike probability $p_k = p$ per bin for which the the CDF of the *z*'s can be calculated analytically and the KS plot determined exactly.

For a discrete time process only a discrete set of ISIs are possible. Specifically $\{n\Delta\}$ where n is an integer greater than zero and Δ is the bin width. In the case of a homogeneous Bernoulli process the rescaled ISIs are $\tau(n) = pn$ and

$$z(n)=1-e^{-pn} \tag{9}$$

and the discrete probability distribution of interspike interval times (and rescaled times) is

$$P_B(n)=(1-p)^{n-1}p \tag{10}$$

As in equation 2, this is merely the product of the probability of no spike for $n-1$ bins followed by the probability of a spike in the last (nth) bin. The 'B' subscript indicates the Bernoulli process. $P_B(n)$ is not an exponential distribution, as would be expected for a homogeneous Poisson process. It is a geometric distribution, although in the limit of small p, it reduces to an exponential distribution. [4] The CDF of this ISI distribution is easily calculated by summing the geometric series and combining terms.

$$\begin{aligned} CDF_B(n)=\sum_{j=1}^{n}P_B(j) &= \tfrac{p}{1-p}\sum_{j=1}^{n}(1-p)^j \\ &= 1-(1-p)^n \end{aligned} \tag{11}$$

To finally get the CDF of the rescaled ISIs *z*, equation 9 is inverted to get $n=-\dfrac{\log(1-z(n))}{p}$ and substituted into equation 11.

$$CDF_B(z)=1-(1-p)^{-\frac{\log(1-z(n))}{p}} \qquad z(n-1)\le z\le z(n) \tag{12}$$

In Figure 2 we use equation 12 to generate the KS plot for various spike per bin probabilities *p*. Even at $p = 0.04$ which would correspond to 40 Hz firing at 1 msec discretization the CDF is highly non-uniform with a step like structure caused by the discrete values which the rescaled ISIs can take. Such "steps" will be smoothed out if a inhomogeneous Bernoulli process is used instead. (see below) There is, however, another more serious divergence from uniformity,

[4] Setting $p = \lambda\Delta$ and $t = n\Delta$

$$P_B(t)=(1-p)^{n-1}p=\frac{\lambda\Delta}{1-\lambda\Delta}(1-\lambda\Delta)^{t/\Delta}\rightarrow\lambda e^{-\lambda t}dt=P_p(t)dt$$

when the limit $\Delta \rightarrow dt$ is taken.

namely *a distinct positive bias at low (close to 0) rescaled ISIs and a distinct negative bias at high (close to 1) rescaled ISIs*. This bias will not disappear if an inhomogeneous Poisson process is used.

The grey band, which is barely visible, is the 95% confidence region of the KS plot assuming 10 minutes of 40 Hz spiking, which translates into 24000 spikes on average. Since the confidence bounds are so close to the 45 degree line, and will be for any spike train with a long recording time and appreciable firing rate, we introduce a new type of plot in Figure 2 that we term a "Differential KS Plot". This is simply a plot of the difference between the distribution we hypothesize that the CDF of rescaled times should follow (in this case uniform) and the CDF of the "experimentally recorded" rescaled ISIs (in this case the rescaled ISIs of the Bernoulli process). E.g.

$$CDF_{hyp}(z) - CDF_{exp}(z) \tag{13}$$

It should be emphasized that the Differential KS plot displays the same information as the KS plot, but does so in a different and more visually accessible manner. The confidence bounds (horizontal dashed lines) are now simply given by $\pm\frac{1.36}{\sqrt{N}}$ where N is again the number of rescaled ISIs. Plotted this way one can clearly see the positive bias at low values of the rescaled ISIs and the negative bias at high values of the rescaled ISIs. We emphasize that since these KS and Differential KS plots are calculated using the exact homogeneous Bernoulli distribution, the biases are not finite sampling effects.

The positive bias at low ISIs is easily understood by noting that the smallest possible rescaled time is not zero but

$$z(1)=1-e^{-p}=p-\frac{p^2}{2}+\ldots>0 \tag{14}$$

What about the negative bias at large (z close to 1) rescaled ISIs ? Consider a homogeneous Poisson process with a firing rate λ. Upon discretizing time into bins of width Δ one might naively expect the probability of a spike per bin to be $p = \lambda\Delta$. However it is in fact slightly less than this as we now show. Assume that there is a spike at time $t = 0$. Then for a homogeneous Poisson process the probability density for the waiting time t_w until the next spike is $\rho(t_w) = \lambda e^{-\lambda t_w}$. Integrating, the probability that the next spike lies within any interval $t < t_w \leq t + \Delta$ can be obtained.

$$P(t<t_w \leq t+\Delta)=\int_t^{t+\Delta}\lambda e^{-\lambda t'}dt'=e^{-\lambda t}(1-e^{-\lambda\Delta}) \tag{15}$$

Defining the bin index n such that $t = (n-1)\Delta$ and discretizing we get

$$\begin{aligned} P(n_w=n) &= e^{-\lambda\Delta(n-1)}(1-e^{\lambda\Delta}) \\ &= [1-(1-e^{-\lambda\Delta})]^{n-1}(1-e^{-\lambda\Delta}) \\ &= (1-p)^{n-1}p \end{aligned} \tag{16}$$

where we have defined $p = 1 - e^{-\lambda\Delta}$ in the last line. Discretizing time transforms the homogeneous Poisson process into a homogeneous Bernoulli process, but with a per bin

probability of a spike $p \neq \lambda\Delta$. In fact, by expanding the exponential as a Taylor series it can be seen that

$$p=1-e^{-\lambda\Delta}=\lambda\Delta-\frac{(\lambda\Delta)^2}{2}+\ldots<\lambda\Delta \tag{17}$$

The continuous Poisson process still has an expected number of spikes per interval of width Δ of $\int_0^\Delta \lambda dt=\lambda\Delta$, but such an interval could have more than one spike in it. In contrast, the discrete Bernoulli process can only have either 0 or 1 spikes per bin. Therefore the per bin "spike probability" p calculated above is not the expected number of spikes of the continuous point process within an interval Δ. It is the expected number of *first* spikes in an interval Δ, which is of course less than the total number of expected spikes. By discretizing into bins any chance of there being more than one spike in a time window Δ has been eliminated.

The breakdown of the first order expansion of the exponent is the source of the negative KS plot bias at high (z close to 1) rescaled ISIs. It is a fundamental consequence of discretizing a continuous time point process and is closely connected how the conditional intensity function is generally defined, that is, as the small bin size limit of a counting process (see (Snyder, 1975) cf. Chapter 5). More specifically the conditional intensity function is the probability density of single spike in an infitesimal interval $[t, t+\Delta)$. As shown above, this probability density is actually $p/\Delta = (1-e^{-\lambda\Delta})/\Delta < \lambda$ and the equality only holds in the limit. Thus p/Δ is not a good approximation for λ when the bin size is too large, and this causes the time rescaling theorem to break down.

2.2 Inhomogeneous Bernoulli Processes

The same positive (negative) bias in the KS plot at low (high) rescaled ISIs remains when the spiking process is not homogeneous Bernoulli. We now we define three inhomogeneous spiking models in continuous time and subsequently discretize them. We use these inhomogeneous discrete time models to simulate spikes and then calculate the rescaled ISIs using the exact discrete time model used to generate the spikes in the first place. The goal is to show that even if the exact discrete time generative model is known, the continuous time rescaling theorem can fail for sufficiently coarse discretizations.

The first model is an inhomogeneous Bernoulli process. One second of the inhomogeneous firing probability is shown in Figure 3 A. The specific functional form was spline based with knots spaced every 50 msec and the spline basis function coefficients chosen randomly. This model firing probability was repeated 600 times for a total of 10 minutes of simulated time. The second and third models were the homogeneous and inhomogeneous Bernoulli models respectively but with the addition of a spike history dependent renewal process shown in Figure 3 B. We used a multiplicative model for the history dependent firing probabilities of the form

$$\lambda(t)=\lambda_0(t)\lambda_{hist}(t-t_{ls}) \tag{18}$$

where $\lambda_0(t)$ is the time dependent firing probability independent of spike history effects and λ_{hist} is the spike history dependent term which is a function of the time since the last spike ($t' = t - t_{ls}$). The functional form of the spike history dependent term was a renewal process, specifically

$$\lambda_{hist}(t') = \frac{1+3e^{-(t'-2)/5}}{1+e^{-4(t'-2)}} \tag{19}$$

where $t' = t - t_{ls}$ is in msec. This form was chosen to mimic a brief refractory period and subsequent rebound. For comparison purposes, all three of these models were constructed so that the mean firing rate remained approximately 40 Hz. Thus the inhomogeneous Bernoulli firing probability had a 40 Hz mean. In the spike history dependent cases, the history independent firing probabilities $\lambda_0(t)$ were adjusted downwards so that when history effects were included the mean firing rate remained approximately 40 Hz. Specifically the history independent firing probability of the homogeneous Bernoulli process was reduced to 29 Hz and a similar reduction was made for the inhomogeneous Bernoulli model.

In Figure 3 we demonstrate the effect on the KS and Differential KS plots when these models are subjected to various temporal discretizations. Specifically we discretized the models at 1, 0.5, and 0.1 msec resolution by averaging $\lambda_0(t)$ over these bin widths, e.g. $p_{k,0} =< \lambda_0(t) >_k$. The spike history dependent term is a function of $t' = t - t_{ls}$ which was also partitioned into bins. Similar averaging was then employed so that $p_{k',hist} =< \lambda_{hist}(t') >_{k'}$. The full discrete conditional spike probability is then $p_k = p_{k,0} p_{k-k_{ls},hist}$ where k_{ls} is defined as the most recent bin prior to bin k that has a spike in it. We then simulated 10 minutes worth of spikes for each model and discretization. [5] After generating the spikes we then calculated the rescaled times and CDF difference plots according to

$$z_i = 1 - e^{-\sum_{k=k_{i-1}}^{k_i} p_k} \tag{20}$$

Figures 3 C and D show the results for the inhomogeneous Bernoulli model. Comparison with Figure 2 reveals that the main effect of inhomogeneity is to smooth out the "steps". The positive (negative) biases at low (high) rescaled times remain, and as expected they are smaller for finer temporal discretizations. Figures 3 E – H show the results when the spike history dependent term is added to both the homogeneous and inhomogeneous Bernoulli models. The important point is that the biases are *worse* for both models when spike history effects are included, even though the models are constructed so that the mean firing rate remains 40 Hz. The reason for this is that the history dependent term is constructed so that the spike train exhibits burst like behavior. Specifically, after a short (2 msec) refractory period there is an increased probability of a spike. This increases the number of short ISIs. It also increases the smallest possible rescaled ISI z because the probability of a spike goes up immediately following a prior spike, and this ends up shifting distributional weight to short ISIs. This is an important point because it implies that in real experimentally recorded spike trains, which may exhibit burst type behavior, the bias in the KS plot will be worse than would be expected by a simple estimate based upon the mean firing rate, as given in equation 14.

2.3 Unbiased Discrete Time Rescaling Test Using Model Simulation

In a previous section we showed analytically that when discrete time models are used, the rescaled ISIs may not be exponentially distributed even if the model is exactly correct and that this manifests in the KS plot as systematic biases. Our first proposed solution (we present a

[5] For the spike history dependent models the generation of a spike in bin k modifies the firing probabilities in bins $k' > k$. Thus the simulation proceeded bin by bin and upon generation of a spike, the firing probabilities in the following bins were updated according to equation 19 before generating the next observation (spike or no spike) in bin $k + 1$.

second in the following section) to the bias problem is to not assume that the rescaled ISIs are exponentially (or uniformly) distributed, but to instead use a procedure similar to bootstrapping. This proceeds by noting that if a candidate model accurately describes the recorded spikes, then the rescaled ISI distribution of the spikes and *the rescaled ISI distribution expected by the fitted model* should be statistically indistinguishable. If instead the model form is inappropriate to describe the spiking data, then the rescaled ISI distribution expected by the candidate model will not match that of the experimentally recorded spikes, because the model does not describe the recorded spikes accurately. Although the expected distribution of rescaled ISIs is implicitly defined by the fitted model, in practice an explicit analytic form for this distribution may be hard to come by. It can however be sampled numerically using the fitted model to generate spikes and rescaling the resulting ISIs as a function of the model used to generate them. [6]

Specifically, after a candidate model is proposed and fit to the recorded spike train data (any type of candidate model may be used as long as it provides an estimate of the conditional intensity function λ) we use the model to simulate spikes, rescale the resulting simulated ISIs and then use a two sample Kolmororov Smirnov test to determine if the sample of estimated rescaled ISIs $\{z_{est}\}$ and the sample of experimentally recorded rescaled ISIs $\{z_{exp}\}$ are consistent with being drawn from the same underlying distribution (Press, 2007). Formally, the null hypothesis of the KS test has been changed from that stated in section 2 and adapted to the case of discrete time data using a model based approach.

H_0(*estimated*) *: Given a model of the conditional intensity function which is statistically adequate, the set of experimentally recorded ISIs can be rescaled so that they are distributed in the same manner as a sample of rescaled ISIs generated by the statistical model itself.*

To determine the number of spikes (or length of time) which must be simulated we use the analytical expression for the confidence bounds of the two sample KS test. These bounds are a function of the sample sizes of the distributions being compared, in our case the size of the empirical distribution N_{exp} dictated by experiment and the size of the simulated distribution N_{sim} which we can chose. The two sample KS test determines the maximum difference between the CDFs of the experimentally recorded rescaled ISIs $\{z_{exp}\}$ and the set of rescaled ISIs simulated using the model $\{z_{sim}\}$. If the maximum difference between the two CDFs is less than a certain value, specifically

$$\max|CDF_{sim}(z) - CDF_{exp}(z)| < 1.36\sqrt{\frac{N_{exp}+N_{sim}}{N_{exp}N_{sim}}} \tag{21}$$

then the null hypothesis is confirmed at the $\alpha = 0.05$ significance level (Press, 2007). Alternatively a Differential KS plot (as discussed in the prior subsection) will have 95% confidence bounds of

$$\pm 1.36\sqrt{\frac{N_{exp}+N_{sim}}{N_{exp}N_{sim}}} = \pm\frac{1.36}{\sqrt{N_{exp}}}\sqrt{\frac{1+\gamma}{\gamma}} \tag{22}$$

[6]Most generally the conditional intensity function will have the form $\lambda(t_k) = \lambda(x(t_k)|H(t_k))$ where $x(t_k)$ is the set of time varying external covariates and $H(t_k)$ is the previous spiking history. As the originally recorded spike train was of length T and the external covariates were defined over this time interval, it is simplest to simulate multiple spike trains the length of the original recording time T. For each spike train simulation $x(t)$ remains the same, but $H(t)$ will differ depending upon the exact spike times.

where we have written $N_{sim} = \gamma N_{exp}$.

Since N_{exp} is fixed by experiment, the test will be most strict (tightest confidence bounds) when $N_{sim} \to \infty$, or equivalently as γ increases. Formally, increasing N_{sim} increases the power of the KS test and reduces the number of false positives, e.g. false rejections of the Null Hypothesis. Fortunately N_{sim} need not be overly large. Already at $\gamma = 20$ the confidence bounds are only a factor of 1.02 wider than they would be in the infinite limit $(\pm 1.36/\sqrt{N_{exp}})$ in which the exact distribution would be known. This implies that a simulated spike train 20 times longer than the original, experimentally recorded, spike train provides a test power close to optimal and is sufficient to approximate the confidence bounds extremely well. In the results section of this paper we use simulated spike trains 100 times the original "experimental" length ($\gamma = 100$) which widens the confidence bounds by a factor of only 1.0005.

Specifically then, this technique proceeds as follows:

Procedure for Numerical Correction

1. Discretize the spike train into bins of width Δ and fit a discrete time statistical model.
2. Rescale the experimentally recorded ISIs using equation 8 to obtain the set of rescaled ISIs $\{z_{exp}\}$.
3. Use the statistical model to estimate the rescaled ISI distribution. Simulate $\gamma > 20$ spike trains of the same length as the original experimentally recorded spike train, and rescale the simulated ISIs $\{z_{sim}\}$.
4. Construct the CDFs of both z_{exp} and z_{sim}, take the difference, and plot it on the interval [0,1] with the confidence bounds $\pm \frac{1.36}{\sqrt{N_{exp}}} \sqrt{\frac{1+\gamma}{\gamma}}$.

2.4 Discrete Time Version of Time Rescaling Theorem

We now prove a discrete time version of the time rescaling theorem which corrects for both sources of KS plot bias. Specifically, we demonstrate how to write a new rescaled time ξ which is exponentially distributed at for arbitrary temporal discretization. The proof given here assumes that an underlying continuous time point process $\lambda(t|H_t)$ is sampled at finite resolution Δ.

Proposition—Suppose a continuous time point process $\lambda(t|H_t)$ is sampled at finite resolution so that the observation interval from $(0|T]$ is partitioned into bins of width Δ. Denote the bin in which the i'th spike is located as k_i and that of the next spike as bin $k_{i+1} = k_i + L_i$. Let $p_{k_i+l} = p(k_i + l|H_{k_i+l})$ be the *discrete time* conditional spike probabilities evaluated in bins k_{i+l} for $l = 1 \ldots L_i$. Define the random variable

$$\xi_i = \sum_{l=1}^{L_i-1} q_{k_i+l} + q_{k_i+L_i} \frac{\delta_i}{\Delta} \tag{23}$$

where

$$q_{k_i+l} = -\log(1 - p_{k_i+l}) \tag{24}$$

and $\delta_i \in [0, \Delta]$ is a random variable determined by first drawing a uniform random variable $r_i \in [0, 1]$ and then calculating

$$\delta_i = -\frac{\Delta}{q_{k_i+L}} \log[1 - r_i(1 - e^{-q_{k_i+L_i}})] \tag{25}$$

Then ξ_i has an exponential PDF with unit rate. For clarity of notation we drop the subscript i in the following proof. It should be taken as implicit.

Proof—Assume the last spike was located in bin k and the next spike in bin $k + L$. If we knew the underlying continuous time conditional intensity function $\lambda(t|H_t)$ and the exact spike time $t_\delta = (k + L - 1)\Delta + \delta$ in bin $k + L$ ($\delta \in [0, \Delta]$) then using the continuous time version of the time rescaling theorem we could write the probability of this event as

$$P(t_\delta)dt = e^{\int_{k\Delta}^{t_\delta} \lambda(u)du} \lambda(t_\delta)dt = e^{-\tau} d\tau \tag{26}$$

which is exponentially distributed in τ. Since we know neither $\lambda(t|H_t)$ nor t_δ precisely we must recast τ in terms of what we do know, the discrete bin-wise probabilities p_{k+l}.

The p_{k+l}'s can be written in terms of the underlying continuous process $\lambda(t|H_t)$. Consider any bin $k + l$. Since discretizaton enforces at most one spike per bin, p_{k+l} does not equal the integral of $\lambda(t|H_t)$ over the bin, but rather the probability (measured from the start of the bin) that the first spike waiting time is less than Δ.

$$\begin{aligned} p_{k+l} &= \int_{(k+l-1)\Delta}^{(k+l)\Delta} e^{\int_{(k+l-1)\Delta}^{t} \lambda(u)du} \lambda(t)dt \\ &= 1 - e^{-\int_{k\Delta}^{(k+1)\Delta} \lambda(u)du} \end{aligned} \tag{27}$$

Partitioning the integral in the exponent of equation 26 into a sum of integrals over each bin allows $P(t_\delta)$ to be written as.

$$P(t_\delta)dt = \exp[-\sum_{l=1}^{L-1} q_{k+l} - \int_{(k+L-1)\Delta}^{t_\delta} \lambda(u)du] \lambda(t_\delta)dt \tag{28}$$

where we've introduced $q_{k+l} = \int_{(k+l-1)\Delta}^{(k+l)\Delta} \lambda(u)du$ as a shorthand. By inverting equation 27 q_{k+l} can be written directly in terms of p_{k+l}.

$$q_{k+l} = -\log(1 - p_{k+l}) \tag{29}$$

Since we have no information about how $\lambda(t|H_t)$ varies over bin $k + L$ we can pick any functional form, as long as it obeys the constraint that its integral over the bin equals $q_{k+L} = -\log(1 - p_{k+L})$. One choice is $\lambda(t|H_t) = \frac{q_{k+L}}{\Delta}$. [7] It then follows that

[7] In fact *any* form for λ within bin $k + L$ could be chosen. Choosing it to be constant merely allows for easier random sampling.

$$P(t_\delta)dt=\exp[-\sum_{l=1}^{L-1}q_{k+l}-\frac{q_{k+L}}{\Delta}\delta]\frac{q_{k+L}}{\Delta}dt=e^{-\xi}d\xi \tag{30}$$

$P(t_\delta)dt$ has now been rewritten in terms of what we know, the q_{k+l}'s (implicitly the p_{k+l}'s). We have defined the rescaled time as

$$\xi=\sum_{l=1}^{L-1}q_{k+l}+\frac{q_{k+L}}{\Delta}\delta \tag{31}$$

(where $d\xi=\frac{q_{k+L}}{\Delta}d\delta$) to distinguish it from the rescaled time of the continuous time version of the theorem which directly sums the p_{k+l} and does not require random sampling of the exact spike time δ.

Random sampling of $\delta \in [0, \Delta]$ must respect the choice of $\lambda(t|H_t)=\frac{q_{k+L}}{\Delta}$ and the fact that we know the spike is in the bin somewhere. For our choice of λ, the probability density of δ conditioned upon there being a spike in the bin is a truncated exponential.

$$P(\delta|\text{ spike })d\delta=\frac{e^{-\frac{q_{k+L}}{\Delta}\delta}}{1-e^{-q_{k+L}}}\frac{q_{k+L}}{\Delta}d\delta. \tag{32}$$

The numerator is simply the event time probability measured from the start of bin $k + L$. The denominator is a normalization obtained by integrating the numerator between 0 and Δ. [8] To draw from this distribution we integrate it to obtain its CDF

$$\text{CDF}(\delta|\text{ spike })=\frac{1-e^{-q_{k+L}\delta/\Delta}}{1-e^{-q_{k+L}}} \tag{33}$$

set this cdf equal to a uniform random variable r and then solve for δ

$$\delta=-\frac{\Delta}{q_{k+L}}\log[1-r(1-e^{-q_{k+L}})] \tag{34}$$

This completes the proof.

There are two differences between the discrete and continuous time versions of the theorem. The first, and most fundamental, difference is that $p=1-\exp[-\int_0^\Delta\lambda(t)dt]$ *not* $\int_0^\Delta\lambda(t)dt$. The latter is only true when Δ is small. Expanding the logarithm of equation 29 we obtain

[8] Or Bayes rule could be used. E.g.

$$P(\delta|\text{ spike })=P(\delta)/P(\text{ spike })=\{\exp[-q_{k+L}\delta/\Delta]\frac{q_{k+L}}{\Delta}\}/p_{k+L}=\{\exp[-q_{k+L}\delta/\Delta]\frac{q_{k+L}}{\Delta}\}/(1-e^{-q_{k+L}})$$

$$q=\int_0^{\Delta}\lambda(t)dt=-\log(1-p)=p-[\frac{p^2}{2}-\frac{p^3}{3}+\ldots] \tag{35}$$

To properly rescale the ISIs when Δ is large, all the terms in the Taylor series must be kept. This can be thought of as introducing a correction term (in the brackets) for the finite bin size. Equivalently the approximation $1-p\approx e^{-p}$ used in the continuous time version of the proof is not valid for large p (or Δ). The second difference is that we randomly choose an exact spike time $t_\delta=(k+L-1)\Delta+\delta$ according to the distribution given in equation 32. This is done because there is simply no information about where exactly in bin $k+L$ the spike is located and for the rescaled time ξ to be exponentially distributed it must be continuously valued. In the continuous time limit, both of these distinctions vanish.

The hypothesis for testing goodness of fit is now exactly the same as that of the original time rescaling theorem, except that the rescaling is modified to take into account the discretization. Reintroducing the subscript i to denote the individual spike times k_i, the procedure for performing the KS test is simply described.

Procedure for Analytic Correction

1. Discretize the spike train into bins of width Δ with the spikes in bins $\{k_i\}$ and fit a discrete time statistical model resulting in the spike per bin probabilities p_k.
2. Generate a new time series of discrete values q_k according to

$$q_k=-\log(1-p_k) \tag{36}$$

3. For each interspike interval calculate the rescaled ISI ξ_i according to

$$\xi_i=\sum_{l=1}^{L_i-1}q_{k_i+l}+q_{k_i+L_i}\frac{\delta_i}{\Delta} \tag{37}$$

where δ is a random variable determined by first drawing a uniform random variable $r_i\in[0,1]$ and then calculating

$$\delta_i=-\frac{\Delta}{q_{k_i+L_i}}\log[1-r_i(1-e^{-q_{k_i+L_i}})] \tag{38}$$

4. Make a final transform to the random variables y_i

$$y_i=1-e^{-\xi_i} \tag{39}$$

If the discrete time statistical model is accurate the y_i will be uniformly distributed. Therefore the y_i can be used to make a KS or Differential KS plot.

3 Results

In this section we fit Generalized Linear Models (GLMs) to spike trains both simulated and experimentally recorded in awake monkey V1 cortex during visual stimulation. We then check goodness of fit using both the standard KS test and our methods. We demonstrate dramatic

and nearly identical improvement in KS test accuracy for both techniques. Although we emphasize that any discrete time statistical model may be used, we chose a GLM, specifically the logistic regression (logit link function) form, because of the discrete binary nature of spike train data. Standard linear regression assumes continuous variables, and is therefore inappropriate for the problem. Further reasons for using GLMs are their already wide application to the analysis of neural spiking activity (Frank, 2002; Truccolo, 2005; Czanner, 2008; Paninski, 2004a; Kass and Ventura, 2001), their optimality properties (Pawitan, 2001) and the ease of fitting them via maximum likelihood. Methods for fitting GLMs exist in most statistical packages including Matlab and R.

3.1 Simulated Data

Using the three continuous time point process models of the previous section (inhomogeneous Poisson, homogeneous Poisson with spike history dependence, inhomogeneous Poisson with spike history dependence) we simulated 10 minutes of spikes from each model at very fine 10^{-10} msec discretization, essentially continuous time. These spike trains are the "experimental" data. We emphasize that all of our simulated data used a realistic mean firing rate of 40 Hz, and that many experimental situations exist for which the mean firing rates are much higher (De Valois, 1982; MacEvoy, 2007). The spikes were then discretized into 1 msec bins and a GLM was fit to each simulated spike train. This procedure mimics the usual approach taken in fitting a GLM to real data. We used a logistic regression type GLM (logit link function) appropriate for discrete time binary data. Each model's spike train was fit using one of the following GLM forms:

- *Inhomogeneous Bernoulli GLM*:

$$\log\left[\frac{\lambda(k)}{1-\lambda(k)}\right]=\sum_{j=1}^{J}\beta_j\mathscr{B}_j(k) \tag{40}$$

- *Homogeneous Bernoulli With Spike History GLM*:

$$\log\left[\frac{\lambda(k)}{1-\lambda(k)}\right]=\beta_0+\sum_{r=1}^{R}\theta_r g(k-r) \tag{41}$$

- *Inhomogeneous Bernoulli With Spike History GLM*:

$$\log\left[\frac{\lambda(k)}{1-\lambda(k)}\right]=\sum_{j=1}^{J}\beta_j\mathscr{B}_j(k)+\sum_{r=1}^{R}\theta_r g(k-r) \tag{42}$$

the $\mathscr{B}_j(k)$ are periodic B-spline basis functions with knots spaced 50 msec apart. These are continuously defined (even though we use discrete time bins) temporally localized basis functions, similar in shape to Gaussians which can be computed recursively (Wasserman, 2004). Their use allows for a "PSTH like" fit, but one which is cubic polynomial smooth. The $g(k-r)$ are indicator functions equal to 1 if the most recent spike is r bins prior to k and 0 otherwise. This functional form of λ_{hist} is standard (Truccolo, 2005; Czanner, 2008). The β's and θ's are parameters to be fit via maximum likelihood.

Next, following the first procedure described in the Methods, we used the fitted GLM to simulate 100 ten minute spike trains, rescaled both the "experimental" and simulated ISIs and constructed both the KS and CDF difference plots. The results are shown in Figure 4 where the blue lines correspond to the comparison of the experimental CDF with the uniform CDF

and the red lines to the comparison of the experimental CDF with the CDF estimated from the GLM as was described in section 2.3. For all three models, the Differential KS plots reveal strong biases when the "experimental" rescaled ISIs are compared with the uniform distribution, and a complete elimination of the bias when the distribution simulated from the GLM is used. Further, use of the GLM simulated distribution makes the difference between the Differential KS plot lying within or outside the 95% confidence bounds. This was true even when spike history effects were included and KS plot biases much worse than in their absence. Finally we applied the analytic discrete time rescaling theorem described in section 2.4 and plotted the results in green. The analytically corrected Differential KS plot is nearly identical to the numerically simulated one. This indicates that the analytical correction, which is simpler to apply, is sufficient to test model goodness of fit.

3.2 Monkey V1 Receptive Field Data

Next we used spiking data recorded in V1 of two adult female rhesus monkeys (*Macaca mulatta* during a fixation task. See Appendix B for details on the experimental procedure. The visual stimuli consisted of a high contrast light bar (50 cd/m^2; bar width, 0.2° or 5 pixels) moving with a constant velocity (v = 14.9°/s or 425 pixels/s). The bar was presented in a square aperture of size (21.8° × 21.8° or 600×600 pixels centered over the receptive fields of the neurons being recorded. During stimulus presentation the monkey was required to maintain fixation within a virtual window (window size, 1°) centered on the fixation point.

In this paper we show data from two monkeys. For each monkey we selected two examples in which the recorded cells exhibited high average firing rates (first column of Figure 5). The data shown was recorded over 9 trials, each of which lasted 2 seconds, during which the bar moved in a single direction. As with the simulated data we used a GLM based logistic regression form (logit link function) for the conditional intensity function, with a temporal discretization of Δ = 1 msec.

$$\log\left[\frac{\lambda(k)}{1-\lambda(k)}\right] = \underbrace{\sum_{j=1}^{J}\xi_j x_j(k)}+\underbrace{\sum_{r=1}^{R}\theta_r g(k-r)} \\ \stackrel{\text{def}}{=} \psi_{\text{stim}}(k)+\psi_{\text{hist}}(k) \tag{43}$$

where x_j represents the jth covariate that encodes the stimulus input (which may be in a form of either a feature of the visual stimulus, or a PSTH profile, or a specific basis function). The $g(k-r)$ is, as in the previous section, an indicator function representing whether the most recent spike was in the past rth temporal window, and θ_r represents the associated weight coefficient (a negative θ_r implies an inhibitory effect that might account for the refractory period of neuronal firing, while a positive θ_r implies an excitatory effect). The first term of right-hand side of equation (1) are summarized by a stimulus-dependent response factor denoted by ψ_{stim} and the last term represents a spiking history-dependent factor denoted by ψ_{hist}.

Similarly to the simulated inhomogeneous Bernoulli process data in the last section, we used semi-parametric cubic B-spline functions (piecewise smooth polynomials) to model the stimulus-induced spiking activity ψ_{stim}.

$$\psi_{\text{stim}}(k)=\sum_{j=1}^{J}\xi_j \mathscr{B}_j(k) \tag{44}$$

where J denotes the number of knots or control points. Note that the values of control points only affect the shape of ψ_{stim} locally due to the piecewise definition. For the data shown here 12 control points are non-evenly placed on the 2-s time interval

As with our simulated data, we see in Figure 5 that the when the standard KS test was used, the KS and Differential KS plots lay outside the 95% confidence bounds. However when temporal discretization was taken into account and our two techniques were used, the plots lay well within confidence bounds and the GLM model was shown to be very well fit to the data. Thus again the simple analytic method is found to be sufficient to account for discretization induced KS plot bias.

4 Discussion

It is vital to check a model's goodness of fit before making inferences from it. The time rescaling theorem provides a powerful, yet simple to implement, statistical test applicable to a spike train, or other point process, for which the data is binary rather than continuously valued. The theorem states that the ISIs of a continuous time point process can be rescaled (through a variable transformation) so that they are independent and exponentially distributed. The rescaled ISIs can then be compared to the exponential distribution using a Kolmogorov Smirnov test, or further rescaled to a uniform distribution and the KS test performed graphically Brown (2001). Each ISI is rescaled as a function of the time varying spike probability over that particular ISI. Thus time rescaling considers the probabilities of individual ISIs and provides a much stronger statistical test than, for example, tests based upon the unscaled ISI distribution. Practical numerical considerations dictate that the fitting of a statistical model usually requires the discretization of time into bins. For the purposes of the time rescaling theorem, if the spike rate is low, ISIs long and the probability per bin of a spike small, the distinction between discrete and continuous time will often not be important. In this paper we addressed the case where the spike rate is high, ISIs short and the probability per bin of a spike large so that the distinction between discrete and continuous time matters.

When the probability per time bin of a spike is not sufficiently small, the standard, continuous time KS plot exhibits biases at both low and high rescaled ISIs. The source of these biases is twofold and originates in the consequences of discretizing a continuous time point process. First, the uncertainty as to where exactly in a bin a spike is located causes discrete time models place a lower bound on the size of the smallest rescaled ISI z This leads to positive KS plot bias at low z. Second, because discrete binary models only allow for a single spike per bin, they estimate per bin spike probabilities p_k which are less than $\int_0^{\Delta}\lambda(t)dt$ with the integral over bin k. We demonstrated both of these points theoretically using a homogeneous Poisson process which we discretized into a homogeneous Bernoulli process, and also in our proof of the discrete time version of the theorem. These biases can be numerically relevant even at moderate spike rates and reasonable temporal discretizations. In this paper we considered mainly 40 Hz spiking at 1 msec discretization, ($p = 0.04$) but under some neurophysiological conditions, the spike rate can be much higher. For example, the awake monkey data presented in the Results exhibited firing rates which at times exceeded 100 Hz.

Under such conditions KS plots will exhibit biases at both low and high rescaled ISIs which can not be removed through more accurate numerical integration techniques or increased data sampling. In fact, sampling a longer spike train will make the issue more critical because the 95% confidence bounds on the KS plot scale as $1/\sqrt{N_{exp}}$ where N_{exp} is the number of experimentally recorded ISIs. In cases of long recording times the confidence bounds can be quite tight and it can be difficult to see variations in the fit using the standard KS plot even if those variations are statistically significant. We therefore introduced a new type of plot, the

"Differential KS plot" in which we plot the difference between the CDFs of the empirical and simulated ISI distributions along with analytical 95% confidence bounds. This new type of plot displays the same information as the original KS plot, but in a more visually accessible manner.

To handle KS plot bias we proposed and implemented two different procedures, both of which are capable of testing the statistical sufficiency of any model which provides a measure of the discrete time conditional intensity function. The first procedure operates purely in discrete time and uses numerical simulation, in a manner similar in spirit to a bootstrap, to estimate the distribution of rescaled ISIs directly from a fitted statistical model. Model goodness of fit is tested by comparing the estimated and experimentally recorded rescaled ISI distributions using a KS test. The confidence bounds on this two sample KS test scale as $\sqrt{N_{exp}N_{sim}/(N_{exp}+N_{sim})}$. This procedure is therefore computationally tractable because a simulated spike train twenty times longer than the original experimentally recorded spike train will result in a KS test with confidence bounds only 1.02 times as wide as if the exact rescaled ISI distribution were known. For the second technique we presented and proved a discrete time version of the time rescaling theorem. This presumes an underlying continuous time point process which is sampled at finite resolution Δ, analytically corrects for the discretization and defines a rescaled time ξ which is exponentially distributed at arbitrary temporal discretizations. We applied these two techniques to both simulated spike trains and also to spike trains recorded in awake monkey V1 cortex and demonstrated an elimination of KS plot bias when our techniques were used. The performance of both techniques was nearly identical, revealing high goodness of fit even when the fitted model failed the standard continuous time application of the KS test. Therefore either method might be used, although the analytical method is perhaps preferable, if only because it is quicker to compute.

The discrete time rescaling theorem is appropriate for point process type data such as spike trains which are equally well described by either their spike times or their interspike intervals. It is, however, a test of model *sufficiency*, namely whether a proposed statistical model is sufficient to describe the data. It does not, in and of itself, address issues of model complexity (over fitting) or whether the model form chosen is appropriate for describing the data in the first place. [9] Over fitting can be guarded against by splitting one's data into training and test data. After fitting the model parameters using the training data, the fitted model and the discrete time rescaling theorem can be applied to the test data. Of course we do not mean to imply that the discrete time rescaling theorem is the *only* statistical test which should be employed for selecting and validating an appropriate model. Other other statistical tests and criteria, for example log likelihood ratio tests, the Akaike and Bayesian Information Criteria and so forth, should also be employed to rigorously judge goodness of fit and model complexity.

One might reasonably ask why not simply fit a statistical model with extremely fine temporal discretization so that the time rescaling theorem applies in its standard form. There are several issues. First, spikes are not instantaneous events but spread out in time on the order of a msec or slightly less. Secondly, experimenters often exclude apparent spikes which occur less than a msec (or thereabouts) apart in a recording as it is difficult to distinguish spike wave forms which essentially lie on top of each other. For both these reasons defining spikes as instantaneous events is physically problematic. Although the continuous time point process framework is theoretical appealing, there is usually no reason not to consider the data in discrete time, fit a discrete time model and perform a discrete time goodness of fit test. Finally there is

[9] In this paper we used GLMs. Such models are widely applied to the analysis of spike train data. It is also interesting to note that they have an interpretation as a sort of integrate and fire neuron, see for example Paninski (2004b). However nothing in the discrete time rescaling theorem precludes its use for testing the fit of a statistical model of the spiking probability which is not a GLM.

the important issue of computation time and computer memory. When recording times are long and the number of spikes large, confidence bounds on the KS test will be very very tight. Extremely fine temporal discretization will then be required for the biases to be less than the width of the confidence bounds. The amount of memory and computation time required under these conditions can rapidly become prohibitive. Further, since using the discrete time rescaling theorem is almost as quick and simple a procedure as the standard KS test, the authors can see no reason *not* to use it. In closing, a failure of the standard KS test does not immediately imply poor model fit. Biases induced by temporal discretization may be a factor, and should be considered before rejecting the model.

Acknowledgments

The authors would like to thank Sergio Neuenschwander and Bruss Lima for the generous use of their data. This work was supported by NIH grants K25 NS052422-02, DP1 OD003646-01, MH59733-07, the Hertie Foundation, the Max Planck Society, and the EU Grant FP6-2005-NEST-Path-043309.

Appendix A: Independence of Rescaled Times

In this appendix we prove that the rescaled times ξ_i (and in the continuous time limit τ_i) are not only exponentially distributed, but also independent. To establish this result it suffices to show that the joint CDF of the ξ_i's can be written as the product of the individual CDF's and that these CDF's are those of independent exponential random variables with rate 1. The CDF of an exponential random variable with rate 1 is

$$F(\xi)=1-e^{-\xi} \tag{45}$$

We recall that the rescaled times ξ_i are defined as

$$\xi_i=\sum_{k=k_{i-1}+1}^{k_i-1} q_k+\frac{q_{k_i}}{\Delta}\delta_i \tag{46}$$

where $\delta_i \in [0, \Delta]$ is a random variable determined by first drawing a uniform random variable $r_i \in [0, 1]$ and then calculating

$$\delta_i=-\frac{\Delta}{q_{k_i}}\log[1-r_i(1-e^{-q_{k_i}})] \tag{47}$$

This definition of the rescaled times ξ_i implicitly defines a spike time $t_i = (k_i-1)\Delta+\delta_i$. Because the transformation from the spike times t_i (or the spike bins k_i) to the ξ_i's is one-to-one we have that the following two events are equivalent

$$\{\Xi_1<\xi_1, \Xi_2<\xi_2, \ldots, \Xi_N<\xi_N\}=\{T_1<t_1, T_2<t_2, \ldots, T_N<t_n\} \tag{48}$$

Therefore the joint CDF of the ξ_i's is

$$\begin{aligned} F(\xi_1, \xi_2, \ldots, \xi_N) &= P\{\Xi_1 < \xi_1, \Xi_2 < \xi_2, \ldots, \Xi_N < \xi_N\} \\ &= P\{T_1 < t_1, T_2 < t_2, \ldots, T_N < t_n\} \\ &= F(t_1, t_2, \ldots, t_N) \\ &= \prod_{i=2}^{N} F(t_i | t_1, \ldots, t_{i-1}) F(t_1 | 0) \end{aligned} \tag{49}$$

The last line follows from the multiplication rule of probability (Miller, 2001).

The conditional CDFs $F(t_i|t_1,\ldots, t_{i-1})$ can be calculated by noting that the probability of any given ISI is equal to 1 minus the probability that there was at least one spike within the epoch defined by the ISI. Formally this can be written as

$$\begin{aligned} P(\text{no spike in } (k_{i-1}\Delta, (k_i - 1)\Delta + \delta_i)) &= 1 - P(\text{at least one spike in } (k_{i-1}\Delta, (k_i - 1)\Delta + \delta_i)) \\ &= 1 - \int_{k_{i-1}\Delta}^{t_i} P(t_i' | k_1, k_2, \ldots, k_{i-1}) dt_i' \\ &= 1 - F(t_i | k_1, k_2, \ldots, k_{i-1}) \end{aligned} \tag{50}$$

The right hand side has the CDF we wish to calculate. It remains to determine the LHS. But this is simply the ISI probability of equation 30

$$P(t = (k_i - 1)\Delta + \delta_i | k_1, k_2, \ldots, k_{i-1}) d\delta_i = P(\xi_i) d\xi_i = e^{-\xi_i} d\xi_i \tag{51}$$

Thus

$$F(t_i | k_1, k_2, \ldots, k_{i-1}) = F(\xi_i) = 1 - e^{-\xi_i} \tag{52}$$

Inserting this result into equation 49 we get

$$\begin{aligned} F(\xi_1, \xi_2, \ldots \xi_N) &= \prod_{i=1}^{N} 1 - e^{\xi_i} \\ &= \prod_{i-1}^{N} F(\xi_i) \end{aligned} \tag{53}$$

which establishes the proof. A similar argument can be made for the continuous time rescaled time τ_i. An intuitive way to understand the independence of the rescaled ISIs is that these were calculated using either $\lambda(t|H_t)$ or $p(k|H_t)$ which are conditioned upon the previous spiking history. This 'pre'-conditioning before calculation of the rescaled times enforces the independence of ξ and/or τ. This independence of the rescaled times is useful because testing for independence provides an additional statistical significance test beyond the testing for exponentiallity by a KS or CDF difference plot. The test is identical to the continuous time case and we refer the reader to (Czanner, 2008) for a discussion.

Appendix B: Experimental Procedures

Experimental procedures were approved by the National Committee on Animal Welfare (Regierungspraesidium Hessen, Darmstadt) in compliance with the guidelines of the European Community for the care and use of laboratory animals (European Union directive 86/609/EEC).

Neuronal spiking activities were recorded in awake and head-fixed monkeys in opercular region of V1 (RFs centers, 2–5° of eccentricity) and, on some occasions, from the superior bank of the calcarine sulcus (8–12° of eccentricity).

Quartz-insulated tungsten-platinum electrodes (diameter 80 μm, 0.3–1.0 MΩ impedance; Thomas Recording) were used to record the extracellular activities from 3 to 5 sites in both superficial and deep layers of the striate cortex (digitally band-pass filtered, 0.7–6.0 kHz; Plexon Inc.). Spikes were detected by amplitude thresholding, which was set interactively based on online visualization of the spike waveforms (typically, 2–3 s.d. above the noise level). Trials with artifacts were rejected during which either monkey did not maintain fixation or monkey showed no response or incorrect behavior.

References

Barbieri R, Matte EC, Alabi AA, Brown EN. A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability. Am. J. Physiol. Heart Circ. Physiol 2005;288:H424–H435. [PubMed: 15374824]

Brown EN, Barbieri R, Ventura V, Kass RE, Frank KV. The time-rescaling theorem and its application to neural spike train data analysis. Neural Computation 2001;14:325–346. [PubMed: 11802915]

Czanner G, Eden UT, Wirth S, Yanike M, Suzuki WA, Brown WA. Analysis of between-trial and within-trial neural spiking dynamics. Journal of Neurophysiology 2008;99:2672–2693. [PubMed: 18216233]

De Valois RL, Yund EW, Hepler N. The orientation and direction selectivity of cells in macaque visual cortex. Vision Research 1982;22:531–544. [PubMed: 7112953]

Frank LM, Eden UT, Solo V, Wilson MA, Brown EN. Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. Journal of Neuroscience 2002;22:3817–3830. [PubMed: 11978857]

Johnson, A.; Kotz, S. Distributions in statistics: continuous univariate distributions. 2nd ed.. New York: Wiley; 1970.

Kass RE, Ventura V. A spike train probability model. Neural Computation 2001;13:1713–1720. [PubMed: 11506667]

MacEvoy SP, Hanks TD, Paradiso MA. Macaque V1 activity during natural vision: effects of natural scenes and saccades. Journal of Neurophysiology 2007;99:460–472. [PubMed: 18077668]

Massey FJ. The kolmogorov-smirnov test for goodness of fit. Journal of the American Statistical Association 1951;46:68–77.

McCullagh, P.; Nelder, JA. Generalized Linear Models. 2rd ed.. New York, New York: Chapman and Hall; 1989. p. 106-114.

Miller, GK. Probability. Hoboken, NJ: Wiley; 2006. p. 82-85.

Paninski L. Maximum likelihood estimation of cascade point process neural encoding models. Network 2004;4:243–262. [PubMed: 15600233]

Paninski L. Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. Neural Computation 2004;16:2533–2561. [PubMed: 15516273]

Pawitan, Y. In all likelihood: statistical modeling and inference using likelihood. London: Oxford University Press; 2001.

Press, WH.; Teukolsky, SA.; Vetterling, WT.; Flannery, BP. Numerical Recipes. 3rd ed.. New York, New York: Cambridge University Press; 2007. p. 736-740.

Song, D.; Chan, RH.; Marmarelis, VZ.; Hampson, RE.; Deadwyler, SA.; Berger, TW. Physiologically plausible stochastic non-linear kernel models of spike tain to spike train transformation; Conf. Proc. IEEE Eng. Med. Biol. Soc; 2006. p. 6129-6132.

Snyder, D. Random Point Processes. New York: John Wiley & Sons; 1975.

Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. Journal of Neurophysiology 2005;93:1074–1089. [PubMed: 15356183]

Wasserman, L. All of Statistics. New York, New York: Springer-Verlag; 2004. p. 223-224.

Wasserman, L. All of Nonparametric Statistics. New York, New York: Springer-Verlag; 2007.

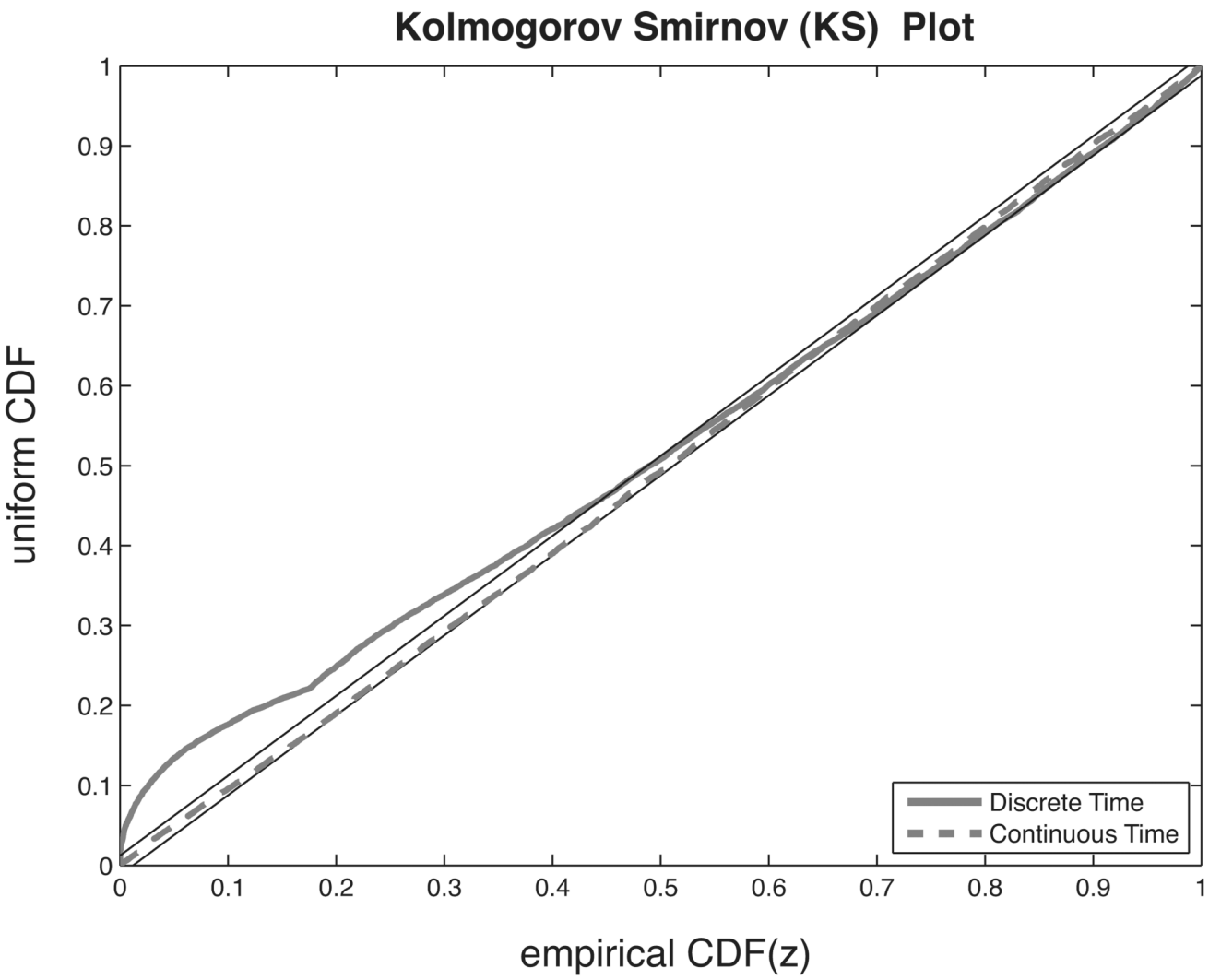

Figure 1.

Example of two simple KS plots demonstrating that temporal discretization induces biases even if the conditional intensity function used to calculate the rescaled times is exactly correct. CDF of the rescaled times *z* is plotted along the x-axis versus the CDF of the uniform (reference) distribution along the y-axis. Spikes were generated from an inhomogeneous Poisson process with a maximum firing rate of 50 Hz. Thick grey dashed line: KS plot of rescaled ISIs generated by a continuous time model. Thick grey line: KS plot of rescaled ISIs calculated from the same model discretized at 5 msec resolution. The discretization was deliberately enhanced to emphasize the effect. Thin black 45 degree lines are 95% confidence bounds upon the KS plots.

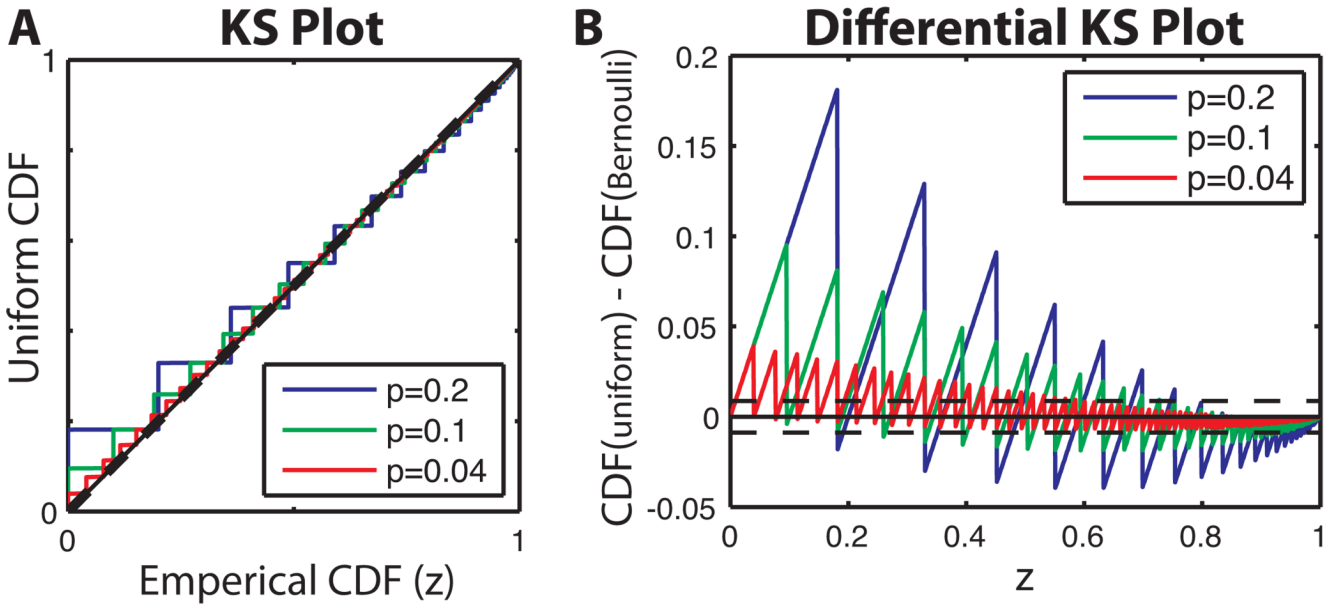

Figure 2.

Illustration of KS plot bias induced when a homogeneous Poisson process is discretized to a homogeneous Bernoulli process. (A) KS plot for various spike per bin probabilities p. Blue: $p = 0.2$, green: $p = 0.1$, red: $p = 0.04$ (40 Hz at 1 msec discretization). The rescaled times are not uniformly distributed but have positive bias at rescaled ISIs close to 0 and negative bias at rescaled ISIs close to 1. (B) Differential KS Plot: $CDF_{uniform} - CDF(z)_{bernoulli}$. Biases are easier to see if the difference between the expected CDF (uniform) and the actual CDF of the rescaled times is plotted. The colors indicate the same spike per bin probabilities p as in A). The horizontal dashed lines are the 95% confidence region assuming a 10 minutes of a 40 Hz Bernoulli process (24000 spikes).

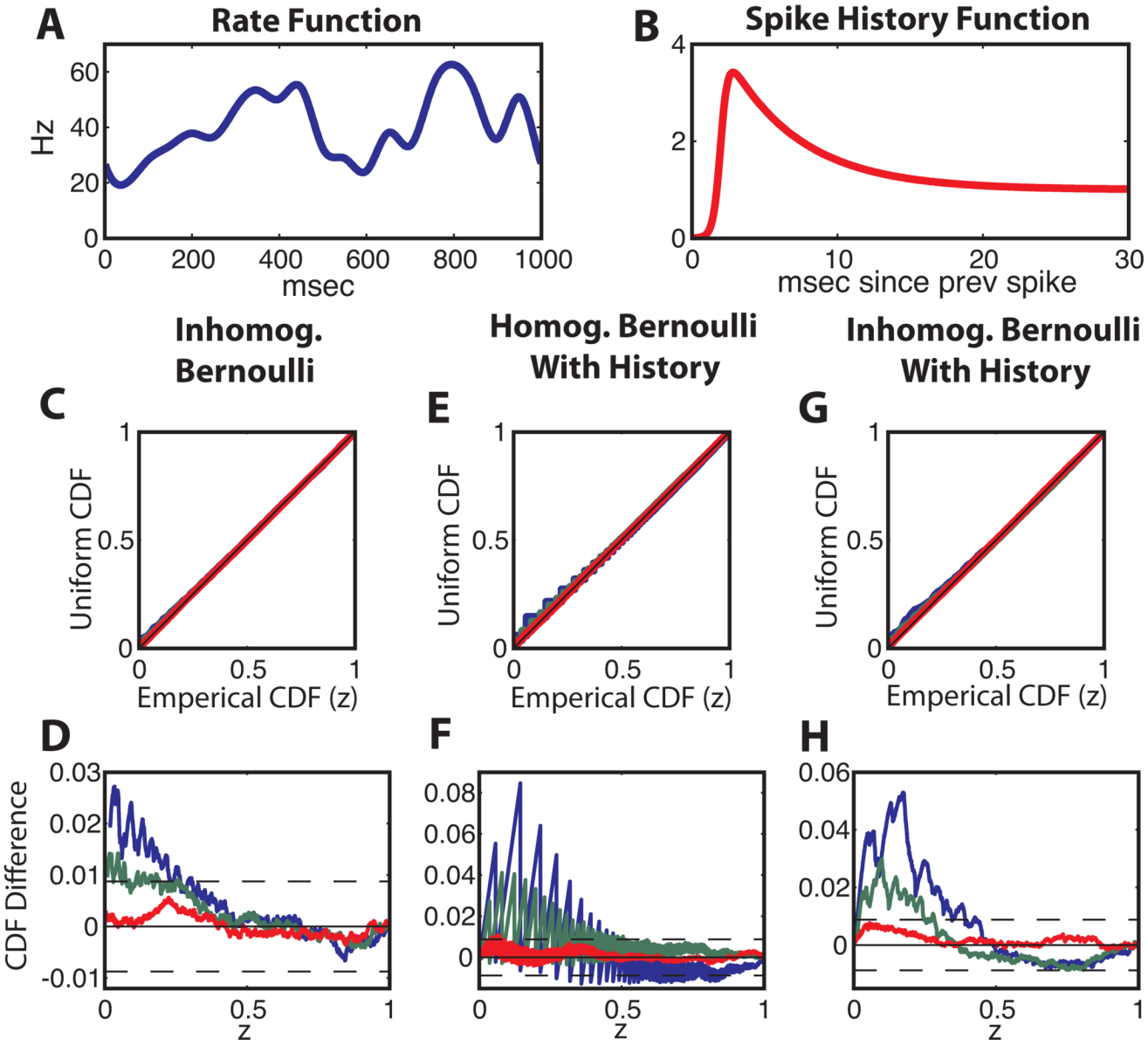

Figure 3.

KS and Differential KS plots for 10 minute long 40 Hz mean firing rate simulated spike trains. Three continuous time models of the conditional intensity function were used for simulation. 1) inhomogeneous Poisson process 2) homogeneous Poisson with a renewal spike history process 3) inhomogeneous Poisson with a renewal spike history process. (See text) The continuously defined processes were discretized at various values Δ and used to simulate spikes. (A) 40 Hz mean inhomogeneous Bernoulli firing rate. (B) Spike history term λ_{hist} as a function of time since the most recent spike. (C) and (D) KS and Differential KS plots for inhomogeneous Bernoulli process. Blue: Δ = 1 msec, green: Δ = 0.5 msec, red: Δ = 0.1 msec. Horizontal dashed lines are 95% confidence bounds. (E) and (F) Homogeneous Bernoulli process with spike renewal history term. (G) and (H) Inhomogeneous Bernoulli process with spike renewal history term. Note that when spike history effects are present the biases are larger both at short and long rescaled ISIs.

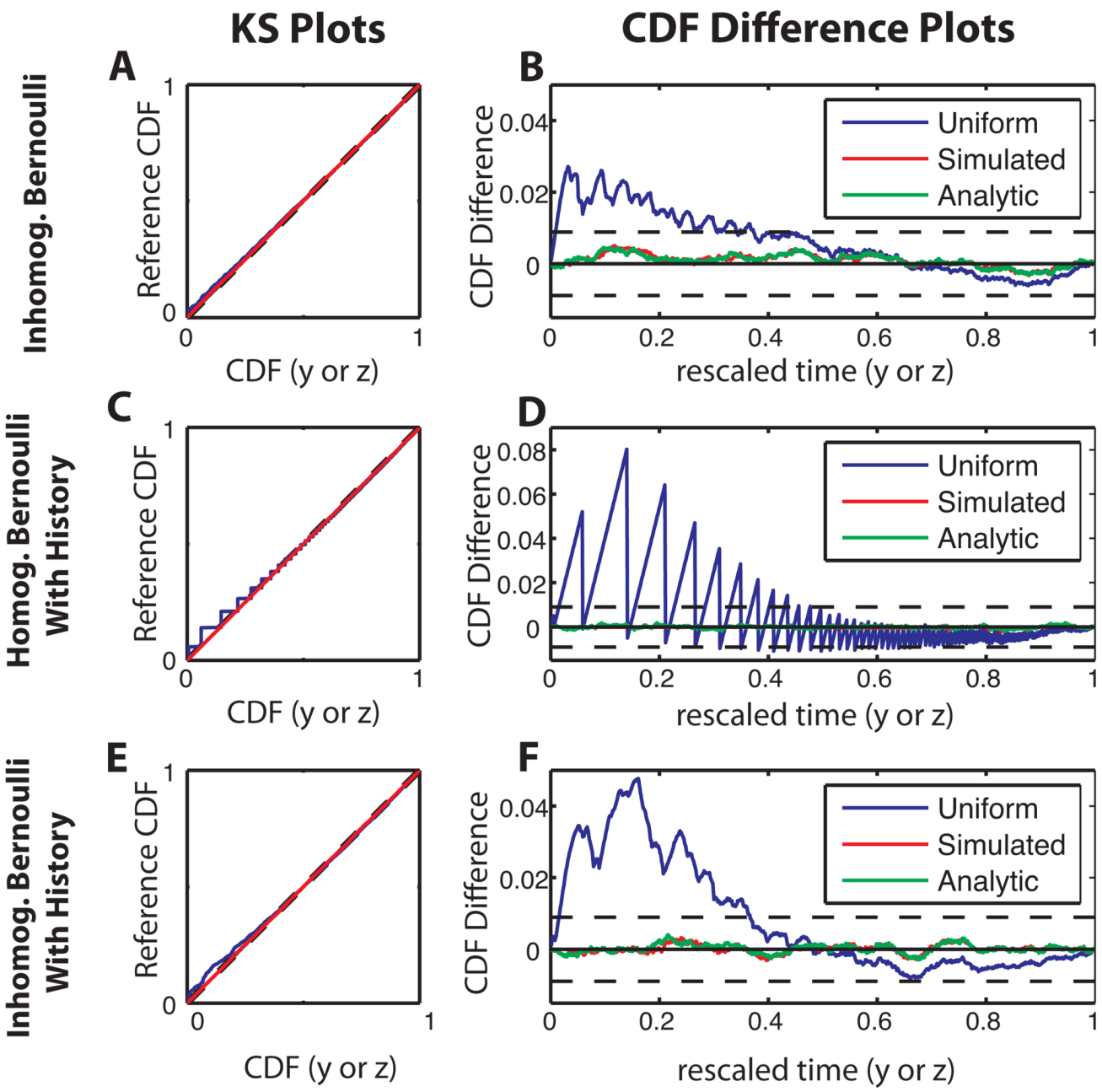

Figure 4.

Comparison of standard KS test, KS test using simulated rescaled ISI distribution, and KS test using the analytically corrected rescaled time. Spike trains were simulated using the same three models as in Figure 2 at fine 10^{-10} msec temporal precision and then discretized at $\Delta = 1$ msec resolution. Logistic GLM models were fit and used to estimate the rescaled ISI distributions (See text.) (A), (C) and (E) KS plots for inhomogeneous Bernoulli, homogeneous Bernoulli with spike history and inhomogeneous Bernoulli with spike history respectively. (B), (D) and (F) Differential KS plots for the same. Blue lines correspond to the standard KS test which plots the CDF of the rescaled time *z* versus the CDF of the uniform distribution, red lines to the numerical simulation method which plots the CDF of the rescaled time *z* versus the CDF

of the numerically simulated reference distribution, and green lines to the analytical method which plots the CDF of the analytically corrected rescaled time *y* versus the CDF of the uniform distribution. The red and green lines essentially overlap in the plots. For all three spike train models, strong KS and Differential KS plot bias was eliminated when either the numerically estimated distribution or the analytical correction were used.

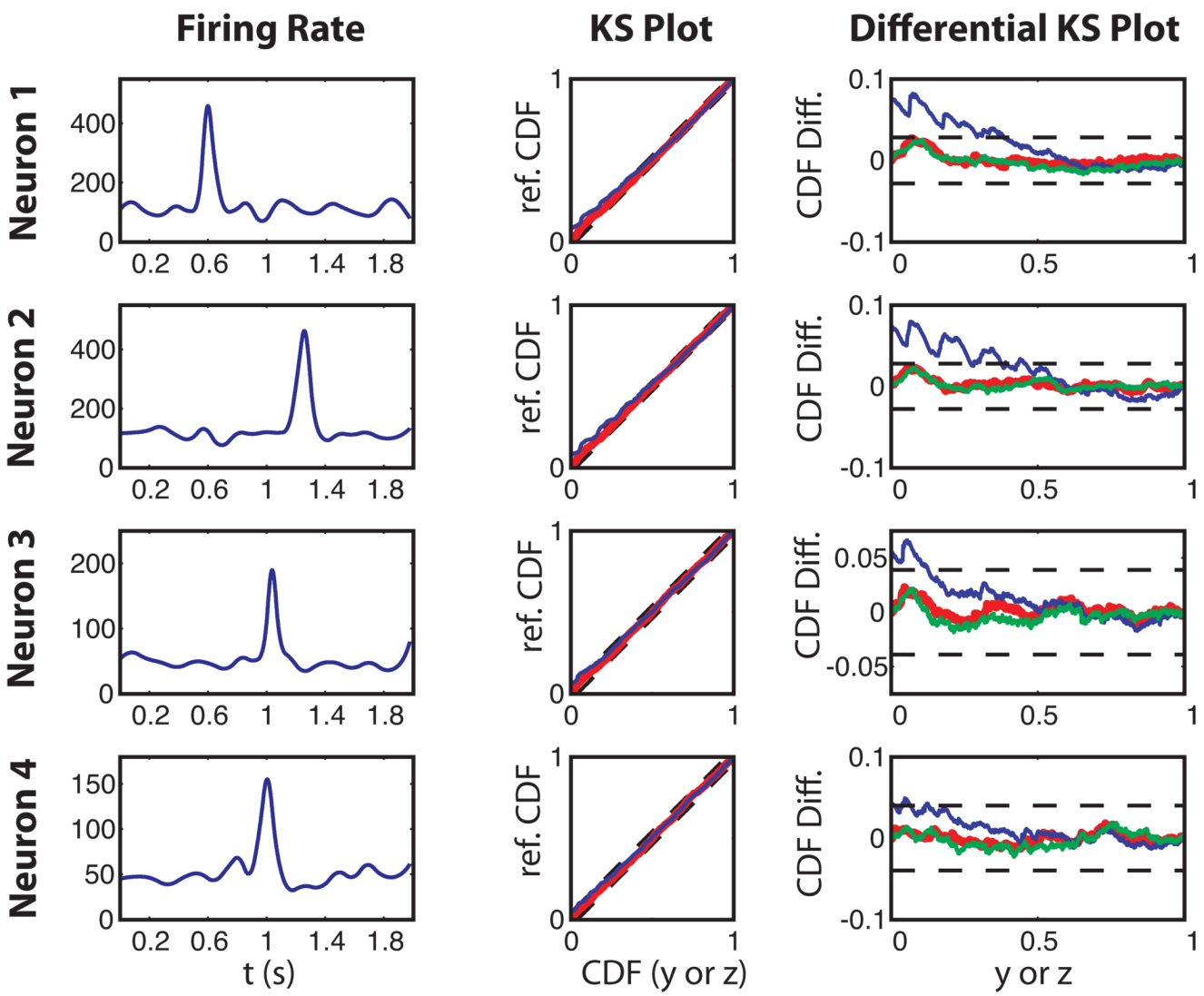

Figure 5.

Four examples of neurons from two different monkeys (top two rows: monkey 1, bottom two rows: monkey 2) for which goodness of fit appears to be poor when the standard KS test is used but revealed to be good when either the numerically estimated reference distribution or the analytically corrected rescaled time *y* are used. First Column: firing rate, second column: KS plot, third column: Differential KS plots. Blue: standard KS test. Red: KS test with numerical simulation of reference distribution. Green: KS test with analytically corrected rescaled time *y*. As with the simulated spike trains of Figure 4, the KS and Differential KS plot biases are eliminated when either the rescaled ISI distribution (*z*) is simulated using the fitted model, or the analytically corrected rescaled time *y* is used.