Testing equality of two functions using BARS[‡]

Sam Behseta[1,*,†] and Robert E. Kass[2]

[1] *Department of Mathematics, California State University Bakersfield, Bakersfield, CA 93311, U.S.A.*
[2] *Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.*

SUMMARY

This article presents two methods of testing the hypothesis of equality of two functions $H_0 : f^1(t) = f^2(t)$ for all t, in a generalized non-parametric regression framework using a recently developed generalized non-parametric regression method called Bayesian adaptive regression splines (BARS). Of particular interest is the special case of testing equality of two Poisson process intensity functions $\lambda^1(t) = \lambda^2(t)$, which arises frequently in neurophysiological applications. The first method uses Bayes factors, and the second method uses a modified Hotelling T^2 test. Both methods are applied to the analysis of 347 motor cortical neurons and, for certain choices of test criteria, the two methods lead to the same conclusions for all but 7 neurons. A small simulation study of power indicates that the Bayes factor can be somewhat more powerful in small samples. The T^2-type test should be useful in screening large number of neurons for condition-related activity, while the Bayes factor will be especially helpful in assessing evidence in favour of H_0. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: Bayes factor; curve-fitting; functional data analysis; inhomogeneous Poisson process; neuronal data analysis; non-parametric regression

1. INTRODUCTION

Bayesian adaptive regression splines (BARS) [1] is a generalized non-parametric regression method that is particularly good at fitting curves with irregular variation. DiMatteo *et al.* showed that BARS could reduce mean squared error below other existing methods, and this new technology has been used in a variety of applications in neurophysiology, imaging, EEG analysis, and genetics [1–7]. Furthermore, BARS has been implemented in *C*, with calling functions in *R* and *S*, in publically available software [8]. A common problem in these curve-fitting scenarios is to determine whether two curves produced under two different but related conditions are the same. Formally, this involves a hypothesis test of the form $H_0 : f^1(t) = f^2(t)$, for all t in some finite interval, where t may represent time or it could

be any explanatory variable to which the response is related. The purpose of this paper is to define and compare two methods of testing this hypothesis in conjunction with BARS.

We are especially interested in applying this new methodology in neurophysiological applications, where the problem becomes one of testing equality of two Poisson process intensity functions $\lambda^1(t)$ and $\lambda^2(t)$. Neurophysiologists examine activity of individual neurons *in vivo* in order to characterize their responsiveness either to some stimulus presented to the animal or to some behaviour the animal is engaged in (such as moving its hand). Typically, many neurons (often hundreds) are examined from a particular brain region during several months of experimentation. The need for a test of $H_0 : \lambda^1(t) = \lambda^2(t)$ arises in two circumstances. First, it may be important to establish that some neurons are not at all responsive to the stimulus or task, so that H_0 might hold in some cases. Second, even when H_0 is not itself interesting, an experimenter still may wish to eliminate from further consideration those neurons for which the data cannot establish conclusively that the neurons *are* responsive. Thus, in this latter case, a screening procedure of some kind is invoked. The methods we discuss here are useful for both purposes.

There is a substantial literature on the general problem of testing equality of two functions, focusing mainly on the special case of non-linear regression. A nice overview is provided by Fan and Lin [9] and a more recent treatment is given by Neumeyer and Dette [10]. The method of Fan and Lin is, in principle, quite general, but in the case of testing equality of Poisson intensity functions would require pre-processing, and furthermore assumes equispaced time points, which is not essential in our approach. An important observation, discussed by Fan and Lin, is that tests based on large numbers of time points may not be very powerful. Our use of BARS, however, drastically reduces the dimensionality of the problem by fitting splines with a relatively small number of knots. After presenting the methodology we report simulation results to substantiate the intuition that the methods we discuss should have good power against smooth but relatively small departures from equality.

BARS uses data $y_1, \ldots, y_p$ obtained at $t = t_1, \ldots, t_p$, with each Y_j assumed to depend probabilistically on $f(t_j)$ through the model

$$
\begin{aligned}
Y_j &\sim p(y_j | \theta_j, \zeta) \\
\theta_j &= f(t_j)
\end{aligned}
\tag{1}
$$

with f being a spline having an unknown number of knots at unknown locations $\xi_1, \ldots, \xi_k$. Model (1) includes a vector of nuisance parameters ζ to indicate generality. In the Poisson case, however, there are no nuisance parameters and in what follows we will, for convenience, drop ζ from our notation. Writing $f(t)$ in terms of basis functions $b_{\xi,h}(t)$ as $f(t) = \sum_h b_{\xi,h}(t)\beta_{\xi,h}$ the function evaluations $f(t_1), \ldots, f(t_p)$ may be collected into a vector $(f(t_1), \ldots, f(t_p))^{\mathrm{T}} = X_\xi \beta_\xi$, where X_ξ is the design matrix and β_ξ is the coefficient vector. For a given knot set $\xi = (\xi_1, \ldots, \xi_k)$ model (1) poses a relatively easy estimation problem; for exponential-family responses (such as Poisson) it becomes a generalized linear model. The hard part of the problem is determining the knot set ξ, and using the data to do so provides the ability to fit a wide range of functions (as reviewed by Hansen and Kooperberg [11]). In Section 2 we briefly review the strategy used by BARS.

Now suppose we were to choose a particular knot set ξ and use it for both of the two curve-fitting problems, namely; determining ξ and utilizing the data to do so. Then H_0 could be re-expressed as $\beta_\xi^1 = \beta_\xi^2$ and the likelihood ratio test could be applied to test H_0 *versus* the

alternative $\beta_\xi^1 \neq \beta_\xi^2$. Although this is, at first glance, an appealing possibility, it is not clear how to carry out such a determination of a single knot set. Furthermore, using the likelihood ratio for a single knot set would ignore the uncertainty due to knot set selection. To solve the first problem we developed a version of BARS that fits $f^1(t)$ and $f^2(t)$ together, under the assumption that they share the same knot set. To circumvent the second problem we defined a suitable Bayes factor for testing H_0 against the same-knot-set alternative H_A, which takes account of the posterior distribution on ξ. Implementation details for this Bayes factor are given in Section 3. While we have obtained very good results using the Bayes factor, it does require both functions to use the same knot set (as would the likelihood ratio test) which may be restrictive. One reason we are concerned about using the same knot set for both functions is that in our recent study of methods for assessing variability among many curves we found [2], somewhat surprisingly, that a random-coefficient hierarchical model that assumed the same knot sets among all curves did not perform as well as an alternative approach based on fitting the curves separately. Therefore, in Section 4, we introduce our second procedure, which begins with separate BARS fits for $f^1(t)$ and $f^2(t)$ and produces a p-value based on an analogue of Hotelling's T^2 for testing equality of two multivariate Normal means.

Figure 1 displays data from 2 primary motor cortical neurons, each recorded under two experimental conditions in a sequential reaching task [12]. These 2 neurons were selected
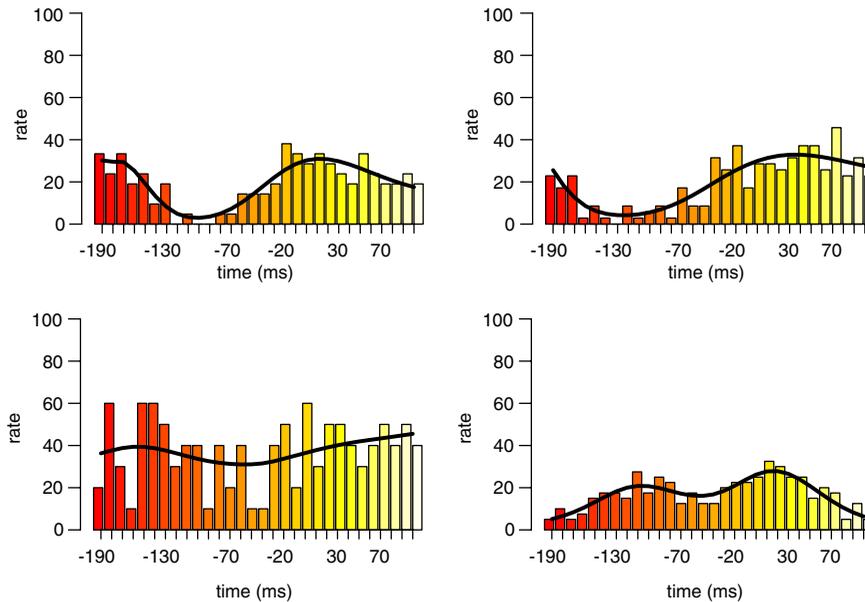

Figure 1. Responses of 2 motor cortical neurons under repeating-mode and random-mode conditions. BARS intensity functions (neuronal spikes/s) are overlaid on peristimulus time histograms. Horizontal axes are experimental time (in milliseconds), with 0 corresponding to the time at which the monkey's hand leaves the previous button and begins to move to the new button. Top left: Neuron 1 in random-mode condition. Top right: Neuron 1 in repeating-mode condition. Bottom left: Neuron 2 in random-mode condition. Bottom right: Neuron 2 in repeating-mode condition. It appears from inspection that neuron 1 responds similarly in the two conditions while neuron 2 responds differently.

from 347 neurons examined in this experiment. Overlaid on the data histograms are the BARS fits. According to our methods, one neuron clearly responds differently to the two experimental conditions, while for the other neuron the data are inconclusive. In Section 5 we discuss the details and we also summarize our comparison of the two methods across all 347 neurons. In Section 6 we present a small simulation study of the two methods.

2. BACKGROUND ON BARS

BARS is an MCMC-based algorithm that samples from a suitable approximate joint posterior distribution on the knot set ξ and number of knots k. This, in turn, produces samples from the posterior on the space of splines. In practice, cubic splines and the natural spline basis have been used in most applications. BARS could be viewed as a powerful engine for searching for an 'optimal' knot set, but because it generates a posterior on the space of splines it produces an improved spline estimate based on model averaging [13] and it also provides uncertainty assessments. Importantly, the uncertainty assessment includes uncertainty about knot placement.

The standard output of BARS is a sample from the joint posterior of (ξ, k) but, for notational convenience, in most of our presentation we leave k implicit and discuss the posterior on ξ (k is the length of the vector ξ). Details of implementation are described in Reference [8]. In brief, key features of the MCMC implementation of BARS include (i) a reversible-jump chain on ξ after integrating the marginal density

$$p(y|\xi) = \int p(y|\beta_\xi, \xi)\pi(\beta_\xi|\xi)\,\mathrm{d}\beta_\xi \qquad (2)$$

(where $y = (y_1, \ldots, y_p)$), the integration being performed exactly for Normal data and approximately, by Laplace's method, otherwise, (ii) continuous proposals for ξ, and (iii) a locality heuristic for the proposals that attempts to place potential new knots near existing knots. BARS explores the space of generalized regression models defined by ξ and k and the prior on k can, in some cases, control the algorithm in important ways [1, 4, 11].

The essential idea of using reversible-jump MCMC to select knots was suggested by Denison *et al.* [14], following the lead of Green [15], who discussed the special case of change-point problems. However, aspects of BARS outlined in (i)–(iii) distinguish it from (and improve upon) the method of Denison *et al.* (see Reference [14]). The first implementation feature, item (i) above, introduces an analytical step within the MCMC partly to simplify the problem of satisfying detailed balance and partly for the sake of MCMC efficiency (which is generally increased when parameters are integrated; see Reference [16]). In addition, BARS takes advantage of the high accuracy of Laplace's method in this context. In doing so the 'unit-information' prior discussed by Kass and Wasserman [17] and Pauler [18] has been used (as π in (2)), and this gives the interpretation that the algorithm is essentially using the Bayes information criterion (BIC) to define a Markov chain on the knot sets. We return to the unit-information prior and BIC below. The importance of performing integral (2), at least approximately, has been stressed by Kass and Wallstrom [4]. Continuous proposals and the locality heuristic (items (ii) and (iii)) together allow knots to be placed close to one another, which is advantageous when there is a sudden jump in the function. It might be thought that non-differentiable curves would require placement of multiple knots at the same location, but

DiMatteo *et al.* found (and showed in their paper) that it was unnecessary to modify BARS in this way: the algorithm places knots essentially on top of each other when necessary.

3. BAYES FACTOR

In conjunction with (1) we consider the null hypothesis $H_0 : \beta_\xi^1 = \beta_\xi^2$ *versus* the alternative $H_A : \beta_\xi^1 \neq \beta_\xi^2$, but the knot set ξ remains unknown. We wish to test H_0 based on data $y^1 = (y_1^1, \ldots, y_{p1}^1)$ and $y^2 = (y_1^2, \ldots, y_{p2}^2)$. The Bayes factor is deceptively easy to write. It is given by

$$B = \frac{\int p(y^1|\beta_\xi, \xi) p(y^2|\beta_\xi, \xi) \pi(\beta_\xi, \xi) \, \mathrm{d}\beta \, \mathrm{d}\xi}{\int p(y^1|\beta_\xi^1, \xi) p(y^2|\beta_\xi^2, \xi) \pi(\beta_\xi^1, \beta_\xi^2, \xi) \, \mathrm{d}\beta_\xi^1 \, \mathrm{d}\beta_\xi^2 \, \mathrm{d}\xi}$$

There are two issues in implementing this Bayesian hypothesis test: First, BARS must be modified so that the same knots are used for both curves. Second, B must be computed from the posterior sample produced by BARS. We discuss the latter computation in the next two subsections, then return to modification of BARS.

3.1. Computation of B via posterior simulation

BARS produces a sample from the joint posterior on ξ and k. The key step that allows the computation to proceed is to assume the two hypotheses are equally likely, rewrite the Bayes factor in terms of the posterior odds

$$B = \frac{P(H_0|y^1, y^2)}{1 - P(H_0|y^1, y^2)}$$

and then write the posterior probability $P(H_0|y^1, y^2)$ in the form

$$P(H_0|y^1, y^2) = \int P(\beta_\xi^1 = \beta_\xi^2|\xi, y^1, y^2) p(\xi|y^1, y^2) \, \mathrm{d}\xi$$

This says that the posterior probability of H_0 may be obtained as the posterior expectation

$$P(H_0|y^1, y^2) = E_{\xi|y^1, y^2}[P(\beta_\xi^1 = \beta_\xi^2|\xi, y^1, y^2)] \tag{3}$$

Thus, from the posterior sample of knot sets $\xi^{(g)}$, for $g = 1, \ldots, G$ we compute

$$P(H_0|y^1, y^2) \approx \frac{\sum_g P(\beta_\xi^1 = \beta_\xi^2|\xi^{(g)}, y^1, y^2)}{G} \tag{4}$$

In other words, we are able to compute the model-averaged posterior probability as an average of model-fixed posterior probabilities. Equation (4) is valuable because, as we show in the next subsection, it is straightforward to compute $P(\beta_\xi^1 = \beta_\xi^2|\xi^{(g)}, y^1, y^2)$, approximately, via BIC.

3.2. BIC and posterior probability

Conditionally on the knot set ξ, $H_0 : \beta^1_\xi = \beta^2_\xi$ defines a model nested within the alternative H_A. The theory in Reference [17] therefore applies and BIC may be used to approximate the posterior probability $P(\beta^1_\xi = \beta^2_\xi | \xi^{(g)}, y^1, y^2)$.

The Bayes information criterion (BIC) for a model M having parameter vector θ is given by

$$\mathrm{BIC}(M) = -2 p_M(y | \hat\theta, M) + d(M) \log n$$

where $d(M) = \dim(\theta)$ and n is the sample size, here, the sum of the two experiments. When applied to H_0 and H_A, conditionally on ξ, we have

$$\mathrm{BIC}_\xi(H_0) = -2 p(y^1 | \hat\beta_\xi) p(y^2 | \hat\beta_\xi) + d_0 \log n$$

and

$$\mathrm{BIC}_\xi(H_A) = -2 p(y^1 | \hat\beta^1_\xi) p(y^2 | \hat\beta^2_\xi) + d_A \log n$$

where d_0 and d_A are the respective dimensionalities of the spline models under H_0 and H_A. This produces the difference $\mathrm{BIC}_\xi(H_0) - \mathrm{BIC}_\xi(H_A)$ given by

$$\mathrm{BIC}_\xi = 2(p(y^1 | \hat\beta^1_\xi) p(y^2 | \hat\beta^2_\xi) - p(y^1 | \hat\beta_\xi) p(y^2 | \hat\beta_\xi)) - (d_A - d_0) \log n \tag{5}$$

The unit-information prior discussed by Kass and Wasserman [17] is a multivariate Normal distribution. It has the interpretation that the amount of information in the prior represented by its covariance matrix is equal to the amount of information in one data observation. Adopting it for β^1_ξ and β^2_ξ yields the approximation

$$P(\beta^1_\xi = \beta^2_\xi | \xi, y^1, y^2) \approx \frac{\exp(\frac{1}{2}\mathrm{BIC}_\xi)}{1 + \exp(\frac{1}{2}\mathrm{BIC}_\xi)} \tag{6}$$

Kass and Wasserman [17] showed that the BIC-based approximation has a theoretical error of order $O(n^{-1/2})$ and is highly accurate in practice. Pauler [18] extended the argument to regression models and again showed the approximation to by highly accurate in practice.

The posterior probability of H_0 may thus be computed, approximately, by combining equations (4), (6), and (5).

3.3. Modification of BARS

It is straightforward to modify BARS so that it may be used to fit two functions simultaneously, constrained to have the same knots. The marginal density, or integrated likelihood, in (2) is replaced with

$$p(y^1, y^2 | \xi) = \int p(y^1 | \beta^1_\xi, \xi) p(y^2 | \beta^2_\xi, \xi) \pi(\beta^1_\xi, \beta^2_\xi | \xi) \, \mathrm{d}\beta_\xi$$

For non-Normal data this integral is approximated using BIC, as in the usual implementation of BARS. In the reversible-jump MCMC what is needed is the ratio of two such integrated likelihoods. In this case the log ratio is approximated by $\frac{1}{2}\mathrm{BIC}_\xi$, defined analogously to (5).

4. A GAUSSIAN PROCESS TEST

In addition to a sample from the posterior from the joint posterior on ξ and k, BARS can produce a sample from the posterior on a set of function evaluations $U = (f(\tilde{t}_1), f(\tilde{t}_2), \ldots, f(\tilde{t}_p))$ along a grid of points $\tilde{t}_1, \tilde{t}_2, \ldots, \tilde{t}_p$. Let us write the posterior mean produced by BARS as $\hat{U}$. For sample sizes typically used with BARS, $\hat{U}$ will be approximately multivariate Normal with covariance matrix Σ, also produced by BARS. Assuming this approximate Normality holds for each of the two functions we write

$$\hat{U}^j \sim \mathrm{N}(U^j, \Sigma^j) \tag{7}$$

for $j = 1, 2$, and the null hypothesis may now be written $H_0 : U^1 = U^2$.

If we had $\Sigma^1 = \Sigma^2$ and we estimated this covariance matrix with a pooled sample covariance under the assumption it was of full rank, then the hypothesis could be tested using Hotelling's T^2. However, not only do we have $\Sigma^1 \neq \Sigma^2$ (and these are given by BARS and thus are not estimated with the sample covariance matrix) but these matrices are not of full rank (the grid size p is likely to be considerably larger than the dimensionality of the spline representation). We therefore modify the T^2 statistic. We first rewrite the variance of $\hat{U}^1 - \hat{U}^2$ in terms of its spectral decomposition

$$\Sigma^1 + \Sigma^2 = P\Lambda P^{\mathrm{T}}$$

where Λ is the diagonal matrix of eigenvalues. We then suppose there are k positive eigenvalues and write the sub-matrices corresponding to these positive eigenvalues as P_k and Λ_k. The modified test statistic becomes

$$T_k^2 = (\hat{U}^1 - \hat{U}^2)^{\mathrm{T}} P_k^{\mathrm{T}} \Lambda_k^{-1} P_k (\hat{U}^1 - \hat{U}^2)$$

Under H_0 we have $P_k(\hat{U}^1 - \hat{U}^2) \sim \mathrm{N}(0, \Lambda_k)$ and therefore we also have

$$T_k^2 \sim \chi_k^2$$

Thus, p-values are obtained immediately. In practice, the value of k may be determined by examining the magnitude of successive eigenvalues λ_i and stopping when

$$\frac{\lambda_{k+1}}{\lambda_1 + \cdots + \lambda_k} < \varepsilon$$

Because the grid of points could be arbitrarily fine, i.e. $\tilde{t}_1, \tilde{t}_2, \ldots, \tilde{t}_p$ could be arbitrarily close together, we are effectively assuming the posterior mean function produced by BARS is a Gaussian process. We therefore call this procedure the Gaussian process test. It is important to note that this test explicitly accounts for the uncertainty in knot selection, because that uncertainty is incorporated into the posterior covariance matrices Σ^j.

5. DATA ANALYSIS

The data in Figure 1 came from a study of primary motor cortex neurons in monkeys during two conditions of a sequential pointing task [19]. Relevant experimental details are summarized

Table I. Comparison of Bayes factor (posterior probability $P(H_0|y^1, y^2)$) and Gaussian process test (T_k^2) among 347 neurons.

	$T_k^2, p < 0.01$	$T_k^2, p > 0.01$
$P(H_0|y^1, y^2) < 0.1$	313	6
$P(H_0|y^1, y^2) > 0.1$	1	27

With the cut-offs $P(H_0|y^1, y^2) < 0.1$ and $p < 0.01$ the two tests disagreed on only 7 neurons.

in the PhD thesis of Behseta [20] and some analysis has been reported in Reference [21]. In brief, firing times of a single neuron were recorded while a monkey completed the task, with increased firing rate of the neuron indicating increased functional activity. The task required the monkey to touch a particular sequence of 3 illuminated buttons, among 5 buttons in all. In the first experimental condition the button-touches were in a predetermined and highly practiced order ('repeating mode'), while in the second condition they were in random order ('random mode'). The experiment was repeated thousands of times, over the course of more than a year, while recordings were made on single neurons. On a given day several new neurons were typically examined. (It was not possible to re-examine neurons across days.) A standard practice is to aggregate firing times for a given neuron across experimental replications into 10 ms time bins, thereby reducing the data to a count for each bin (the information lost being negligible), and this is the form in which we have analysed the data here. Figure 1 displays the resulting histograms for 2 of the neurons (out of a total of 347) over a 300 ms period, with BARS fits overlaid. (The histograms have been normalized by dividing by the number of experimental replications for each neuron, thereby making the units events/s/replication, which are the units associated with the Poisson process intensity functions.) For a general discussion of statistical methods in a related neurophysiological context, see Reference [22]. Among other things, that work verified that it is safe to treat such aggregated and binned data as generating Poisson-distributed counts.

The fitted intensity functions from the first neuron, in the top two panels of Figure 1, do not appear to be much different, while those from the second neuron, in the bottom two panels, do seem to be different. It is useful to have formal hypothesis tests for several reasons. First, there is non-trivial noise due to relatively limited sample sizes: the number of trials for the first neuron in the repeating and random conditions were $r = 21$ and 19 while for the second neuron there were $r = 46$ and 12. Such relatively small numbers of trials are common in neurophysiological research. Second, with large numbers of neurons to examine (347 in this case) it is helpful to have an automated procedure. Third, as we indicated earlier it is sometimes of interest to evaluate the evidence *in favour of* H_0. For this purpose the Bayes factor is particularly helpful because the *p*-value cannot separate failure to reject due to insufficient data from failure to reject due to evidence in favour of H_0. (See Reference [13], for further discussion of this point.)

We applied the Bayes factor and Gaussian process tests to each of the 2 neurons. For the first neuron, we obtained $k = 8$ for the degrees of freedom of the χ^2 statistic as explained in Section 4 followed by $P(H_0|y^1, y^2) = 0.43$ and $T_8^2 = 1.6, p = 0.93$. Taking account of the standard guidance for interpreting Bayes factors [13], these two methods both lead to the

conclusion that there is no evidence against H_0. In addition, the Bayes factor indicates no evidence in favour of H_0. Thus, scientifically, there are insufficient data to draw any conclusion for the first neuron, but it would be reasonable to eliminate this neuron from any further analysis that involved only neurons that discriminated between the two conditions. For the second neuron we obtained $P(H_0|y^1, y^2) = 0.040$ and $T_8^2 = 20.1$, $p = 0.010$, supporting the apparent conclusion of significant difference based on visual examination of the figure.

The agreement of the two methods for these 2 neurons begs the question as to how often the methods agree for the whole data set. The results are displayed in Table I, where it is seen that the two methods almost always lead to the same conclusion (for the criteria $P(H_0|y^1, y^2) < 0.1$ and $p < 0.01$).

6. SIMULATION STUDY OF POWER

To examine the relative merits of the Bayes factor and Gaussian process tests we conducted a small power study in the context of testing equality of Poisson process intensity functions. We created two sets of distinct true functions and used three data scenarios. The functions were chosen for convenience but were designed to reflect the kind of modest effects we often see in neuronal data. The sample sizes are also realistic.

Figure 2 displays the two sets of intensity functions we chose, prior to scaling. In the figure they are shown as probability densities. In the simulations these densities were multiplied by
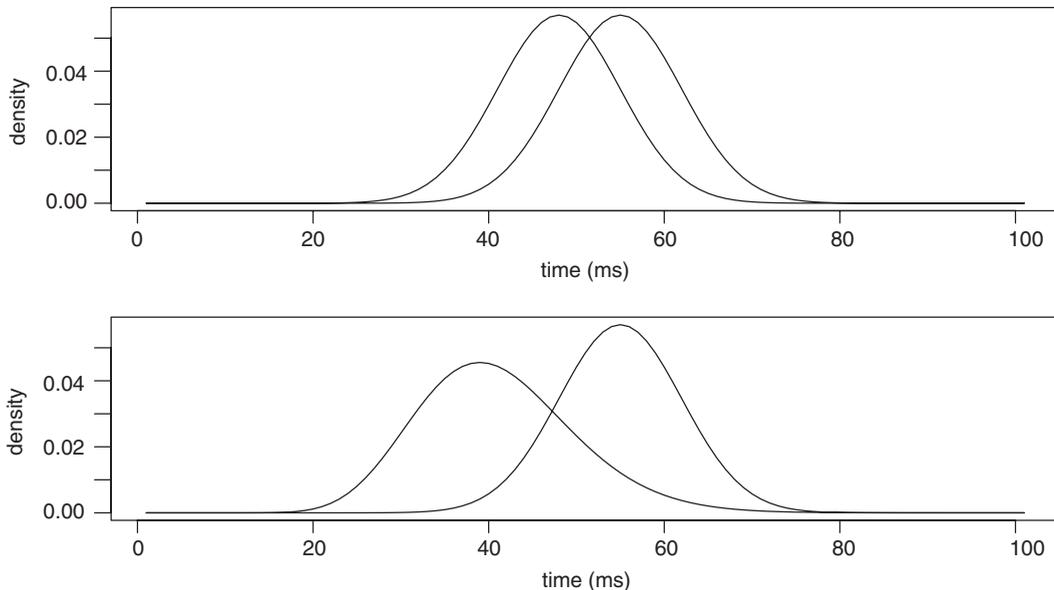

Figure 2. Curves used for the power study. Top, scenario A: Pictured Normal(47,49) and Normal(57,49) densities. Bottom, scenario B: Pictured are $\chi^2(40)$ and Normal(57,49) densities. The densities were multiplied by 1000 to produce realistic neuronal intensity functions.

Table II. Probability of rejecting H_0 for scenario A with $\alpha = 0.05$, for various sample sizes (number of trials) r.

	$r = 15$	$r = 30$	$r = 50$
Bayes factor	0.73	0.82	0.91
T_k^2	0.58	0.79	0.88

Table III. Probability of rejecting H_0 for scenario A with $\alpha = 0.01$, for various sample sizes (number of trials) r.

	$r = 15$	$r = 30$	$r = 50$
Bayes factor	0.69	0.79	0.86
T_k^2	0.43	0.76	0.82

Table IV. Probability of rejecting H_0 for scenario B with $\alpha = 0.05$, for various sample sizes (number of trials) r.

	$r = 15$	$r = 30$	$r = 50$
Bayes factor	0.86	0.98	0.99
T_k^2	0.79	0.96	0.99

Table V. Probability of rejecting H_0 for scenario B with $\alpha = 0.01$, for various sample sizes (number of trials) r.

	$r = 15$	$r = 30$	$r = 50$
Bayes factor	0.75	0.89	0.95
T_k^2	0.68	0.91	0.94

1000, which produces the realistic maximal firing rate of 57 spikes/s. We computed the power of each test when each was adjusted to have size (type I error) $\alpha = 0.05$ or $\alpha = 0.01$ for sample sizes (number of trials) $r = 15, 30, 50$. Specifically, given intensity functions $\lambda^1(t)$ and $\lambda^2(t)$, for $j = 1, 2$ we computed r replications of Poisson process data based on $\lambda^j(t)$. For each such pair of sets of r Poisson event times we performed the Bayes factor and Gaussian process (T_k^2) tests at level α. This process was repeated 1000 times to determine the probability of rejecting H_0.

The results are given in Tables II–V. They show that both of these tests have good power against realistic alternatives. For sample sizes of $r = 30$ or more the power is very high. The Bayes factor has consistently higher power than the Gaussian process test. However, for sample sizes of $r = 30$ or more the distinction becomes negligible.

We also conducted additional simulation studies, trying more complex (wiggly) alternative scenarios in an attempt to find cases for which the Gaussian process test had substantially

greater power than the Bayes factor. We were unsuccessful. In the overwhelming majority of cases, the results were close to the ones we reported here. In one case the power of the T^2 procedure was negligibly (0.02) larger than the power of the Bayes factor test, when $r = 15$.

7. DISCUSSION

We have presented two methods of testing equality of two functions non-parametrically, based on BARS. Both methods appear to work well, and they agreed closely in the data set we examined. For small samples the Bayes factor is slightly more powerful. On the other hand, the Gaussian process test has the advantage (an advantage at least to many practitioners) that it is based on familiar p-values. Its avoidance of the same-knot-set assumption might help it adapt better to certain functions. This becomes an important point when one considers the variability among many functions: our experience with multiple curve-fitting [2] suggests that same-knot-set assumption may undersmooth parts of data by imposing too many knots. However, we were unable to construct a pair of functions for which the Gaussian process test had substantially greater power than the Bayes factor.

Our emphasis has been on testing equality of two Poisson process intensity functions in the context of neurophysiology. For screening of large numbers of neurons, based on this work, we would recommend the use of the Gaussian process test with $p = 0.01$. Cases involving more delicate scientific questions, where equality of intensity functions is considered an important possibility, are different. There it would likely be preferable to rely on the Bayes factor, which provides an assessment of the probability in favour of H_0. An additional possibility would be to devise a method of computing the Bayes factor that would be applicable even if the knot sets were distinct. We leave that to future efforts.

We used the BIC to define our Bayes factor, citing previous work to justify our equation (6). Equation (4) involves a substantial summation, so that accumulation of errors is possible. We did not investigate the accuracy of our approximation. Instead, we relied on our simulation study as a strong indication that the approach is useful.

REFERENCES

1. DiMatteo I, Genovese CR, Kass RE. Bayesian curve fitting with free-knot splines. *Biometrika* 2001; **88**: 1055–1073.
2. Behseta S, Kass RE, Wallstrom G. Hierarchical models for assessing variability among functions. *Biometrika*, in press.
3. Kass RE, Ventura V, Cai C. Statistical smoothing of neuronal data. *NETWORK: Computation in Neural Systems* 2003; **14**:5–15.
4. Kass RE, Wallstrom G. Invited comment on 'Spline Adaptation in Extended Linear Models' by Mark H. Hansen and Charles Kooperberg. *Statistical Science* 2002; **17**:24–29.
5. Kaufman CG, Ventura V, Kass RE. Spline-based nonparametric regression for periodic functions and its application to directional tuning of neurons. *Statistics in Medicine*, in press.
6. Wallstrom GL, Kass RE, Miller A, Cohn JF, Fox NA. Correction of ocular artifacts in the EEG using Bayesian adaptive regression splines. In *Case Studies in Bayesian Statistics*, vol. VI. Gatsonis C, Carriquiry A, Higdon D, Kass RE, Pauler D, Verdinelli I (eds). 2002, in press.
7. Zhang X, Roeder K, Wallstrom G, Devlin B. Integration of association statistics over genomic regions using Bayesian adaptive regression splines. *Human Genomics* 2003; **1**:20–29.
8. Wallstrom GL, Kass RE, Liebner J. A portable C implementation of BARS with R and S wrappers. *Journal of Statistical Software*, accepted.
9. Fan J, Lin S. Test of significance when data are curves. *Journal of the American Statistical Association* 1998; **93**:1007–1021.

10. Neumeyer N, Dette H. Nonparametric comparison of regression curves: an empirical process approach. *Annals of Statistics* 2003; **31**:880–920.
11. Hansen MH, Kooperberg C. Spline adaptation in extended linear models. *Statistical Science* (with discussion) 2002; **17**:2–51.
12. Kakei S, Hoffman DS, Strick PL. Muscle and movement representation in the primary motor cortex. *Science* 2001; **285**:2136–2139.
13. Kass RE, Raftery AE. Bayes factors. *Journal of the American Statistical Association* 1995; **90**:773–795.
14. Denison D, Mallick B, Smith A. Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society, Series B* 1998; **60**:333–350.
15. Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 1995; **82**:711–732.
16. Liu JS, Wong WH, Kong A. Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* 1994; **81**:27–40.
17. Kass RE, Wasserman LA. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association* 1995; **90**:928–934.
18. Pauler DK. The Schwarz criterion and related methods for normal linear models. *Biometrika* 2001; **85**:13–27.
19. Matsuzaka Y, Picard N, Strick PL. Sequence learning in monkeys. *Society for Neuroscience Abstracts* 2001; **24**:174.
20. Behseta S. Bayesian multiple curve fitting in the analysis of neuronal data. *Ph.D. Thesis*, Carnegie Mellon University, 2003.
21. Behseta S, Matsuzaka Y, Picard N, Kass RE, Strick PL. Muscle-like activity of M1 neurons during multi-joint movements. *Society for Neuroscience Abstracts* 2002; 61.13.
22. Ventura V, Carta R, Kass RE, Gettner SN, Olson CR. Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics* 2002; **3**:1–23.