Bayesian Nonparametric Modeling for Comparison of Single-Neuron Firing Intensities

Athanasios Kottas

Department of Applied Mathematics and Statistics, University of California,

Santa Cruz, California 95064, U.S.A. (Email: thanos@ams.ucsc.edu)

and

Sam Behseta

Department of Mathematics, California State University Fullerton,

Fullerton, California 92834, U.S.A.

SUMMARY. We propose a fully inferential model-based approach to the problem of comparing the firing patterns of a neuron recorded under two distinct experimental conditions. The methodology is based on non-homogeneous Poisson process models for the firing times of each condition with flexible nonparametric mixture prior models for the corresponding intensity functions. We demonstrate posterior inferences from a global analysis, which may be used to compare the two conditions over the entire experimental time window, as well as from a pointwise analysis at selected time points to detect local deviations of firing patterns from one condition to another. We apply our method on two neurons recorded from the primary motor cortex area of a monkey's brain while performing a sequence of reaching tasks.

KEY WORDS: Beta mixtures; Dirichlet process; Neuronal data; Non-homogeneous Poisson process; Primary motor cortex area.

1. Introduction

Stochastic modeling and statistical estimation techniques for the analysis of data from neurophysiological studies have received considerable attention in the literature. One of the key techniques in neuroscience involves recording of electrical activity of neurons in laboratory animals. The technique studies action potentials (spikes) generated by the neuron and measured using an electrode inserted into the animal's brain. The times at which spikes occur (referred to as firing times) are recorded to provide the neuronal data. In this context, the focus of statistical modeling approaches is on the temporal evolution of the neuronal firing activity. Reviews of related work, for data from either a single neuron or from multiple neurons, can be found in, e.g., Brillinger (1992), Brown, Kass and Mitra (2004), and Kass, Ventura and Brown (2005). See also West (1997) and Rigat, de Gunst and van Pelt (2006), and further references therein, for Bayesian modeling approaches under different neuronal data structures.

This paper develops modeling for neuronal data arising as firing times from a single neuron under two distinct experimental conditions. In particular, the motivating neurophysiological study involves neurons recorded from the primary motor cortex area of a Macaque monkey's brain while performing the sequential task of reaching a series of illuminating targets on a touch-sensitive screen (Matsuzaka, Picard and Strick, 2007).

We develop a Bayesian nonparametric modeling framework that allows full inference for comparison of neuronal firing intensities. In particular, a key feature of the proposed approach is that it enables inference for both global and local differences in the firing intensities under the two experimental conditions. From a methodological point of view, we seek a flexible prior model for non-homogeneous Poisson process intensities over time. Such a model is formulated using a class of nonparametric mixtures for densities supported by bounded intervals, in the process, leading to a modeling approach

2

for temporal point patterns.

The outline of the paper is as follows. Section 2 provides background on the motivating neurophysiological study and corresponding data as well as discussion of some of the related literature. The methodology for modeling the evolution of neuronal firing rates over time is presented in Section 3. Section 4 considers the inferential problem of comparing neuronal firing intensities under two experimental conditions, using the Bayesian nonparametric model developed in Section 3. Section 5 applies the modeling approach to the data discussed in Section 2. Finally, Section 6 concludes with a summary and discussion of possible extensions.

2. Motivating Application and Literature Review

Section 2.1 provides details on the data from the neurophysiological study that motivates our method for comparison of neuronal firing intensities. The stochastic framework, based on Poisson processes, we utilize for modeling neuronal firing times is discussed in Section 2.2. Section 2.3 reviews relevant existing work in order to place our contribution within the literature.

2.1 Background on the Primary Motor Cortex Neuronal Data

The primary motor cortex area (or M1) is long known to be involved in execution of movements by guiding muscle activities (e.g., Gazzaniga, Ivry and Mangun, 2002). Single neuron recordings in monkeys suggested that premotor or supplementary motor cortical regions of the brain are in charge of the planning phase of movements, while the execution of movements are implemented through M1 (Ashe et al., 1993). This is usually referred to as the motor hierarchy (see, e.g., Gallistel, 1980).

To investigate this theory further, Matsuzaka et al. (2007) studied the activity of monkeys' M1 neurons during two conditions of a sequential pointing task. They recorded the firing times of a single neuron while a monkey was trained to reach three illuminated targets among the total of five targets on a touch-sensitive screen. The target triplets were illuminated under two conditions. First, the target appearances were in a repeating order, whereas in the second condition, the targets appeared based on a pseudo-random sequence. (The two conditions will be referred to as *repeating* and *random*, respectively.) The task was highly practiced so that the monkey was able to "learn" the repeating patterns to a degree that, at times, the monkey could predict the upcoming illuminated target prior to its appearance on the screen. Matsuzaka et al. (2007) suggested: "M1 may be a site of storage for the internal representation of skilled sequential movements." This conclusion questions the role of M1 as a mere executioner of activities in the motor hierarchy. Additionally, to quantify movement kinematics associated with this experiment, the authors recorded the electromyography activity of a large number of both axial and proximal arm muscles.

Behseta, Kass and Wallstrom (2005) distinguished 16 neurons whose firing activities highly resonated with the patterns of the electromyography recordings of two distal muscles: Extensor Carpi Radialis (ECR), a wrist muscle, and Abductor Pollicis Longus (API), a digit or a finger muscle.

To illustrate our methodology, we consider neurons 29 and 32 from the group of neurons studied by Behseta et al. (2005). The activities of these two neurons are worth contemplating; the firing pattern of neuron 29 mimics the activity of the ECR muscle, whereas neuron 32 demonstrates high concordance with the API muscle. Scientifically, the behavior of these neurons is interesting, mainly because their activities may contribute to the debate over the mimicry between motor neurons and hand muscles. In

4

Figure 1, we provide a raster-plot and a Peri-Stimulus Time Histogram (PSTH) for each neuron under both conditions. In the raster-plot, rows are trials and spike times are shown by tickmarks. A time interval of 300 milliseconds (ms) was considered, spanning 200 ms prior to reaching the target (negative time) and 100 ms after reaching the target on the screen (positive time), where 0 denotes the time of reaching the screen. Spikes last for about 1 ms. Thus, by slicing the time interval into 10 ms bins (as in the original study of Matsuzaka et al., 2007) and by pooling the spikes within bins, we create the PSTH plots that provide a graphical tool to explore the overall firing pattern of each neuron. In both conditions, 20 trials were recorded for neuron 32, achieving a total of 52 and 102 firing times under the random and repeating condition, respectively. For neuron 29, 21 and 32 trials were recorded under the random and repeating mode, respectively, resulting in a total of 108 firing times for the random condition and 224 for the repeating condition. The firing times of neuron 32 (top row of Figure 1) behave unimodally for the first 250 ms followed by a burst of activity in the last 50 ms; the pattern appears to be similar under both conditions. This can be contrasted with neuron 29 (bottom row of Figure 1), which demonstrates a bimodal firing pattern in the repeating mode, whereas its firing activity subsides substantially under the random condition, predominantly after the reaching time 0.

A reliable statistical methodology should be able to calibrate the overall similarity of the two firing patterns and to identify sharp differences between the two conditions. Moreover, the methodology should allow comparison on two fronts: general global analysis, enabling the neuroscientist to decide whether the neuron should be considered for further study; and pointwise analysis, to pinpoint differential patterns at specific time points in the experimental time interval.

2.2 The Probability Model for Neuronal Data

The neuronal data we consider can be represented in their most general form through vectors $\{y_{ij}^{(\ell)} : i = 1, ..., N^{(\ell)}; j = 1, ..., n_i^{(\ell)}\}$, where $y_{ij}^{(\ell)}$ is the j-th firing time in the i-th trial under condition ℓ, with $\ell = 1, 2$. However, given our inferential objective of comparison of firing intensities under the two conditions, it suffices to consider modeling for the firing times aggregated over all trials. Hence, for each condition $\ell = 1, 2$, the data vector we work with consists of $\boldsymbol{t}^{(\ell)} = \{t_k^{(\ell)} : k = 1, ..., K^{(\ell)}\}$, where $K^{(\ell)} = \sum_{i=1}^{N^{(\ell)}} n_i^{(\ell)}$ is the total number of firing times from all trials, and $t_k^{(\ell)}$ is the k-th spike time in the aggregated set of firing times for condition ℓ. We assume, without loss of generality, that the point pattern under each trial is observed in the unit time interval, and thus, $0 < t_1^{(\ell)} \leq t_2^{(\ell)} \leq ... \leq t_{K^{(\ell)}}^{(\ell)} < 1$, for $\ell = 1, 2$. (Inference over the time interval where the firing times are recorded can be readily obtained through transformation.)

The postulated probability model for the underlying point process of firing times, corresponding to condition $\ell = 1, 2$, is a non-homogeneous Poisson process (NHPP) with intensity function $\lambda^{(\ell)}(\cdot)$, a non-negative and locally integrable function (i.e., $\int_D \lambda^{(\ell)}(u)\mathrm{d}u < \infty$ for any bounded subset D of the positive real line). The NHPP provides a plausible model for the aggregated firing times based on both empirical evidence as well as theoretical results, which yield that pooled point patterns across a large number of replicated trials follow approximately a NHPP model (see, e.g., the related discussion in Ventura et al., 2002). Section 3.4 discusses an approach to assessment of goodness-of-fit for the NHPP model. In Section 5, the model checking approach is applied to the data from the two neurons discussed in Section 2.1.

2.3 Discussion of Related Literature

In the context of neurophysiological studies as the one discussed in Section 2.1, the

question of main scientific interest is to explore similarities and differences of the neuronal firing rates under the two experimental conditions. In this spirit, analyses of the data of Section 2.1 have been reported in Behseta and Kass (2005) and Behseta et al. (2005). In particular, Behseta and Kass (2005) proposed two methods (based on Bayes factors or a modified Hotelling T^2 statistic) for hypothesis testing of equality of firing intensity functions corresponding to the two conditions. Both methods rely on asymptotic properties of posterior curves fitted to the histograms of firing rates, using Bayesian adaptive regression splines (DiMatteo, Genovese and Kass, 2001). Moreover, the literature on comparing neuronal activities of the motor cortex recordings includes a number of articles on neuronal population coding and the comparison of firing rates under directional movements (e.g., Georgopoulos, 1995). These articles share a common theme: to distinguish a condition (direction) in which the neuron seems to be active the most (preferred direction).

Our modeling framework addresses the more general problem of comparing overall patterns of neuronal activity. We develop a Bayesian nonparametric approach that allows full inference for comparison of neuronal firing intensities, including inference for both global and local differences in the firing intensities under the two conditions. Hence, the approach yields a more general comparison framework than hypothesis testing for equality of the two intensity functions (as in Behseta and Kass, 2005).

An alternative modeling framework is based on an approximation of the NHPP likelihood for the firing times by discretizing time into small intervals so that the data are converted into a binary sequence, with 1/0 indicating presence/absence of a spike in each interval. This approach dates back at least to Brillinger (1988) where the binary sequence is fitted using generalized linear model techniques. More recently, Roca-Pardiñas et al. (2006) and Faes et al. (2008) proposed flexible regression-based

7

methods under this framework. The former paper utilizes bootstrap methods to construct a generalized additive model in which factor-by-curve interactions are examined. The latter article uses a pseudo-likelihood technique to fit a binary regression on each condition over a basis matrix, either a cubic-spline, or a matrix of orthonormal polynomials. Conditions then are distinguished as covariates with the aid of a dummy variable. In contrast to regression-based approaches, our proposed method is based on the actual NHPP likelihood for the point pattern of firing times over continuous time. Moreover, it may be viewed as a practically useful alternative to classical estimation techniques, since it has the capacity to provide full inference for the firing intensities without resorting to asymptotic results, which may be suspect for small to moderate firing time point patterns such as the ones studied in this paper.

3. Bayesian Nonparametric Modeling for Neuronal Firing Intensities

Here, we present the methodology for modeling the intensity of neuronal firing times. We work with a NHPP for the firing times, and in Section 3.1 develop a nonparametric prior model for the NHPP intensity function. Approaches for prior specification and posterior inference are discussed in Sections 3.2 and 3.3, respectively. Section 3.4 describes an approach to assessing goodness-of-fit for the NHPP model.

3.1 The Modeling Approach

In this section, for notational simplicity, we suppress the superscript ℓ included in Section 2.2 to indicate the particular condition. Hence, from a modeling perspective, interest revolves around the NHPP intensity function $\lambda(\cdot)$. The likelihood for $\lambda(\cdot)$

based on the observed firing times $\boldsymbol{t} = \{t_k : k = 1, ..., K\}$ can be expressed as

$$\exp\{-\int_0^1 \lambda(u)\mathrm{d}u\} \prod_{k=1}^{K} \lambda(t_k). \tag{1}$$

We seek a flexible prior model for the intensity function $\lambda(\cdot)$ that can provide full inference for the neuronal firing intensity over the experimental time interval without relying on specific parametric forms or asymptotic arguments. Hence, we argue for the utility of a Bayesian nonparametric approach to modeling $\lambda(\cdot)$. In particular, we use a mixture formulation for a density function directly connected with the intensity function. By casting the modeling in a density estimation framework, we can use a flexible class of nonparametric mixture models that allows relatively easy prior specification and posterior simulation. The approach was developed in Kottas and Sansó (2007) in the context of spatial NHPPs. Here, to address our problem in the context of neuronal data analysis, we formulate a model for NHPPs evolving over time.

The modeling proceeds by working with an equivalent representation for $\lambda(t)$, $t \in (0, 1)$, through the density function $f(t) = \lambda(t)/\gamma$, $t \in (0, 1)$, where $\gamma = \int_0^1 \lambda(u)\mathrm{d}u$. Because the parameter γ provides only the scale for $\lambda(\cdot)$, it is the density function $f(\cdot)$ that controls the shape of the intensity function $\lambda(\cdot)$. Hence a flexible nonparametric prior model for $f(\cdot)$ can capture non-standard intensity shapes, and, in conjunction with a prior for γ, induces a semiparametric prior for $\lambda(\cdot)$.

We employ a Dirichlet process (DP) mixture of Beta densities model for $f(\cdot)$,

$$f(t; G) = \int \mathrm{be}(t; \nu, \tau)\mathrm{d}G(\nu, \tau), \quad G \sim \mathrm{DP}(\alpha, G_0). \tag{2}$$

Here, $\mathrm{DP}(\alpha, G_0)$ denotes the DP prior (Ferguson, 1973) for the random mixing distribution G, with precision parameter α and centering distribution G_0, and $\mathrm{be}(\cdot; \nu, \tau)$ denotes the density of the Beta distribution parameterized in terms of its mean $\nu \in (0, 1)$ and a scale parameter $\tau > 0$, i.e., $\mathrm{be}(t; \nu, \tau) \propto t^{\nu\tau-1}(1 - t)^{\tau(1-\nu)-1}$, $t \in (0, 1)$.

9

Mixtures of Beta densities can approximate arbitrarily well any density defined on a bounded interval (e.g., Diaconis and Ylvisaker, 1985), and thus, in our context, provide a natural modeling framework for the firing intensity function. Moreover, the choice of the Beta density addresses edge effects that could arise, say, with a normal kernel, for point patterns with firing times close to the boundaries of the experimental time interval; for instance, this is the case for at least one of the data sets considered in Section 2.1 (see the lower right panel of Figure 1).

Regarding the DP parameters, we place a gamma prior distribution on α. For, $G_0 \equiv G_0(\nu, \tau)$, we consider independent components, specifically, a uniform distribution on $(0, 1)$ for ν and an inverse gamma distribution for τ (with mean $\beta/(c-1)$, provided $c > 1$). We assign an exponential prior distribution to the scale parameter β, and work with fixed shape parameter c, e.g., $c = 2$, which yields an inverse gamma distribution with infinite variance for the component of G_0 corresponding to τ.

Combining the likelihood in (1) with the prior structure for $\lambda(\cdot)$ discussed above, the full Bayesian model becomes

$$\exp(-\gamma)\gamma^K \left\{ \prod_{k=1}^{K} \int \mathrm{be}(t_k; \nu, \tau)\mathrm{d}G(\nu, \tau) \right\} p(\gamma)p(G \mid \alpha, \beta)p(\alpha)p(\beta), \qquad (3)$$

where $p(G \mid \alpha, \beta)p(\alpha)p(\beta)$ denotes the DP prior for G and for its hyperparameters, and $p(\gamma)$ is the prior for γ. In Section 3.2, we discuss an approach to prior specification for α, β and γ, and in Section 3.3 the posterior simulation method for model (3).

3.2 Prior Specification

Regarding parameter γ, a convenient form for its prior is given by a gamma distribution; the prior parameters can be specified based on the role γ plays as the mean of the NHPP over the observation time interval. Alternatively, and in the interest

of obtaining an *automatic* prior for γ, we use the reference prior approach (see, e.g., Bernardo, 2005). The reference prior for γ is based on the Fisher information from the marginal likelihood for γ, which can be obtained from (3) by integrating out all other parameters over their (proper) priors. Specifically, the marginal likelihood for γ is proportional to $\exp(-\gamma)\gamma^K$, yielding $p(\gamma) \propto \gamma^{-1}1_{(\gamma>0)}$ as the reference prior for γ.

Let $1/d$ be the mean of the exponential prior for β. Having specified c, to choose d, we consider only the kernel (i.e., a single component) of mixture (2), for which the variance is $\nu(1 - \nu)/(\tau + 1)$. A proxy for this variance is $(c - 1)/\{4(c - 1 + d^{-1})\}$, based on marginal prior means for ν and τ under G_0. Let r (≤ 1) be a prior guess at the range of observed firing times ($r = 1$ is the natural default choice). Then, we specify d through $(c - 1)/\{4(c - 1 + d^{-1})\} = (r/6)^2$, where $(r/6)^2$ serves as a range-based estimate of the variance. This simple approach requires a small amount of prior input, and, indeed, yields a fairly noninformative specification, since it is based on the special case of the mixture with a single component, whereas, in applications, more components are needed to capture the intensity function shape.

Finally, prior choice for the DP precision parameter α is facilitated by the role that α plays in controlling the number, $K^* \leq K$, of distinct components in the DP mixture (2) (e.g., Escobar and West, 1995). For instance, for moderately large K, $\mathrm{E}(K^* \mid \alpha) \approx \alpha \log\{(\alpha + K)/\alpha\}$, which can be averaged over the prior for α to obtain $\mathrm{E}(K^*)$.

3.3 Posterior Simulation Method

Model (3) can be expressed in a hierarchical form by introducing mixing parameters, $\boldsymbol{\theta} = \{(\nu_k, \tau_k) : k = 1, ..., K\}$, to break the mixture, i.e., given (ν_k, τ_k), the t_k are independent $\mathrm{be}(t_k; \nu_k, \tau_k)$, and, given G, the (ν_k, τ_k) are i.i.d. from G. The full posterior corresponding to this hierarchical version of model (3) is given by $p(\gamma, G, \boldsymbol{\theta}, \alpha, \beta \mid \boldsymbol{t}) =$

$p(\gamma \mid \boldsymbol{t}) \, p(G, \boldsymbol{\theta}, \alpha, \beta \mid \boldsymbol{t})$, where the marginal posterior $p(\gamma \mid \boldsymbol{t})$ is a gamma distribution, in particular, a gamma$(K, 1)$ under the reference prior for γ.

To explore the posterior distribution $p(G, \boldsymbol{\theta}, \alpha, \beta \mid \boldsymbol{t})$, we use a combination of posterior simulation methods for DP mixtures from Neal (2000) and Gelfand and Kottas (2002). Based on a key result from Antoniak (1974),

$$p(G, \boldsymbol{\theta}, \alpha, \beta \mid \boldsymbol{t}) = p(G \mid \boldsymbol{\theta}, \alpha, \beta) p(\boldsymbol{\theta}, \alpha, \beta \mid \boldsymbol{t}),$$

where the distribution of $G \mid \boldsymbol{\theta}, \alpha, \beta$ is a DP, with precision parameter $\alpha + K$ and centering distribution $G_0^*(\nu, \tau \mid \boldsymbol{\theta}, \alpha, \beta) = (\alpha + K)^{-1} \{ \alpha G_0(\nu, \tau \mid \beta) + \sum_{k=1}^{K} \delta_{(\nu_k, \tau_k)}(\nu, \tau) \}$. Moreover, $p(\boldsymbol{\theta}, \alpha, \beta \mid \boldsymbol{t}) \propto p(\alpha) p(\beta) p(\boldsymbol{\theta} \mid \alpha, \beta) \prod_{k=1}^{K} \text{be}(t_k; \nu_k, \tau_k)$ is the posterior that results after marginalizing G over its DP prior in the hierarchical model. The corresponding joint prior for the mixing parameters, $p(\boldsymbol{\theta} \mid \alpha, \beta)$, can be constructed according to a generalized Pólya urn scheme, which yields, for each k, a prior full conditional distribution for (ν_k, τ_k) with point masses $(\alpha + K - 1)^{-1}$ at the (ν_m, τ_m), $m \neq k$, and the remaining probability $\alpha(\alpha + K - 1)^{-1}$ assigned to G_0 (Blackwell and MacQueen, 1973). We obtain posterior samples from $p(\boldsymbol{\theta}, \alpha, \beta \mid \boldsymbol{t})$ using Markov chain Monte Carlo (MCMC) algorithm 5 from Neal (2000).

Next, given posterior draws $\{\boldsymbol{\theta}_b, \alpha_b, \beta_b\}$ from $p(\boldsymbol{\theta}, \alpha, \beta \mid \boldsymbol{t})$, we can sample G_b from $p(G \mid \boldsymbol{\theta}_b, \alpha_b, \beta_b)$ using the DP constructive definition (Sethuraman, 1994) with a truncation approximation. Specifically, we take $G_b = \sum_{l=1}^{L} w_{lb} \delta_{(\nu'_{lb}, \tau'_{lb})}$, with $w_{1b} = z_{1b}$, $w_{lb} = z_{lb} \prod_{s=1}^{l-1}(1 - z_{sb})$, $l = 2, ..., L-1$, $w_{Lb} = 1 - \sum_{l=1}^{L-1} w_{lb} = \prod_{s=1}^{L-1}(1 - z_{sb})$, where the z_{sb} are independent draws from a Beta$(1, \alpha_b + K)$ distribution, and, independently, the (ν'_{lb}, τ'_{lb}) are independent draws from $G_0^*(\nu, \tau \mid \boldsymbol{\theta}_b, \alpha_b, \beta_b)$. The truncation approximation can be made arbitrarily accurate through appropriate choice of L (Kottas, 2006). For instance, it can be shown that $\text{E}(\sum_{l=1}^{L-1} w_{lb} \mid \alpha_b) = 1 - \{(\alpha_b + K)/(\alpha_b + K + 1)\}^{L-1}$, and thus one

approach would be to choose L that makes, say, $\{(K+\max_b\alpha_b)/(K+1+\max_b\alpha_b)\}^{L-1}$ small to any desired accuracy.

Now, $f_{b0} = \int \mathrm{be}(t_0;\nu,\tau)\mathrm{d}G_b(\nu,\tau) = \sum_{l=1}^{L} w_{lb}\mathrm{be}(t_0;\nu'_{lb},\tau'_{lb})$ is a realization from the posterior of $f(t_0;G)$, for any time point t_0 in $(0,1)$. Hence, if γ_b is a draw from $p(\gamma \mid \boldsymbol{t})$, $\gamma_b f_{b0}$ is a posterior draw for $\lambda(t_0;\gamma,G) = \gamma f(t_0;G)$, the intensity function at t_0. Moreover, the cumulative intensity function is given by $\Lambda(t_0;\gamma,G) = \int_0^{t_0}\lambda(u;\gamma,G)\mathrm{d}u = \gamma\int_0^{t_0} f(u;G)\mathrm{d}u = \gamma\int\mathrm{Be}(t_0;\nu,\tau)\mathrm{d}G(\nu,\tau)$, where $\mathrm{Be}(t_0;\nu,\tau)$ is the cumulative distribution function at t_0 of the $\mathrm{be}(\cdot;\nu,\tau)$ density. Hence, posterior samples for $\Lambda(t_0;\gamma,G)$ arise from $\gamma_b\sum_{l=1}^{L} w_{lb}\mathrm{Be}(t_0;\nu'_{lb},\tau'_{lb})$. Therefore, the computational technique described above yields full posterior inference for the intensity and cumulative intensity functions at any collection of points in the time interval $(0,1)$.

3.4 Model Checking

A possible approach to checking the NHPP assumption for the aggregated neuronal firing times, under the model for the intensity function developed in Section 3.1, is based on the "Time-Rescaling" theorem (e.g., Daley and Vere-Jones, 2003).

Assume that the point pattern $\boldsymbol{t} = \{t_k : k = 1,...,K\}$, with ordered firing times $0 < t_1 \leq t_2 \leq ... \leq t_K < 1$, is a realization from a NHPP with intensity function $\lambda(\cdot)$ and cumulative intensity function $\Lambda(t) = \int_0^t \lambda(u)\mathrm{d}u$. Then, based on the Time-Rescaling theorem, the transformed point pattern $\{\Lambda(t_k) : k = 1,...,K\}$ is a realization from a homogeneous Poisson process with unit rate. Hence, defining $\Lambda(t_0) \equiv \Lambda(0) = 0$, the rescaled times $\Lambda(t_k) - \Lambda(t_{k-1})$, $k = 1,...,K$, are independent exponential random variables with mean 1, and thus, the

$$x_k = 1 - \exp\{-(\Lambda(t_k) - \Lambda(t_{k-1}))\}, \quad k = 1,...,K$$

are independent uniform random variables on $(0, 1)$. Because the above transformations are one-to-one, one can apply a graphical or formal model checking technique that measures agreement between the x_k and the uniform distribution on $(0, 1)$ to evaluate agreement between the firing times t_k and the NHPP model. Note that the approach involves checking of both the NHPP probability model assumption as well as of the statistical model used for the NHPP intensity function, and thus, it can also be used to compare the fit of different models for the NHPP intensity function. On the other hand, it does not seem feasible to distinguish in the results the role of the NHPP assumption from the form of the model for the NHPP intensity. However, the Time-Rescaling theorem offers a practically useful approach to identifying strong deviations from the NHPP model, and, in fact, as discussed in Daley and Vere-Jones (2003), is applicable to more general point process models, where the NHPP intensity function is replaced with the conditional intensity function for the point process.

In the context of neuronal data analysis, Brown et al. (2001) used standard tests and quantile-quantile plots to measure agreement of the estimates for the x_k with the uniform distribution on $(0, 1)$. For a similar illustration with our data, in Section 5 we provide a simple graphical diagnostic of the NHPP assumption working with posterior means for the x_k, which are computed using the posterior samples for the cumulative intensities $\Lambda(t_k)$ obtained as discussed in Section 3.3.

Note that under any Bayesian modeling approach, we obtain an entire posterior distribution for the x_k, which thus increases the scope for either formal or graphical assessment of model goodness-of-fit. In particular, the Bayesian nonparametrics literature for point process modeling appears to be lacking formal model checking methods. Current work is exploring the utility of the Time-Rescaling theorem approach in this direction.

4. Inference for Comparison of Neuronal Firing Intensities Under Two Experimental Conditions

Our focus in applying the methodology of Section 3 is on comparison of firing rates resulting from the same neuron under two distinct experimental conditions. Recall from Section 2.2 that the data consists of $\boldsymbol{t}^{(\ell)} = \{t_k^{(\ell)} : k = 1, ..., K^{(\ell)}\}$, $\ell = 1, 2$. Again, we assume a NHPP model, with intensity function $\lambda^{(\ell)}(\cdot)$, independently under each condition $\ell = 1, 2$. Our interest is in full inference for each $\lambda^{(\ell)}(\cdot)$, and in comparison of $\lambda^{(1)}(\cdot)$ and $\lambda^{(2)}(\cdot)$ over the experimental time interval (again, taken to be $(0, 1)$).

We employ the Bayesian nonparametric approach developed in Section 3 to model the two intensity functions. Hence, letting for $\ell = 1, 2$, $\gamma^{(\ell)} = \int_0^1 \lambda^{(\ell)}(u)\mathrm{d}u$, the prior probability model for the firing intensities, $\lambda^{(\ell)}(t) \equiv \lambda^{(\ell)}(t; \gamma^{(\ell)}, G^{(\ell)}) = \gamma^{(\ell)} f^{(\ell)}(t; G^{(\ell)})$, $t \in (0, 1)$. Here, the prior model for the densities $f^{(\ell)}(t; G^{(\ell)})$ is given by the DP mixture in (2) with independent DP priors for the mixing distributions $G^{(1)}$ and $G^{(2)}$. Therefore, posterior realizations over $(0, 1)$ for each of the two firing intensities result by applying the approach described in Section 3.3 to the full model for $(\gamma^{(1)}, f^{(1)}(\cdot))$ and $(\gamma^{(2)}, f^{(2)}(\cdot))$ given the data $\boldsymbol{t}^{(1)}$ and $\boldsymbol{t}^{(2)}$, respectively.

Note that direct comparison of the intensities $\lambda^{(1)}(\cdot)$ and $\lambda^{(2)}(\cdot)$, using their posterior realizations, is hindered by their different scales, captured by $\gamma^{(1)}$ and $\gamma^{(2)}$, respectively. Our modeling framework provides a natural way to address this issue by working instead with the densities $f^{(1)}(\cdot)$ and $f^{(2)}(\cdot)$. Posterior realizations for $f^{(1)}(\cdot)$ and $f^{(2)}(\cdot)$ can be used in several ways to obtain inference for comparison of the respective firing intensities $\lambda^{(1)}(\cdot)$ and $\lambda^{(2)}(\cdot)$. For instance, in Section 5 we illustrate with posterior point estimates and associated uncertainty bands for the function $f^{(1)}(\cdot) - f^{(2)}(\cdot)$ as well as with the entire posterior $p\{f^{(1)}(t_0; G^{(1)}) - f^{(2)}(t_0; G^{(2)}) \mid \boldsymbol{t}^{(1)}, \boldsymbol{t}^{(2)}\}$ for specific points t_0 in the experimental time interval.

We note that the above inference includes uncertainty assessment that does not rely on asymptotic considerations. Moreover, it can uncover both local and global differences in the neuronal firing intensities and thus provides a more general scope for comparison than formal testing of the hypothesis $\lambda^{(1)}(t) = \lambda^{(2)}(t)$, for all $t \in (0, 1)$, which was addressed in Behseta and Kass (2005).

5. Analysis of Primary Motor Cortex Neuronal Data

We applied the model to the neuronal data described in Section 2.1 following the approaches of Section 3.2 and 3.3 for prior specification and posterior simulation, respectively. The results reported below are based on 10,000 posterior samples obtained after a (conservative) burn-in period of 20,000 MCMC iterations with thinning every 50-th iteration to eliminate autocorrelations. For both neurons, we used the reference prior for γ, set $c = 2$, and also $r = 1$ resulting in an exponential prior for β with mean 8. Moreover, we used gamma priors for α that imply, for both the random and repeating condition, an approximate prior expectation of 4 for the number of distinct mixture components K^*.

For both neurons, we report results from global as well as pointwise analyses. Focusing first on neuron 32, in the top panels of Figure 2, we construct the curves associated with the posterior mean and 95% interval estimates for the intensity functions under the random and the repeating condition. Both intensity functions peak up around the same time interval before the reaching time, followed by a gradual decay, hence depicting a fairly unimodal firing activity. Note that there is an increase in the neuronal firing intensity at the end of the time window that may be due to the anticipation of the upcoming target reach. These posterior estimates provide clear indication for

16

similarity of the firing patterns between the two conditions. However, accurate quantification of the extent of similarity is hindered by the different scale in the intensity function estimates, resulting from the larger number of firing times under the repeating condition. As discussed in Section 4, our approach enables formal comparison between the firing patterns by working on the density scale associated with the two neuronal firing intensities. In this spirit, the lower panel of Figure 2 plots the posterior mean and 95% interval bands for the difference of density functions between the random and repeating conditions (with densities $f^{(1)}(\cdot)$ and $f^{(2)}(\cdot)$, respectively). Arguably, this plot provides a major inferential tool to the neurophysiologist. Note that the 95% posterior bands cover the zero difference for all but a short interval at about 100 ms prior to reaching the screen. This difference may be taken as a sign for distinct firing patterns in the anticipatory phase of the experiment. Even though both neurons were recorded from the primary cortex area of the monkey's brain, their response to the experimental tasks is differential. Matzusaka et al. (2007) argued that "differentially active neurons were more strongly influenced by task mode than by kinematics". In particular, the authors hypothesized that the distinguishable firing patterns of recorded neurons under the two modes can be interpreted as a sign of "functional reorganization" of M1 during the performance of highly practiced tasks.

Figure 3 contains histograms of posterior samples from the difference of density functions between the two conditions, $f^{(1)}(t_0) - f^{(2)}(t_0)$, at nine specified points t_0 in the experimental time interval. These results are useful when the analysis revolves around pointwise comparison between the two conditions. Despite the overall similarity of the two conditions, one can observe a relatively higher neuronal activity in the repeating condition prior to the reaching time followed by a higher firing rate in the random condition after the reaching time.

Figures 4 and 5 provide the same inferential tools for the study of the firing activity from neuron 29. The bottom panel of Figure 4 includes the posterior curves for the difference between the random and the repeating mode that show a significant decrease towards the end of the experimental window, thus capturing the burst of activity after the reaching time in the repeating mode. Also notable is the difference during the first 50 ms of the experiment, which is more readily identified from this plot than through the estimates for the individual firing intensities (upper panels of Figure 4). Posterior distributions for the difference of density functions between the random and repeating modes at 12 specified time points are shown in Figure 5. Again, these posteriors facilitate pointwise analysis, including depiction of increasing levels of posterior uncertainty for time points closer to the boundaries of the experimental time interval.

To study sensitivity to the prior choice for α and β, we also applied the model using priors for α that yield $\mathrm{E}(K^*) \approx 10$, and exponential priors for β with means 35 (corresponding to $c = 2$, $r = 0.5$), 152 (for $c = 20$, $r = 1$), and 665 (for $c = 20$, $r = 0.5$). In all cases, inferences involving the density and intensity functions were largely unaffected. For example, Figures 6 and 7 show results for the difference of density functions between the random and repeating conditions for neuron 32.

Finally, we applied the approach discussed in Section 3.4 to obtain a simple graphical diagnostic for the NHPP assumption under the DP mixture model for the NHPP intensity function. For each of the four point patterns of aggregated firing times (i.e., for both neurons and both conditions), Figure 8 plots the empirical cumulative distribution function of the posterior means for the $x_k = 1 - \exp\{-(\Lambda(t_k) - \Lambda(t_{k-1}))\}$ against the uniform distribution function on $(0, 1)$. The NHPP assumption seems plausible, especially, taking into account the small to moderate sample sizes for the four sets of neuronal firing times.

18

6. Discussion

We have developed a nonparametric Bayesian modeling framework for temporal point patterns assumed to arise from a non-homogeneous Poisson process (NHPP). The approach utilizes a flexible Dirichlet process mixture model for a density function that, up to a scale parameter, specifies the NHPP intensity function.

Our motivating application arises from the problem of obtaining full (and exact) inference for comparison of firing rates of a neuron recorded under two conditions. We achieve this task by aggregating the firing times over all experimental trials, and using the Bayesian nonparametric NHPP model for the aggregated firing times under each of the two conditions. Since the density functions that define the neuronal firing intensities operate on the same scale, the modeling approach enables calibration of posterior differences between the two conditions. Moreover, our fully inferential framework casts the problem in a more general setting than hypothesis testing. The proposed modeling approach has direct application in a large class of neuroscience problems in which comparative studies are performed specifically to measure the nuances in the firing patterns between two (or more) experimental conditions. We have illustrated the methodology with the analysis of neuronal data recorded from the primary motor cortex area of a monkey's brain in the course of an experiment involving a sequence of reaching tasks with the target appearing under two distinct modes.

A practically important extension would be to elaborate the modeling for comparison of neuronal firing activities recorded under more than two conditions. The crux of the inference, which is to build the posterior around the density functions, remains intact. Here, we would also seek an appropriate metric, such as the one arising under Kullback-Leibler distances (e.g., Johnson et al., 2001), to calibrate pairwise dissimilarity between the firing intensities for all conditions of interest.

Moreover, current work is exploring methodological extensions that arise by incorporation of dependent nonparametric prior structures in the modeling for the neuronal firing intensities. One such possibility involves dependent prior models for the mixing distributions in the mixture representation for the densities associated with the firing intensity functions. For instance, if plausible for a particular application, some type of probability order constraint (such as stochastic ordering) can be built in the prior model for the densities under two conditions, or for certain pairs of densities under more than two conditions. (See, e.g., Hanson, Kottas and Branscum, 2008, and further references therein, for nonparametric prior models for stochastic order.)

Another research direction involves model formulations that relax the NHPP assumption for settings where trial-to-trial variation is of interest, and hence one would seek a point process model for the trial-specific firing times. See, e.g., the review paper by Kass et al. (2005) for references to related work. Development of Bayesian nonparametric models in this context is of considerable methodological interest.

Finally, to our knowledge, Bayesian nonparametric methods in the context of the regression-based framework, discussed in Section 2.3, do not exist in the literature. Although, again, this is a different modeling paradigm than the one developed in Sections 3 and 4, it would be of practical interest to compare Bayesian nonparametric inference results under these two distinct modeling scenarios.

ACKNOWLEDGEMENTS

Picard (Center for the Neural Basis of Cognition and University of Pittsburgh) for their permission to use the data analyzed in Section 5.

REFERENCES

Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to nonparametric problems. *Annals of Statistics* **2**, 1152-1174.

Ashe, J., Taira, M., Smyrnis, N., Pellizzer, G., Georgakopoulos, T., Lurito, J.T., and Georgopoulos, A.P. (1993). Motor cortical activity preceding a memorized movement trajectory with an orthogonal bend. *Experimental Brain Research* **95**, 118-130.

Behseta, S. and Kass, R.E. (2005). Testing equality of two functions using BARS. *Statistics in Medicine* **24**, 3523-3534.

Behseta, S., Kass, R.E., and Wallstrom, G.L. (2005). Hierarchical models for assessing variability among functions. *Biometrika* **92**, 419-434.

Bernardo, J.M. (2005). Reference analysis. In *Handbook of Statistics 25*, D.K. Dey and C.R. Rao (eds), 17-90. Amsterdam: Elsevier.

Blackwell, D. and MacQueen, J.B. (1973). Ferguson distributions via Pólya urn schemes. *Annals of Statistics* **1**, 353-355.

Brillinger, D.R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics* **59**, 189-200.

Brillinger, D.R. (1992). Nerve cell spike train data analysis: A progression of technique. *Journal of the American Statistical Association* **87**, 260-271.

Brown, E., Kass, R.E., and Mitra, P.P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature Neuroscience* **7**, 456-461.

Brown, E.N., Barbieri, R., Ventura, V., Kass, R.E., and Frank, L.M. (2001). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation* **14**, 325-346.

Daley, D.J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*, Second Edition. New York: Springer.

Diaconis, P. and Ylvisaker, D. (1985). Quantifying prior opinion (with discussion). In *Bayesian Statistics 2*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (eds), 133-156. Amsterdam: North-Holland.

DiMatteo, I., Genovese, C.R., and Kass, R.E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88**, 1055-1071.

Escobar, M. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577-588.

Faes, C., Geys, H., Molenberghs, G., Aerts, M., Cadarso-Suárez, C., Acuña, C., and Cano, M. (2008). A flexible method to measure synchrony in neuronal firing. *Journal of the American Statistical Association* **103**, 149-161.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209-230.

Gallistel, C.R. (1980). *The Organization of Action: A New Synthesis.* Hillsdale, NJ: Erlbaum.

Gazzaniga, M.S., Ivry, R.B., and Mangun, G.R. (2002). *Cognitive Neuroscience, the Biology of the Mind*, Second Edition. New York: Norton.

Gelfand, A.E. and Kottas, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **11**, 289-305.

Georgopoulos, A.P. (1995). Motor cortex and cognitive processing. In *The Cognitive Neurosciences*, M.S. Gazzaniga (ed), 507-517. Cambridge: MIT Press.

Hanson, T.E., Kottas, A., and Branscum, A.J. (2008). Modeling stochastic order in the analysis of receiver operating characteristic data: Bayesian nonparametric approaches. *Applied Statistics* **57**, 207-225.

Johnson, D.H., Gruner, C.M., Baggerly, K., and Seshagiri, C. (2001). Information-theoretic analysis of neural coding. *Journal of Computational Neuroscience* **10**, 47-69.

Kass, R.E., Ventura, V., and Brown, E.N. (2005). Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology* **94**, 8-25.

Kottas, A. (2006). Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. *Journal of Statistical Planning and Inference* **136**, 578-596.

Kottas, A. and Sansó, B. (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Journal of Statistical Planning and Inference* **137**, 3151-3163.

Matsuzaka, Y., Picard, N., and Strick, P.L. (2007). Skill representation in the primary motor cortex after long-term practice. *Journal of Neurophysiology* **97**, 1819-1832.

Neal, R.M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249-265.

Rigat, F., de Gunst, M., and van Pelt, J. (2006). Bayesian modelling and analysis of spatio-temporal neuronal networks. *Bayesian Analysis* **1**, 733-764.

Roca-Pardiñas, J., Cadarso-Suárez, C., Nácher, V., and Acuña, C. (2006). Bootstrap-based methods for testing factor-by-curve interactions in generalized additive models: assessing prefrontal cortex neural activity related to decision-making. *Statistics in Medicine* **25**, 2483-2501.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639-650.

Ventura, V., Carta, R., Kass, R.E., Gettner, S.N., and Olson, C.R. (2002). Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics* **3**, 1-20.

West, M. (1997). Hierarchical mixture models in neurological transmission analysis. *Journal of the American Statistical Association* **92**, 587-606.
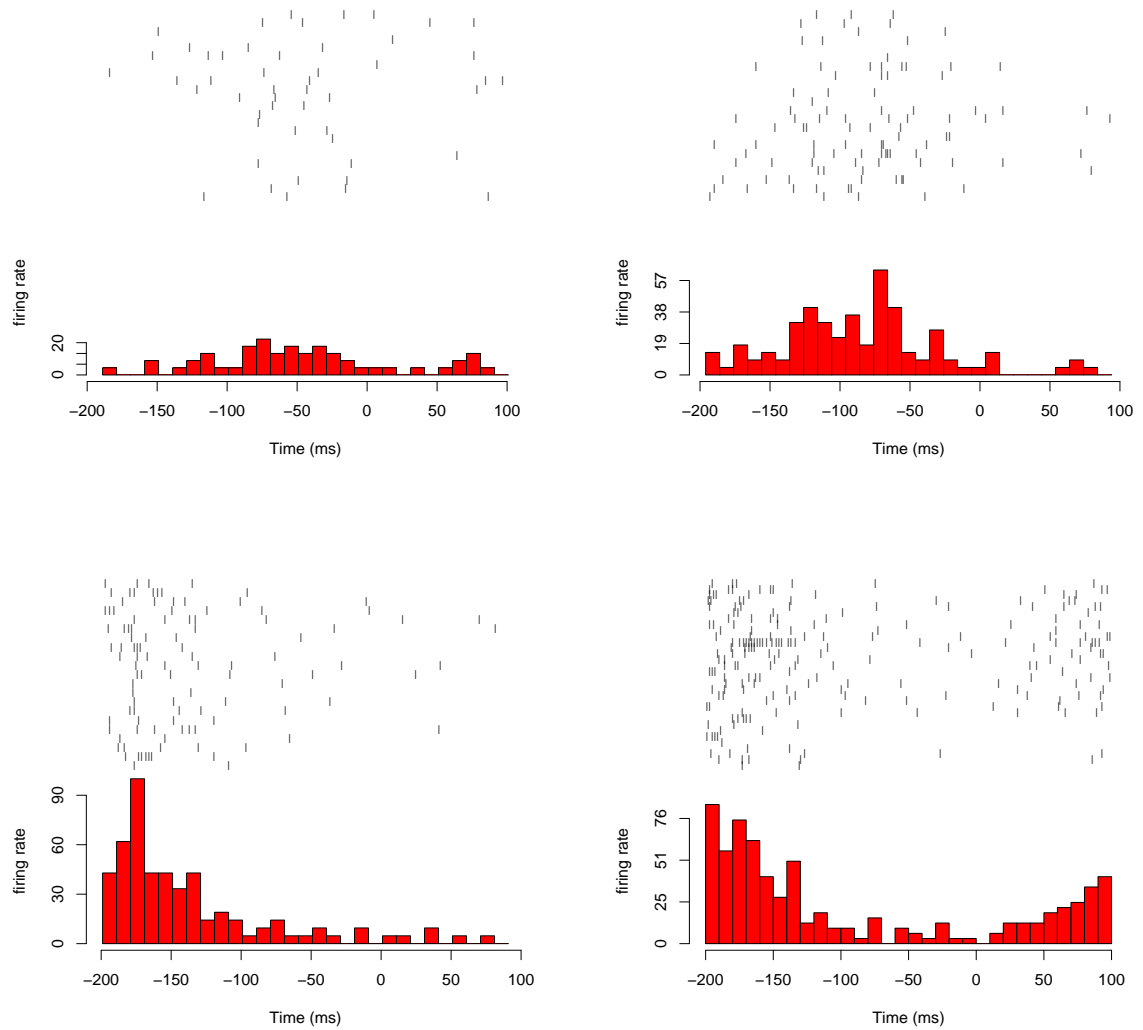
Figure 1: Raster and PSTH plots for the firing times under neurons 32 and 29 (top and bottom row, respectively). The left panels correspond to the random condition, and the right panels to the repeating condition. See Section 2.1 for details.
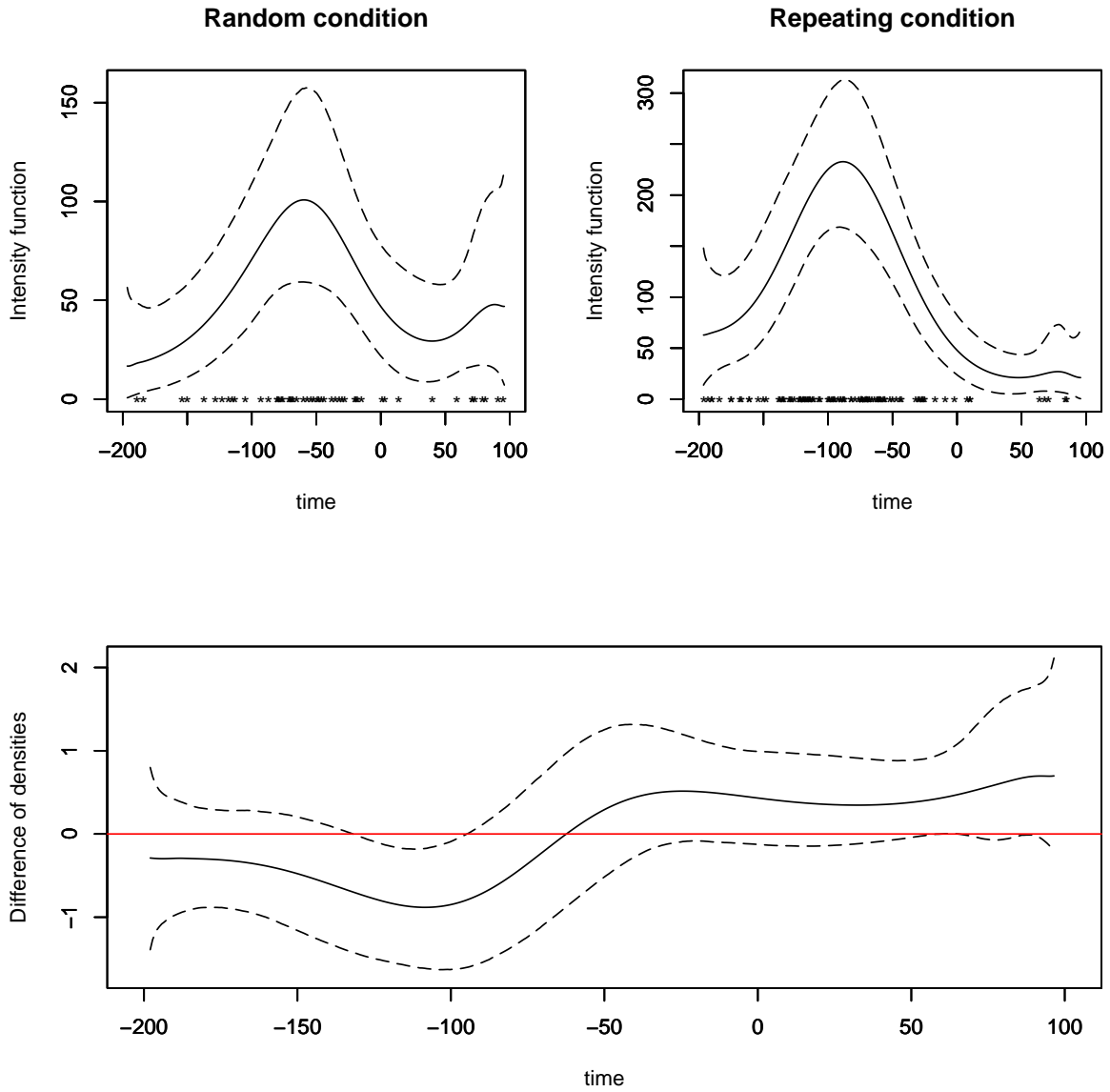
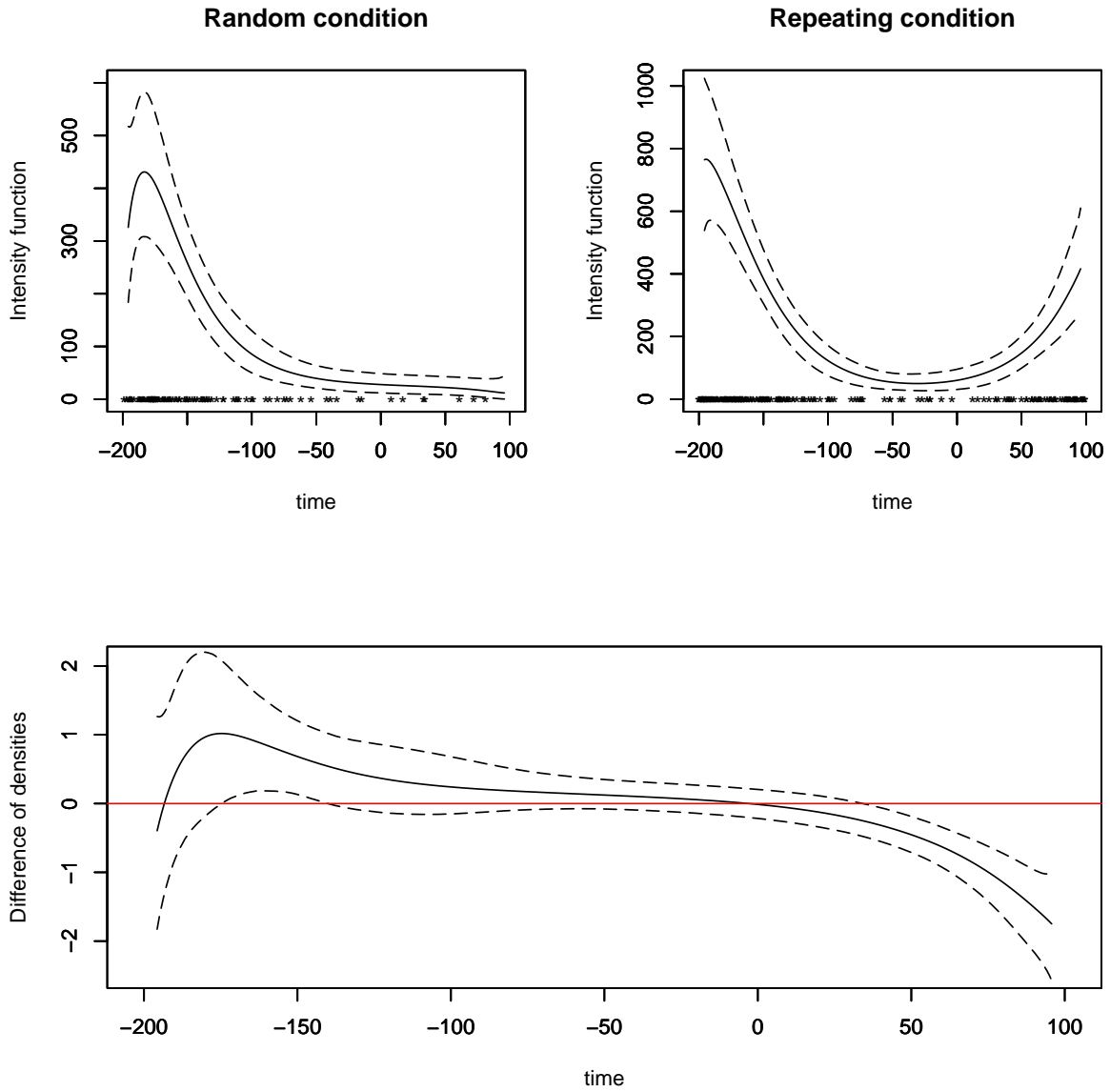Figure 2: Neuron 32. The top panels include posterior mean and 95% interval estimates for the intensity functions under the random and repeating conditions; the corresponding observed firing times are plotted on the horizontal axes. The lower panel plots the posterior mean and 95% interval band for the difference of density functions between the random and repeating condition.
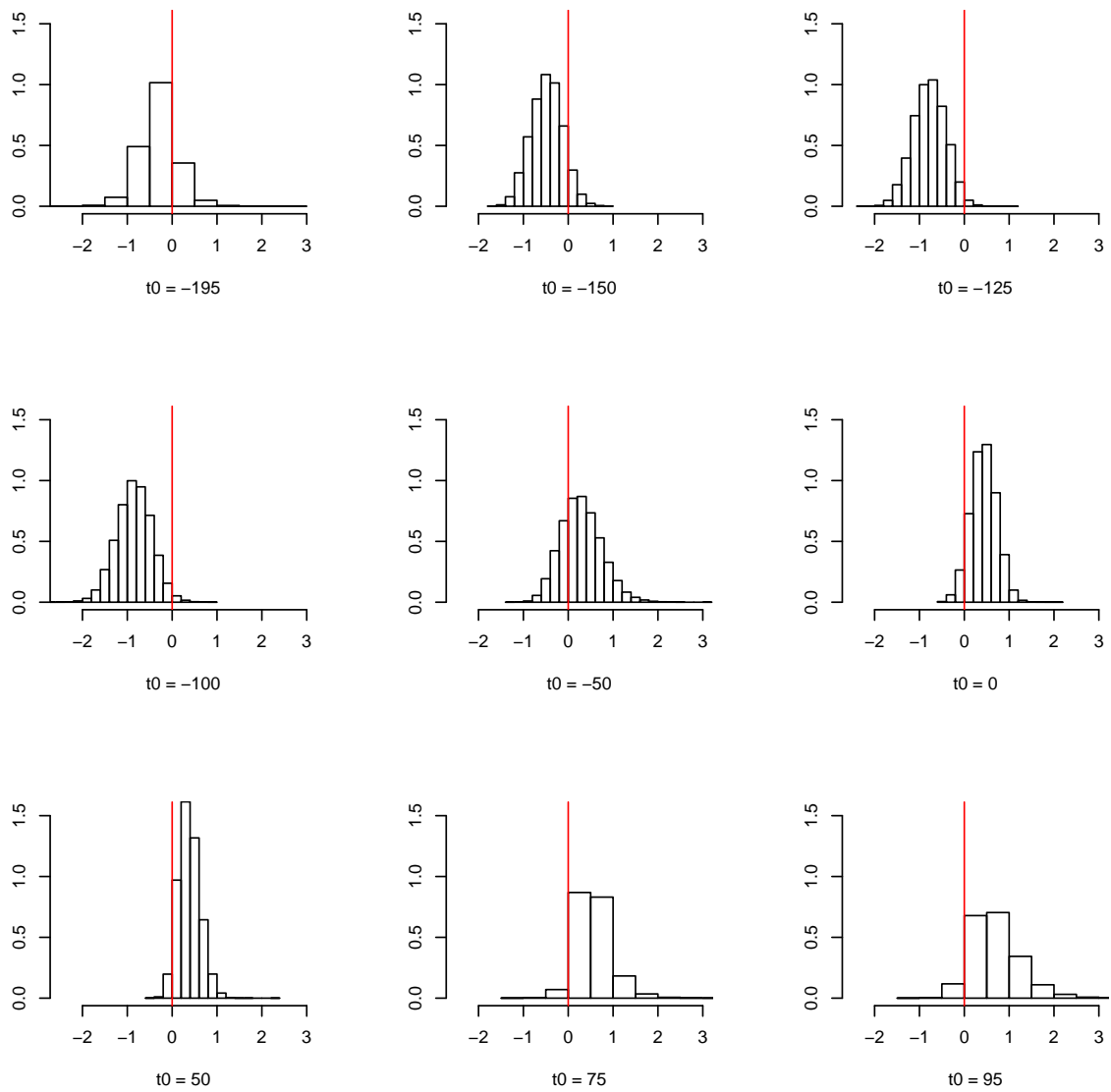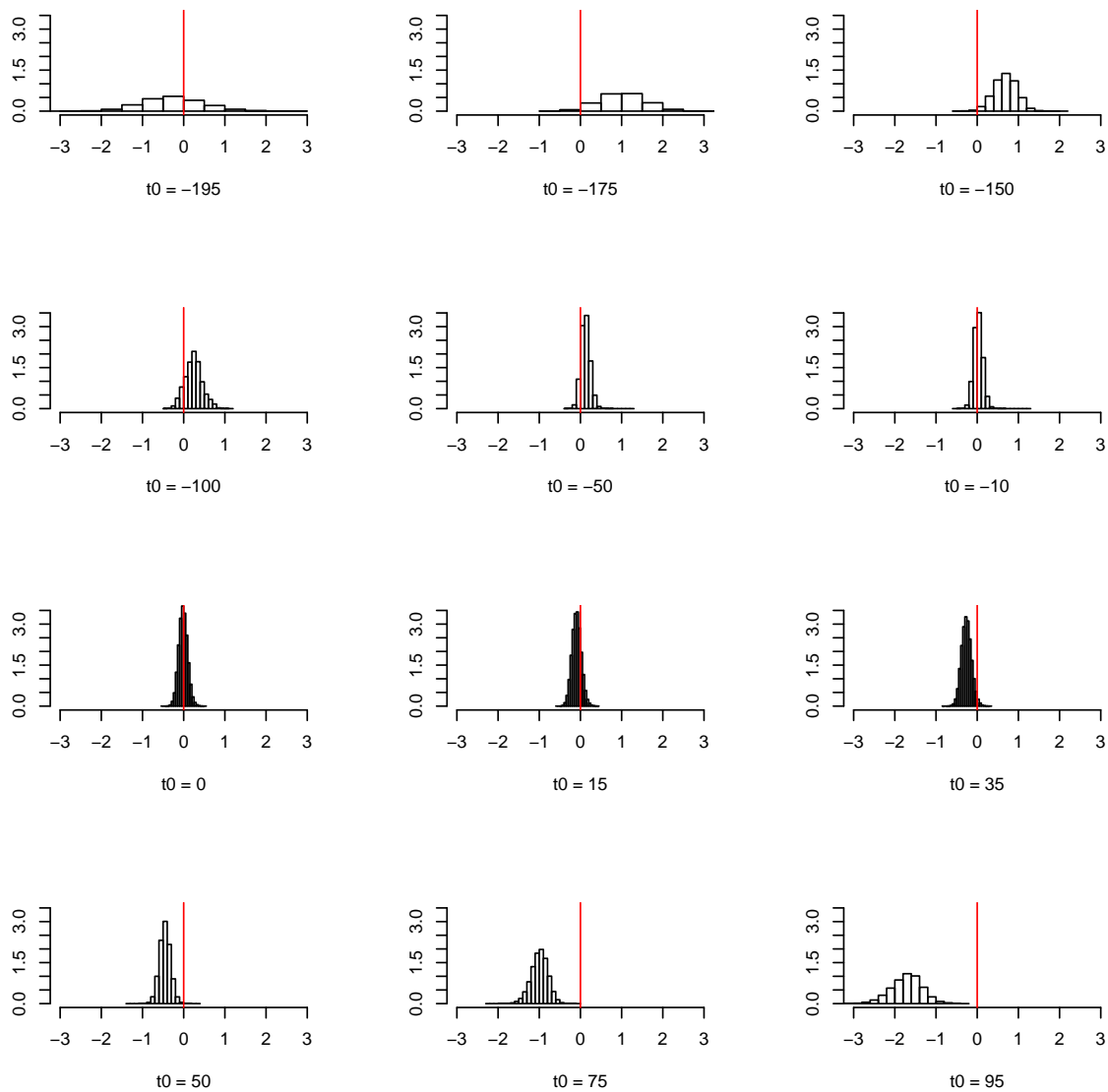
Figure 3: Neuron 32. Posterior distributions for the difference of density functions between the random and repeating condition at 9 time points.

Figure 4: Neuron 29. The top panels include posterior mean and 95% interval estimates for the intensity functions under the random and repeating conditions; the corresponding observed firing times are plotted on the horizontal axes. The lower panel plots the posterior mean and 95% interval band for the difference of density functions between the random and repeating condition.

Figure 5: Neuron 29. Posterior distributions for the difference of density functions between the random and repeating condition at 12 time points.
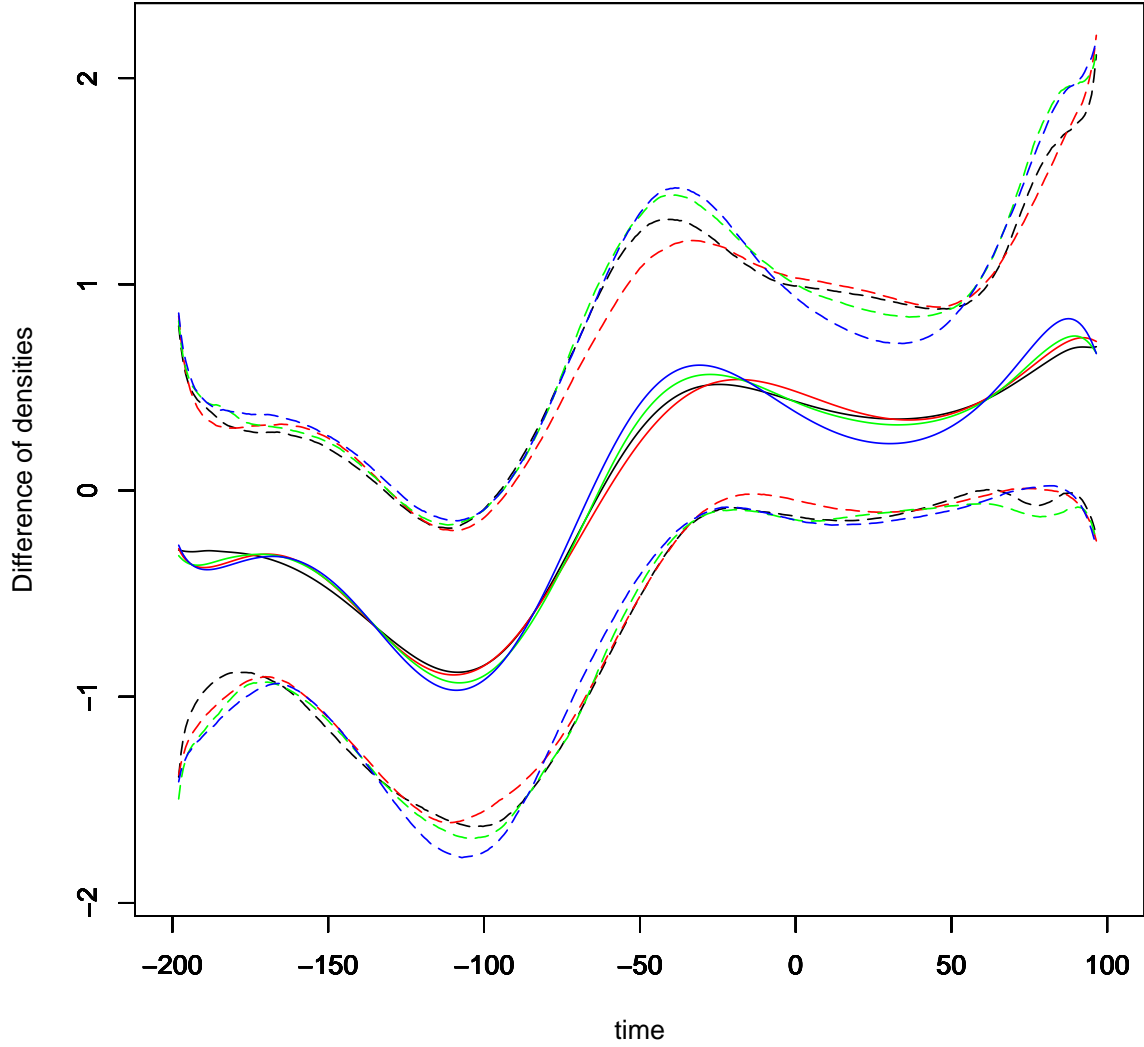
Figure 6: Prior sensitivity analysis for neuron 32. Posterior means (solid lines) and 95% interval bands (dashed lines) for the difference of density functions between the random and repeating condition, under four prior choices for β (for different pairs of c and r). Black color corresponds to the original specification with $c = 2$, $r = 1$; red color to $c = 20$, $r = 1$; green color to $c = 2$, $r = 0.5$; and blue color to $c = 20$, $r = 0.5$.
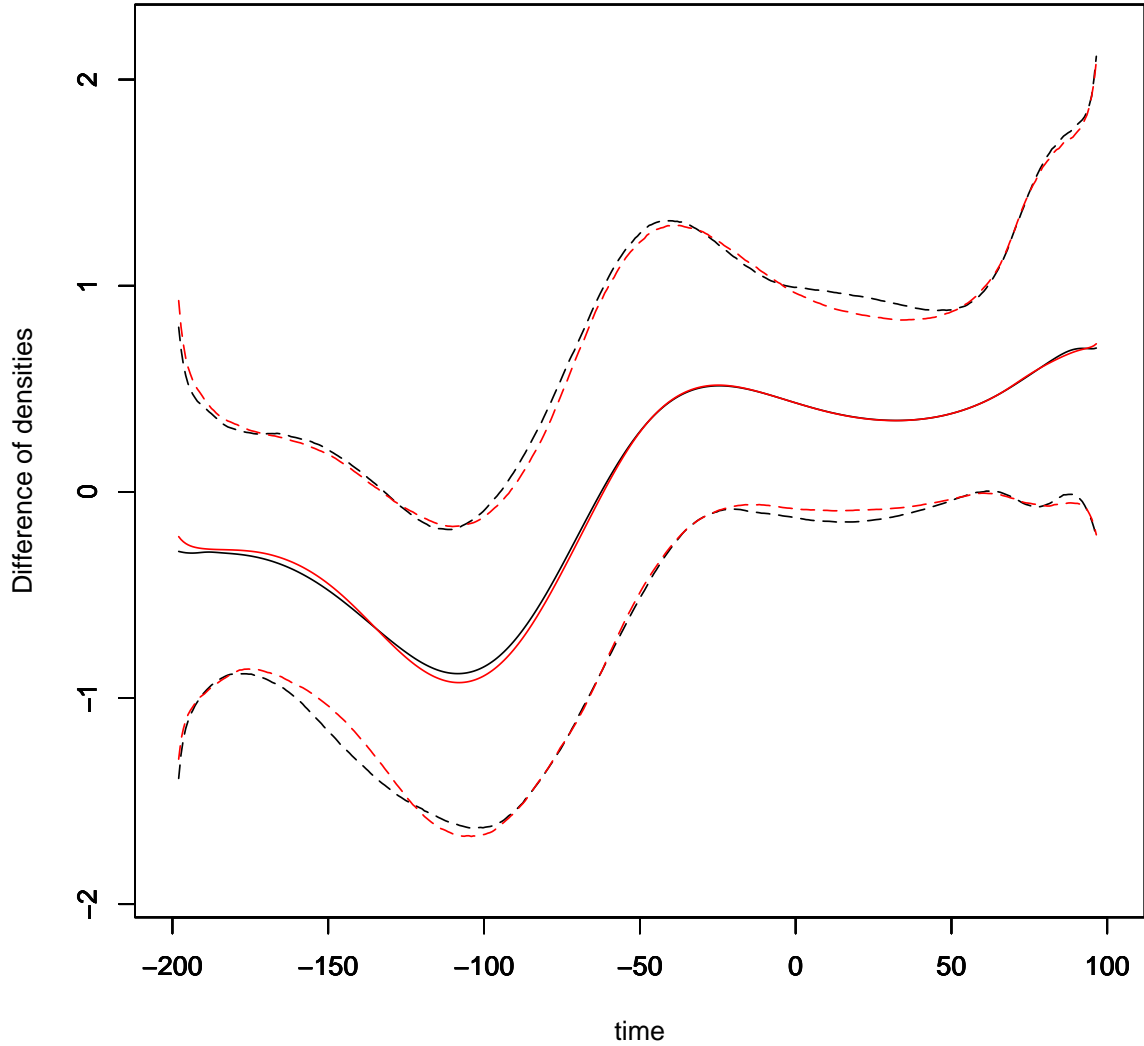
Figure 7: Prior sensitivity analysis for neuron 32. Posterior means (solid lines) and 95% interval bands (dashed lines) for the difference of density functions between the random and repeating condition, under two prior choices for α. Black color indicates results under the original specification, which corresponds to $\mathrm{E}(K^*) = 4$. Red color denotes results under the alternative prior choice that yields $\mathrm{E}(K^*) = 10$.
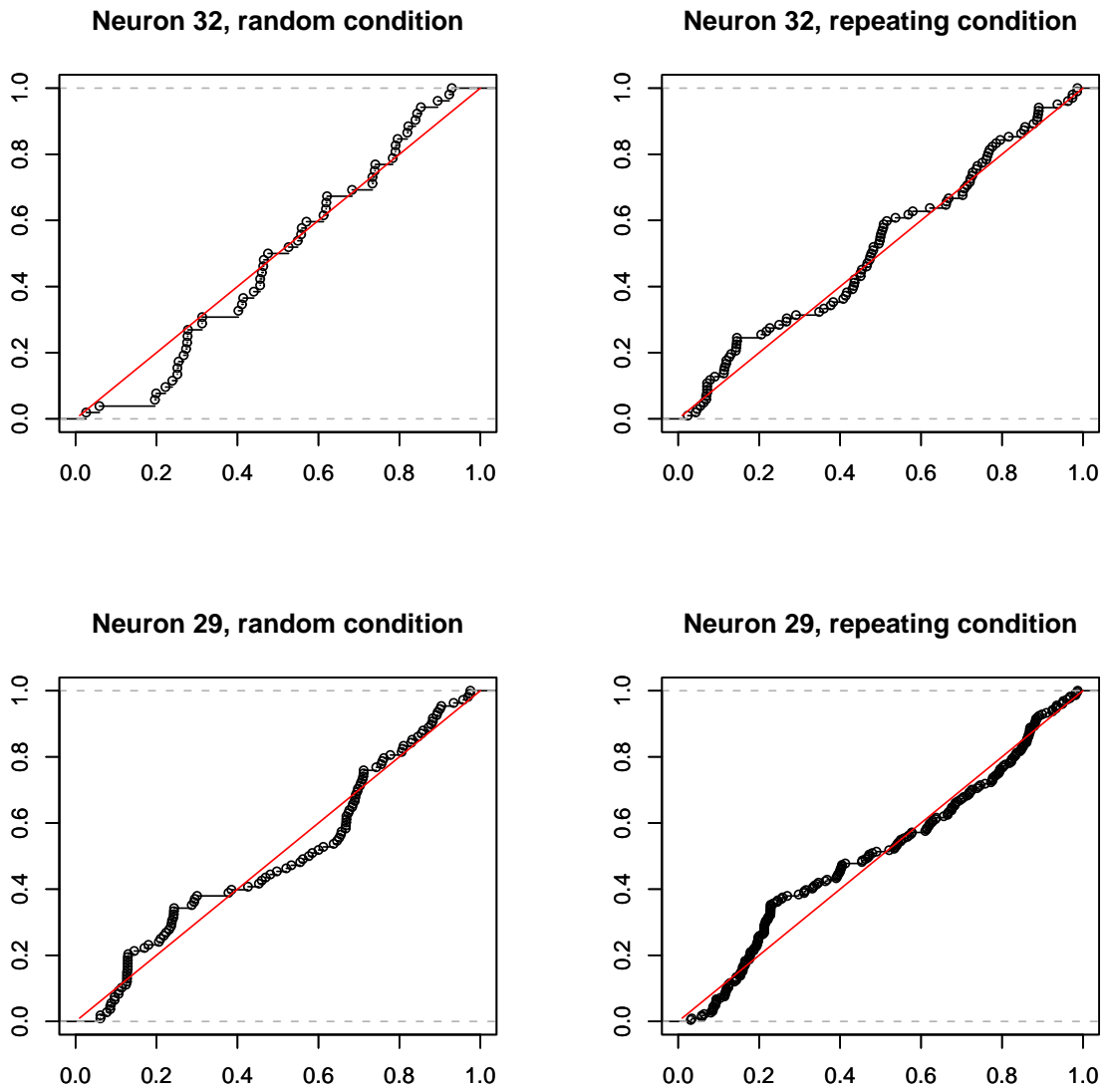
Figure 8: Model checking results for neurons 32 and 29. Empirical cumulative distribution function plots of the posterior means for the $x_k = 1 - \exp\{-(\Lambda(t_k) - \Lambda(t_{k-1}))\}$ against the uniform distribution function on $(0, 1)$ (given by the red line). See Section 3.4 for details.