Sequential change-point detection for time series models: assessing the functional dynamics of neuronal networks.

Fabio Rigat*

July 16, 2007

Abstract

This paper illustrates a sequential method to detect significant parameter changes for time series models. Rather than relying on an explicit state equation, the parameters' dynamics are assessed as a change-point problem by combining Bayesian estimation with a non-parametric test of hypothesis. The Kullback-Leibler divergence between the posterior probability densities given two different sets of data is proposed as a test statistic. Markov chain Monte Carlo posterior simulation is used to approximate in general the value of the Kullback-Leibler statistic and its critical region under the null hypothesis. For exponential family models we show that the statistic has a closed form. We also report the results of a simulation study demonstrating empirically that for the Bernoulli model the power of the change-point test is not affected by the difference in the sample

*Research fellow, CRiSM, Department of Statistics, University of Warwick; f.rigat@warwick.ac.uk

sizes of the data involved. The method is applied to detecting changes of the functional connectivity of neuronal networks using in-vivo multiple spike trains recordings obtained via multi-electrode arrays.

Introduction

This paper illustrates a sequential method for detecting the occurrence of changes in the parameter values of time series models under mild assumptions about the form of their evolution. This method is closely related to state-space models and to change-point regression models. In the former (Harrison and Stevens [1976], West et al. [1985], West and Harrison [1986], West and Harrison [1997]), changes of the model parameters are governed by an explicit state equation indexed by a set of evolution coefficients. Least squares and Bayesian estimates are available in closed form for the Gaussian dynamic linear model (Kalman [1960], Harrison and Stevens [1976]), whereas estimation for non-Gaussian and non-linear dynamic models is currently carried out using sequential importance resampling methods (Gordon et al. [1993], Doucet et al. [2000], Doucet et al. [2001], Gilks and Berzuini [2001], Crisan and Doucet [2002], Del Moral et al. [2006]).

Change-point detection for regression models was introduced by Quandt [1958] and Quandt [1960], who developed a likelihood ratio test detecting a switch between two regimes for linear regression coefficients. Brown et al. [1975] proposed a cumsum test based on a definition of recursive residuals for linear regression models. Muller [1992], Loader [1996] and Mira and Petrone [1996] develop methods for estimating the occurrence of change-points in nonparametric regression models. Carlin et al. [1992] adopt a hierarchical Bayesian approach using Markov chain Monte Carlo posterior

simulation to detect a single change-point for generalised linear regression models. Bélisle et al. [1998] considered a multivariate single change-point problem for neuron spike trains using integer-valued autoregressive time series models. Change-point estimation for partially observed non-linear dynamic systems via particle filters is illustrated in Vaswami [2004]. Applications to tracking adaptive neural responses in the auditory cortex are illustrated in Jain et al. [2007].

The distinctive feature of the approach proposed in this paper is that it does not rely on any parametric formulation neither of the evolution of the model coefficients nor of the number of their changes. Significant departures from a null hypothesis of no change result in atipical values of a Kullback-Leibler divergence (KL; Kullback and Leibler [1951], Kullback [1997]) between the posterior distributions of the model parameters given different sets of data. This non-parametric Kullback-Leibler test is integrated in a sequential procedure alternating Bayesian estimation with change-point testing. We apply this methodology to assess the functional dynamics of networks of neurons using in-vivo experimental multiple spike trains recordings (Buzsáki [2004]).

This paper is organised as follows. Section 1 illustrates the Kullback-Leibler statistic, a Markov chain Monte Carlo strategy to estimate its value and its critical region and the sequential algorithm integrating data fitting and non-parametric change-point testing. Section 2 gives a closed form expression of the Kullback-Leibler statistic for exponential family models. A simulation study is presented investigating empirically the power of the test for comparing datasets having different samples size using the Bernoulli model. Section 3 briefly summarises the state-of-the-art in dynamic modelling of neuronal networks. A Bayesian model for binary networks is in-

troduced along the lines of Rigat et al. [2006] and a set of in-vivo multiple spike trains recordings from sheep temporal cortex is analysed.

1 Sequential time series modelling and Kullback-Leibler change-point testing

For $i = 1, ..., N$ let Y_i be a $(K \times n_i)$ dimensional time series of random variables $Y_{i,k,t}$ with $k = 1, ..., K$, $t \in [t_{i,1} < t_{i,2} < ... < t_{i,n_i}]$ and $t_{i,n_i} < t_{i+1,1}$. When $n_i = 1$, for all values of i we entertain only one K-dimensional time series $(Y_1, ..., Y_N)$. Otherwise, the distinction between the N sets of data is relevant as we admit the possibility of time gaps occurring between pairs of data sets. We assume that the initial conditions y_0 and the form of the sampling distribution $P(Y_i \mid \theta_i, y_0, y_1, ..., y_{i-1})$ are fixed over time. Dependence on the past data $(y_0, y_1, ..., y_{i-1})$ is thought of as corresponding to an autoregressive model and the parameters θ_i represent the autoregressive coefficients for Y_i. Here we do not consider the case where the distribution of Y_i also depends on other covariates, although our methodology can be extended to such a case.

When $i < N$, upon observing a new realisation $Y_{i+1} = y_{i+1}$ we wish to detect whether Bayesian updating of the joint posterior distribution of the model parameters is supported by the data. Our interest focuses on this particular null hypothesis because it represents the optimal solution for the sequential learning problem in the sense of Bellman (Bellman [1957], Zellner [2002]). Under this definition of change-point, if no changes are detected prior to observing y_{i+1} inferences for θ_i are reflected by its conditional

posterior density

$$f(\theta_i \mid y_0, ..., y_i) \propto f(\theta_i) \prod_{j=1}^{i} P(y_j \mid \theta_i, y_0, ..., y_{j-1}). \tag{1}$$

Conditionally on $(y_0, ..., y_i)$ and (1), upon observing y_{i+1} the null hypothesis of no change is

$$H_0 : y_{i+1} \sim P(Y_{i+1} \mid \theta_i, y_0, ..., y_i),$$

and its alternative is

$$H_1 : y_{i+1} \sim P(Y_{i+1} \mid \theta_{i+1}, y_0, ..., y_i),$$

where the probability density of the new parameter θ_{i+1} conditional on $(y_0, ..., y_i)$ is not equal to the right-hand side of (1). Among the many possible specifications of the alternative hypothesis, in this work we consider

$$f(\theta_{i+1} \mid y_0, ..., y_{i+1}) \propto f(\theta_{i+1}) P(Y_{i+1} \mid \theta_{i+1}, y_0, ..., y_i). \tag{2}$$

Under (2), when Bayesian sequential updating is not supported by the data the learning process is reset at the prior, which is taken as fixed for all $i = 1, ..., N$.

Conditionally on $(y_0, ..., y_i)$, the evidence in the data y_{i+1} against H_0 can be thought of as measured by an appropriate notion of discrepancy between the joint posterior densities $f(\theta_i \mid y_0, ..., y_i)$ and $f(\theta_i \mid y_0, ..., y_{i+1})$ derived under the null hypothesis. When y_{i+1} supports H_0, their discrepancy should assume a range of values reflecting the typical concentration of the latter posterior due to the accrual of new data. When y_{i+1} suggests that a departure from H_0 has taken place, the value of their discrepancy should become more extreme. In the next Section we use the Kullback-Leibler divergence to construct such a measure of discrepancy.

1.1 A Kullback-Leibler test statistic

The divergence proposed by Kullback and Leibler [1951] has many applications in statistics, among whom density estimation (Hall [1987], Hastie [1987]), model selection (Akaike [1978], Akaike [1981], Carota et al. [1996] Goutis and Robert [1998]), experimental design (Lindley [1956], Stone [1959]) and the construction of prior distributions (Bernardo [1979]). Its geometric properties have been explored by Critchley et al. [1994]. Here we illustrate a form of the Kullback-Leibler divergence suitable for measuring the evidence against the null hypothesis of no change-point when the data is accrued sequentially over time. This statistic is:

$$KL(y_0, ..., y_{i+1}) = \int_{\Theta_i} \log\left(\frac{f(\theta_i|y_0,...,y_i)}{f(\theta_i|y_0,...,y_{i+1})}\right) f(\theta_i \mid y_0, ..., y_i) d\theta_i, \quad (3)$$

$$= \log(E(P(y_{i+1} \mid \theta_i, y_0, ..., y_i))) - E(\log(P(y_{i+1} \mid \theta_i, y_0, ..., y_i))), \quad (4)$$

where the expectations in (4) are taken with respect to the posterior density $f(\theta_i \mid y_0, ..., y_i)$. The right-hand side of (3) exists finite when the likelihood is bounded away from zero for all values of θ_i and the posterior $f(\theta_i \mid y_0, ..., y_i)$ is proper. In such a case (3) is a non-negative convex function measuring the discrepancy between the posterior densities $f(\theta_i \mid y_0, ..., y_i)$ and $f(\theta_i \mid y_0, ..., y_{i+1})$. The discrepancy (3) is null if and only if the likelihood function $P(y_{i+1} \mid \theta_i, y_0, ..., y_i)$ does not vary with θ_i over the range of values associated to a non-negligible posterior density $f(\theta_i \mid y_0, ..., y_i)$. Prior to observing $Y_{i+1} = y_{i+1}$, (3) is a random variable $KL(y_0, ..., y_i, Y_{i+1})$ which distribution under H_0 depends on that of the future data Y_{i+1} via model $P(Y_{i+1} \mid \theta_i, y_0, ..., y_i)$.

Let $\alpha \in (0, 1)$ be a fixed type-1 error probability, that is the probability of rejecting H_0 when it is true, for the change-point test using the

statistic (3). Let $(l_{i,\alpha}, u_{i,\alpha})$ be the critical cutoff values for the random variable $KL(y_0, , ..., y_i, Y_{i+1})$ under H_0 prior to observing $Y_{i+1} = y_{i+1}$. When $KL(y_0, ..., y_{i+1}) < l_{i,\alpha}$, under H_0 the joint posterior densities $f(\theta_i \mid y_0, ..., y_i)$ and $f(\theta_{i+1} \mid y_0, ..., y_{i+1})$ are too similar whereas if $KL(y_0, ..., y_{i+1}) > u_{i,\alpha}$ the two joint posterior densities are too different to be consistent with H_0.

1.2 Computation of the test statistic

Since (2) is a non-linear functional of the ratio of two posterior densities, in general neither its value nor its critical cutoffs under H_0 can be derived in closed form. However, using the representation (4) suggests that its Monte Carlo approximation,

$$KL(y_0, ..., y_{i+1}) \approx \log\left(\frac{\sum_{m=1}^{M} P(y_{i+1}|\theta_i^m, y_0, ..., y_i)}{M}\right) - \frac{\sum_{m=1}^{M} \log(P(y_{i+1}|\theta_i^m, y_0, ..., y_i))}{M},$$

can be computed using the Gibbs sampler (Gelfand and Smith [1990], Tierney [1994], Smith and Roberts [1993]) draws $\{\theta_i^m\}_{m=1}^M$. Also a Monte Carlo approximation of the critical region of (3) for varying values of Y_{i+1} under H_0 can be constructed using the same Gibbs sampler draws as follows:

i) for each draw θ_i^m generate a realisation y_{i+1}^m using the joint sampling probability $P(Y_{i+1} \mid \theta_i^m, y_0, ..., y_i)$;

ii) compute the Monte Carlo approximation of the statistic $KL(y_0, ..., y_{i+1}^m)$.

The empirical distribution of the sequence $\{KL(y_0, ..., y_{i+1}^m)\}_{m=1}^M$ approximates that of (4) under the null hypothesis. For instance, for any given type-1 error probability α, the empirical $(\frac{\alpha}{2}, 1-\frac{\alpha}{2})$th percentiles of $\{KL(y_0, ..., y_{i+1}^m)\}_{m=1}^M$ approximate the values of the equal tails critical cutoffs $(l_{i,\alpha}, u_{i,\alpha})$.

1.3 Sequential estimation and change-point testing algorithm

Given the initial conditions y_0 and a sampling model $P(Y_{i+1} \mid \theta_i, y_0, ..., y_i)$, the dynamics of the model parameters can be assessed by integrating an estimation step with the change-point test using the Kullback-Leibler statistic (3). We illustrate the sequential Bayesian fitting and testing algorithm starting from the first sample y_1:

i) upon observing y_1, derive the posterior $f(\theta_1 \mid y_0, y_1)$;

ii) having observed y_2, compute $KL(y_0, y_1, y_2)$ and its critical cutoffs under H_0, $(l_{1,\alpha}, u_{1,\alpha})$ given the type-1 error probability α, as described in Section 1.2.

iii.1) if $KL(y_0, y_1, y_2) \in (l_{1,\alpha}, u_{1,\alpha})$, return to i) and derive the updated posterior density

$$f(\theta_1 \mid y_0, y_1, y_2) \propto f(\theta_1 \mid y_0, y_1) P(y_2 \mid \theta_1, y_0, y_1),$$

iii.2) otherwise, return to i) and derive the posterior density for the new parameter

$$f(\theta_2 \mid y_0, y_1, y_2) \propto f(\theta_2) P(y_2 \mid \theta_2, y_0, y_1).$$

iv) return to step ii).

2 Exponential family models

In this Section we assume that each set of data Y_i is a $1 \times n_i$ dimensional sample of conditionally independent observations with joint density (Diaconis

and Ylvisaker [1979])

$$P(Y_i \mid \theta_i) = \prod_{j=1}^{n_i} a(Y_{i,j}) e^{Y_{i,j}\theta_i - b(\theta_i)}, \tag{5}$$

where θ_i is the canonical parameter. The properties of the Kullback-Leibler divergence within the exponential family have been explored by McCulloch [1988]. Here we show that also the Kullback-Leibler (3) has a closed form for exponential familiy models. Therefore, in this case stochastic simulation is necessary to approximate the value of the critical cutoffs $(l_{i,\alpha}, u_{i,\alpha})$ under H_0 but the not value of the test statistic.

Diaconis and Ylvisaker [1979] show that each element of Y_i has moments

$$E(Y_{i,j} \mid \theta_i) = \frac{\partial b(\theta_i)}{\partial \theta_i}, V(Y_{i,j} \mid \theta_i) = \frac{\partial^2 b(\theta_i)}{\partial \theta_i' \partial \theta_i}.$$

Using the conjugate prior

$$f(\theta_i \mid n_0, y_0) = c(n_0, S_0) e^{S_0\theta_i - n_0 b(\theta_i)},$$

where $S_0 = n_0 y_0$ for scalars n_0 and y_0, the posterior for θ_i given the past samples $(y_1 ..., y_i)$ is

$$f(\theta_i \mid n(i), y(i)) = c(n(i), S(i)) e^{n(i)\left(\frac{S(i)}{n(i)}\theta_i - b(\theta_i)\right)} \tag{6}$$

where $n(i) = \sum_{j=0}^{i} n_j$, $S(i) = \sum_{j=0}^{i} n_j \bar{y}_j$ and $\bar{y}_j$ is the arithmetic mean of sample y_j. Using the results of Gutiérrez-Peña [1997], the posterior mean and variance of θ_i under (5) are

$$E(\theta_i \mid n(i), S(i)) = \frac{\partial H(n(i), S(i))}{\partial S(i)}, V(\theta_i \mid n(i), S(i)) = \frac{\partial H(n(i), S(i))}{\partial S(i)^2},$$

where $H(n(i), S(i)) = -\log\left(c(n(i), S(i))\right)$, and the posterior mean and variance of the function $b(\theta_i)$ are

$$E(b(\theta_i) \mid n(i), S(i)) = \frac{\partial H(n(i), S(i))}{\partial n(i)}, V(b(\theta_i) \mid n(i), S(i)) = \frac{\partial H(n(i), S(i))}{\partial n(i)^2}.$$

When the posterior for θ_i has form (6), given the data up to and including y_{i+1} the Kullback-Leibler statistic (3) has form

$$\begin{aligned} KL(y_0, ..., y_{i+1}) &= \log\left(\frac{c(n(i), S(i))}{c(n(i+1), S(i+1))}\right) - S_{i+1}\frac{\partial H(n(i), S(i))}{\partial S(i)} + \\ &+ n_{i+1}\frac{\partial H(n(i), S(i))}{\partial n(i)}. \end{aligned} \quad (7)$$

Proof: by letting the posterior densities $f(\theta_i \mid n_0, n_i, S_0, S_i)$ and $f(\theta_i \mid n_0, n_i, n_{i+1}, S_0, S_i, S_{i+1})$ have form (6), the Kullback-Leibler divergence (3) becomes

$$KL(y_0, ..., y_{i+1}) = \log\left(\frac{c(n(i), S(i))}{c(n(i+1), S(i+1))}\right) - S_{i+1}E(\theta_i) + n_{i+1}E(b(\theta_i)). \quad (8)$$

For exponential family models, the expectations of θ_i and $b(\theta_i)$ with respect to $f(\theta_i \mid y_0, ..., y_i)$ are given in Gutiérrez-Peña [1997], as reported above. By substituting these expressions in (8), equation (7) obtains. $\diamond$

Example 2.1: when Y_i is a scalar Gaussian random variable with mean μ_i and precision λ_i, its distribution can be written in the Diaconis canonical form using the two-dimensional statistic

$$Y_i^* = [Y_i, Y_i^2],$$

and the two-dimensional canonical parameter

$$\theta_i = [\theta_{1,i}, \theta_{2,i}] = \left[\lambda_i\mu_i, -\frac{\lambda_i}{2}\right].$$

with

$$\begin{aligned} a(Y_i^*) &= (2\pi)^{-\frac{1}{2}}, \\ b(\theta_i) = &= -\frac{1}{2}\log(-2\theta_{2,i}) - \frac{\theta_{1,i}^2}{\theta_{2,i}}. \end{aligned}$$

The conjugate prior for (μ_i, λ_i) is the Normal-Gamma $N(\mu_i \mid \gamma, \lambda_i(2\alpha - 1))Ga(\lambda_i \mid \alpha, \beta)$ with coefficients $\alpha > 0.5, \beta > 0, \gamma \in \mathcal{R}$ and normalising

constant (Bernardo and Smith [2007])

$$c(n_0, S_0) = \left(\frac{2\pi}{n_0}\right)^{\frac{1}{2}} \frac{\frac{1}{2} S_{2,0}^{\frac{S_{1,0}}{2}}}{\Gamma\left(\frac{n_0+1}{2}\right)},$$

where $n_0 = 2\alpha - 1$, $y_0^* = [y_{1,0}^*, y_{2,0}^*] = [\gamma, \frac{2\beta}{2\alpha-1} + \gamma^2]$, $S_{1,0} = n_0 y_{1,0}^*$ and $S_{2,0} = n_0 y_{2,0}^*$. Upon observing the realisations $(y_1, ..., y_i)$, the normalising constant of the conjugate posterior (5) is

$$c(n(i), S(i)) = \left(\frac{2\pi}{n(i)}\right)^{\frac{1}{2}} \frac{\frac{1}{2} S(2,i)^{\frac{S(1,i)}{2}}}{\Gamma\left(\frac{n(i)+1}{2}\right)},$$

where $n(i) = n_0 + i$, $S(1,i) = S_{1,0} + \sum_{j=1}^{i} y_j$ and $S(2,i) = S_{2,0} + \sum_{j=1}^{i} y_j^2$. When also y_{i+1} is observed, using (7) the Kullback-Leibler statistics can be written as

$$KL(y_0, ..., y_{i+1}) = \log\left(\frac{\Gamma\left(\frac{n(i+1)+1}{2}\right)}{\Gamma\left(\frac{n(i)+1}{2}\right)}\right) + \frac{1}{2}\log\left(\frac{n(i+1)}{n(i)}\right) + \log\left(\frac{S(2,i)^{\frac{S(1,i)}{2}}}{S(2,i+1)^{\frac{S(1,i+1)}{2}}}\right) -$$
$$-\frac{y_{i+1}}{2}\log\left(\frac{S(2,i)}{2}\right) - y_{i+1}^2 \frac{S(1,i)}{S(2,i)} + \frac{1}{2n(i)} + \Gamma\left(\frac{n(i)+1}{2}\right) \frac{\partial\Gamma\left(\frac{n(i)+1}{2}\right)}{\partial n(i)}.$$

Example 2.2: let Y_i be a sample of size n_i of conditionally independent Bernoulli random variables with success probability π_i. The canonical representation of the Bernoulli probability mass function obtaines by letting $\theta_i = \log\left(\frac{\pi_i}{1-\pi_i}\right)$, $b(\theta_i) = \log\left(1 + e^{\theta_i}\right)$ and $a(Y_i) = 1$. The conjugate prior for π_i is $Beta(S_0, m_0)$ where $m_0 = n_0 - S_0$. Upon observing $(y_1, .., y_i)$ the conjugate posterior is $Beta(S(i), m(i))$, where $S(i) = \sum_{j=0}^{i} S_j$, $n(i) = \sum_{j=0}^{i} n_j$ and $m(i) = n(i) - S(i)$. When also y_{i+1} is observed, (7) has form

$$KL(y_0, ..., y_{i+1}) = \log\left(\frac{\prod_{k=1}^{n_i}(n(i)+k)\prod_{w=1}^{n_i-S_i}(n(i)-S(i)+w)}{\prod_{j=1}^{S_i}(S(i)+j)}\right) -$$
$$-S_{i+1}\frac{\Gamma(S(i))}{\Gamma(n(i)-S(i))}\frac{\partial\frac{\Gamma(n(i)-S(i))}{\Gamma(S(i))}}{\partial S(i)} + n_{i+1}\frac{\frac{\partial(\Gamma(n(i))\Gamma(n(i)-S(i)))}{\partial n(i)}}{\Gamma(n(i))\Gamma(n(i)-S(i))}.$$

Example 2.3: let Y_i represent the random number of events of a given kind observed within a time interval $(t_{i,1}, t_{i,n_i}]$. For this example we assume that the length of the latter is identical for all samples $i = 1, ..., N$. Let the random event times be generated by a homogeneous Poisson process with intensity λ_i, so that the distribution of Y_i is Poisson with parameter $\lambda_i^* = \lambda_i(t_{i,n_i} - t_{i,1})$. The canonical form of the Poisson distribution has parameter $\theta_i = \log(\lambda_i^*)$ and functions $a(Y_i) = \frac{1}{Y_i!}, b(\theta_i) = e^{\theta_i}$. The conjugate prior for λ_i^* is $Gamma(S_0, n_0)$ with mean y_0 and variance $\frac{y_0}{n_0}$. Upon observing $(y_1, .., y_i)$ the conjugate posterior for λ_i^* is $Gamma(S(i), n(i))$ with $S(i) = S_0 + \sum_{j=1}^{i} y_j$, $n(i) = n_0 + i$. When also y_{i+1} is observed, using (7) the Kullback-Leibler statistic has form

$$KL(y_0, ..., y_{i+1}) = \log\left(\frac{S(i)n(i)^{S(i)}}{n(i+1)^{S(i+1)}}\right) + y_{i+1}\left(\log(n(i)) - \frac{\frac{\partial\Gamma(S(i))}{\partial S(i)}}{\Gamma(S(i))}\right) - \frac{S(i)}{n(i)}.$$

2.1 Power and sample sizes

In general, the time series $\{Y_i\}_{i=1}^N$ may have substantially different lenghts $\{n_i\}_{i=1}^N$. To demonstrate the applicability of the KL-based change-point test in this case, this Section reports the results of a simulation study evaluating empirically the power of the test for varying sample sizes (n_1, n_2). We use the Bernoulli model presented in example 2.2 for its similarity with the neuronal network model developed in Section 3.

One hundred thousand independent simulations were run. For each simulation, two sample sizes (n_1, n_2) were independently generated using the discrete uniform distribution on the integers $(1, ..., 100)$. A success probabiliy π was also independently generated for each simulation using a uniform distribution on $(0, 1)$. Conditionally on (n_1, n_2, π), two independent Bernoulli samples were generated $Y_1 \sim Ber(\pi, n_1)$ and $Y_2 \sim Ber(\pi, n_2)$.

For each simulation, a sample of size 5000 was generated from the posterior $Beta(1 + \sum_{j=1}^{n_1} Y_{1,j}, 1 + n_1 - \sum_{j=1}^{n_1} Y_{1,j})$ to compute the Monte Carlo approximations for the Kullback-Leibler statistic (3) under H_0.

We fixed the type-1 error probability of the test to $\alpha = 0.2$. Out of the one hundred thousand simulations the null hypothesis was accepted 79743 times, which closely approximates the true value of the power $1 - \alpha$. The top plot in Figure 1 shows the exact value of the Kullback-Leibler statistic and of the end-points of its 80% equal tails frequency intervals for the first fifty simulations. The Kullback-Leibler is represented with a dot when it lies within its critical region and as a plus sign when it violates H_0. The plot on the bottom row Figure 1 shows the statistic $Z = \log\left(\frac{n_1^*}{n_2^*}\right)$, where $n_1^* = n_1 - 1$ and $n_2^* = n_2 - 1$, for the same fifty simulations. The value of the Kullback-Leibler statistic and that of its critical cutoffs are large for simulations where n_1 is small compared to n_2 and vice versa. In the former case, large cutoff values reflect the fact that under H_0 a larger sample Y_2 is expected to provide more information about π with respect to the small sample Y_1. In the latter case, small cutoff values reflect the opposite expectation. Since the random variables $\frac{n_1-1}{99}$ and $\frac{n_2-1}{99}$ are independent and approximately uniform on $(0, 1)$, the distribution of $Z = \log\left(\frac{n_1^*}{n_2^*}\right)$ is approximately standard double exponential with probability density $f(Z) = \frac{e^{-|Z|}}{2}$ for $Z \in \mathcal{R}$. If the power of the Kullback-Leibler change-point test is not affected by the values of (n_1, n_2, π), the distribution of Z for the group of simulations where H_0 is accepted should also be standard double exponential. Figure 2 represents with a solid line the empirical cumulative distribution function of Z for the 79743 simulations where H_0 was accepted. The two dashed lines in the same Figure represent the end-points of the point-wise 99% sample frequency intervals for the cumulative distribution

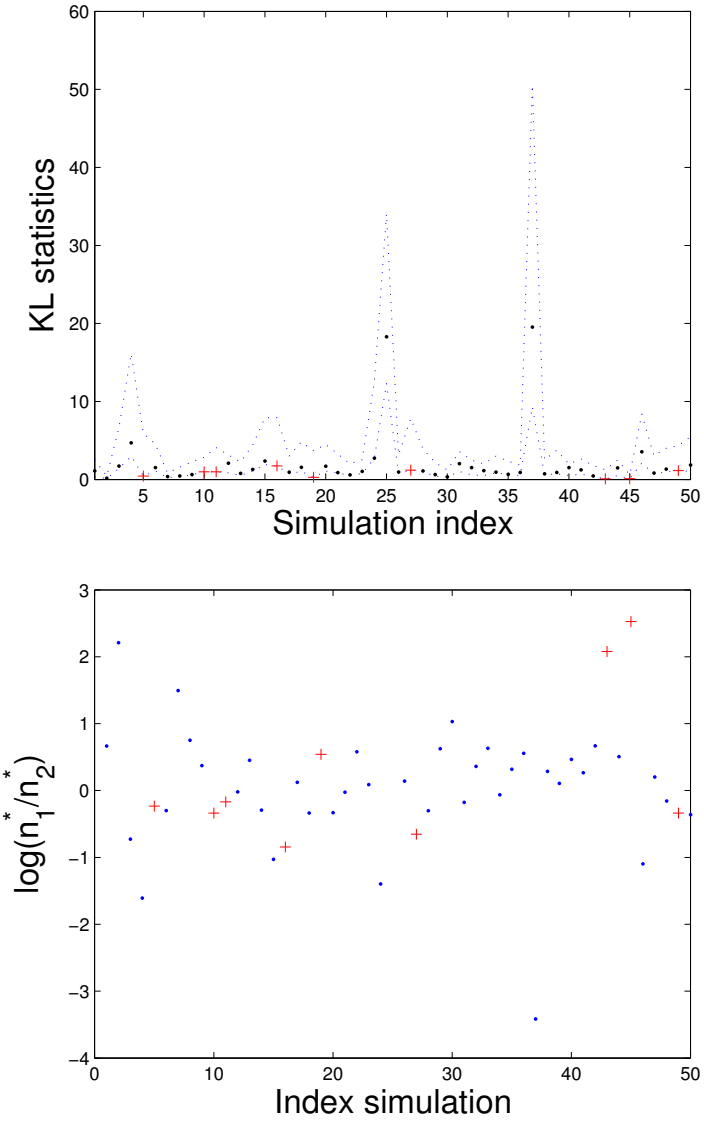

Figure 1: on top, values of the Kullback-Leibler statistic and of the end-points of its 80% probability intervals for the first fifty simulations. At the bottom, values of the statistic $Z = \log\left(\frac{n_1-1}{n_2-1}\right)$ for the same simulations. Dots indicate simulations where H_0 was accepted and plus signs mark the simulations where H_0 was rejected. The two plots show that under H_0 the value of the discrepancy between $f(\pi_1 \mid y_1)$ and $f(\pi_1 \mid y_1, y_2)$ and those of its critical cutoffs adapt to the sample sizes (n_1, n_2).

function of a standard double exponential random variable. The Figure suggests that for the Bernoulli model the power of the change-point test is not significantly affected by different sample sizes (n_1, n_2).

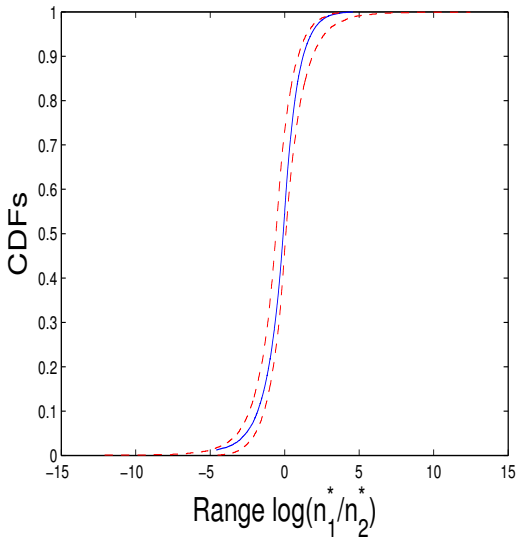

Figure 2: the solid line represents the empirical cumulative distribution function (CDF) of the random variable $Z = \log\left(\frac{n_1^*}{n_2^*}\right)$ for the 79743 simulations where H_0 was accepted. The dashed lines represent the end-points of the point-wise 99% sample frequency intervals for the CDF of a standard double exponential random variable. Acceptance of the null hypothesis H_0 did not cause a significant departure of the distribution of Z from that of a standard double exponential distribution, suggesting that the power of the change-point test is not significantly affected by different sample sizes (n_1, n_2).

3 Detecting neuronal networks functional dynamics

Introductions to the neuronal physiology and to neuronal modelling are presented in Fienberg [1974] and Brillinger [1988]. Recent surveys of the state-of-the-art in multiple spike trains modelling can be found in Iyengar [2001], Brown et al. [2004], Kass et al. [2005], Okatan et al. [2005], Rao [2005] and Rigat et al. [2006]. Latham et al. [2000] illustrate the physiology of spontaneous firing in networks of mammalian neurons and several modelling strategies. Dynamic point process neuronal models based on state-space representations have been proposed by Eden et al. [2004], Truccolo et al. [2005], Brown and Barbieri [2006] and Srinivansan et al. [2006]. State-space point process models relating the neural activity with a learning process are presented in Smith et al. [2004] and Smith et al. [2006].

During the experiment considered in this Section a sheep is shown first a dark screen and then two sheep images side by side. The sheep is then rewarded if it identifies one of the two images as part of a set of "familiar faces" by pressing a button. The available data are the multi-electrode array (MEA) recordings of 59 neurons in the sheep temporal cortex, the experimental phase (pre-trial, i.e. dark screen, or during-trial, i.e. images shown) and whether the familiar images were recognised for each of the during-trial experiments. Further details about the experiment are given in Kendrick et al. [2001].

In this Section we aim at detecting changes in the functional connectivity of the 14 highest firing neurons during the first two experimental phases. Changes occurring before the onset of the visual stimulus can be imputed to spontaneous activities of the sheep such as head movements. If a change

is detected immediately after the stimulus onset, it is most likely related to the processing of the visual input. If a change is detected well after the onset of the stimulus, then it is explained by the motor activity involved in the button pressing. Although the possible biological mechanisms tuning the neuronal functional connectivity are well understood, the expected magnitude and smoothness of its changes are unknown. In such a context the MEA recordings can be conveniently analysed using the sequential method illustrated in Section 1 because it does not involve any explicit parametrization of the neuronal dynamics.

3.1 A binary network model

In what follows $\{Y_i\}_{i=1}^N$ represents a sequence of N binary matrices of dimensions $K \times n_i$ with elements $Y_{i,k,t_{i,j(i)}} = 1$ if neuron k spikes at time $t_{i,j(i)}$ during trial i and $Y_{i,k,t_{i,j(i)}} = 0$ otherwise. We model the joint sampling distribution of the multiple spike data for trial i, Y_i, as a Bernoulli process with renewal as in Rigat et al. [2006]. The differences between the neuronal network model adopted in this paper and that of Rigat et al. [2006] are the absence of a multi-level hierarchical prior and the adoption of a different autoregressive structure. Under this model the joint probability of a given realisation y_i is

$$P(Y_i = y_i \mid \pi_i) \quad = \quad \prod_{t=t_{i,1}}^{t_{i,n_i}} \prod_{k=1}^{K} \pi_{i,k,t}^{y_{i,k,t}} (1 - \pi_{i,k,t})^{1-y_{i,k,t}}. \tag{9}$$

For model (9) to be biologically interpretable, the spiking probability of neuron k at time $t_{i,j(i)}$ during trial i, $\pi_{i,k,t_{i,j(i)}}$, is defined as a one-to-one non-decreasing mapping of a real-valued voltage function $v_{i,k,t_{i,j(i)}}$ onto the

interval $(0, 1)$. In this work we adopt the logistic mapping

$$\pi_{i,k,t_{i,j(i)}} = \frac{e^{v_{i,k,t_{i,j(i)}}}}{1 + e^{v_{i,k,t_{i,j(i)}}}}.$$

Let $\tau_{i,k,t_{i,j(i)}}$ be the last spiking time of neuron k prior to time $t_{i,j(i)}$ during trial i, that is

$$\tau_{i,k,t_{i,j(i)}} = \begin{cases} 1 & \text{if } \sum_{\tau=1}^{t_{i,j(i)}} Y_{i,k,\tau} = 0 \text{ or } t_{i,j(i)} = 1, \\ \max\{1 \leq \tau < t_{i,j(i)} : Y_{i,k,\tau} = 1\} & \text{otherwise.} \end{cases}$$

The voltage function is modelled as

$$v_{i,k,t_{i,j(i)}} = \eta_{i,k} + \sum_{l=1}^{K} \beta_{i,k,l} \sum_{w=\tau_{i,k,t_{i,j(i)}}}^{t_{i,j(i)}-1} y_{i,l,w} e^{-(t-w)}. \tag{10}$$

The neuron-specific coefficients $\eta_{i,k}$ reflect the baseline firing rate of neuron k during trial i. The coefficient $\beta_{i,k,l}$ represents the strenght of the functional relationship from neuron l to neuron k during trial i. When $\beta_{i,k,l}$ is positive, the spiking activity of neuron l promotes that of neuron k whereas when it is negative firing of l inhibits that of k. The last term in the voltage equation, $\sum_{w=\tau_{i,k,t_{i,j(i)}}}^{t_{i,j(i)}-1} y_{i,l,w} e^{-(t-w)}$, defines the auto-regressive structure of the network model. The effect of the spikes produced by neuron l on the voltage function of neuron i from its last spike decrease exponentially over time, mimicking the occurrence of leakage currents across the neuronal membrane (Plesser and Gerstner [2000]).

3.2 Analysis of sheep multiple spike trains

The MEA recordings from the first two experimental phases are binned in contiguous windows of 300 milliseconds each. For each time window we use the Gibbs sampler to compute approximate posterior inferences for

the $K(K+1)$ parameters $\theta_i = (\eta_{i,1}, ..., \eta_{i,K}, \beta_{i,1,1}, ..., \beta_{i,K,K})$ of model (9). We use a component-wise update with independent Gaussian random-walk proposals for twenty-five thousand iterations per time window. The prior for the connectivity parameters of all experiments is Gaussian with mean zero and variance 10. The prior for each intercept η is Gaussian with mean -5 and standard deviation 1.0. The corresponding prior predictive mean firing rate for each neuron has 95% sample frequency interval $(0.004, 0.02)$, which is a biologically plausible range. Given a set of posterior estimates for θ_i conditionally on $(y_0, ..., y_i)$, we use the Kullback-Leibler statistic (3) to test whether a significant change occurred in any of the parameters during the ith+1 time window.

Each dot in the top plot of Figure 3 represents the number of active neurons for each millisecond during the two contiguous experimental phases. A vertical dotted line separates the two phases. Two bursts of activity can be noted at around 2300 and 6300 milliseconds, the first occurring during the pre-trial phase and the second occurring late during the trial. The plot at the bottom of Figure 3 shows as dashed lines the estimated critical cutoffs for the Kullback-Leibler statistic for the type-1 error $\alpha = 0.05$. The values of the test statistic for each time window are reported as a dot for the first experimental phase and as a plus sign for the second phase. Two change-points are dectected, roughly matching the two bursts of spiking activity. Since the first change-point occurs prior to the onset of the visual stimulus and the second occurs towards the end of the second experimental phase, both changes are most likely related to the sheep motor activity rather than to the processing of the visual input. Figure 4 shows the posterior estimates of the baselines firing coefficients of neurons 36 and 50. The former estimates are significantly decreased by the occurrence of the second change-point

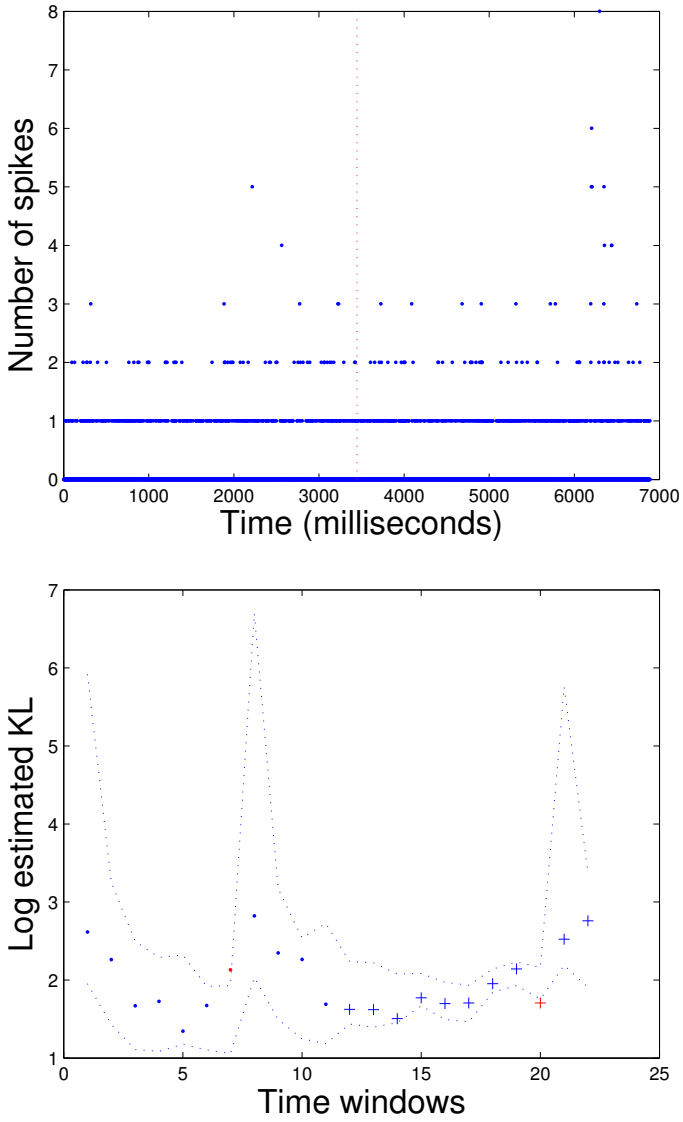

Figure 3: on top, number of spiking neurons for each millisecond during the two experimental phases. At the bottom, estimated KL statistic and its critical cutoffs for each time window. The latter represent contiguous intervals of length 300 milliseconds. Bursts of spiking activity produce two peaks in the number of active neurons on the left-hand side of the figure. The KL statistic detects two change-points in correspondence to these two bursts.

whereas those of the latter decrease at the first change-point. A change of a parameter is considered significant if its estimated 95% highest posterior density interval prior to the occurrence of a change-point does not overlap with the same estimate after the change. None of the neuronal baseline firing coefficients is significantly affected by the occurrence of both change-points. Significant changes of 33 and 27 network coefficients characterise respectively the first and the second change-point. The posterior estimates of the only three network coefficients affected by both change-points are shown in Figure 5. In both cases the pair-wise functional connectivities defined by these three coefficients become excitatory, which partially explains the increased network activity taking place in correspondence to the change-points.

4 Discussion

This paper proposes a method for detecting significant changes in the parameters of time series models which combines Bayesian parametric modelling with a sequential non-parametric test based on the Kullback-Leibler statistic (3). Unlike state-space models, this approach does not involve an explicit parametrization of the dynamics of the model parameters. The difference between this work and the Kullback-Leibler divergences constructed for model selection is that instead of testing which of two competing model formulations best predicts one given set of data, we focus on testing whether successive sets of data accrued over time are adequately explained by a common set of parameters under the same model formulation.

Although inferences for the model parameters are derived using Bayes theorem, the distribution of the change-point statistic under the null hy-

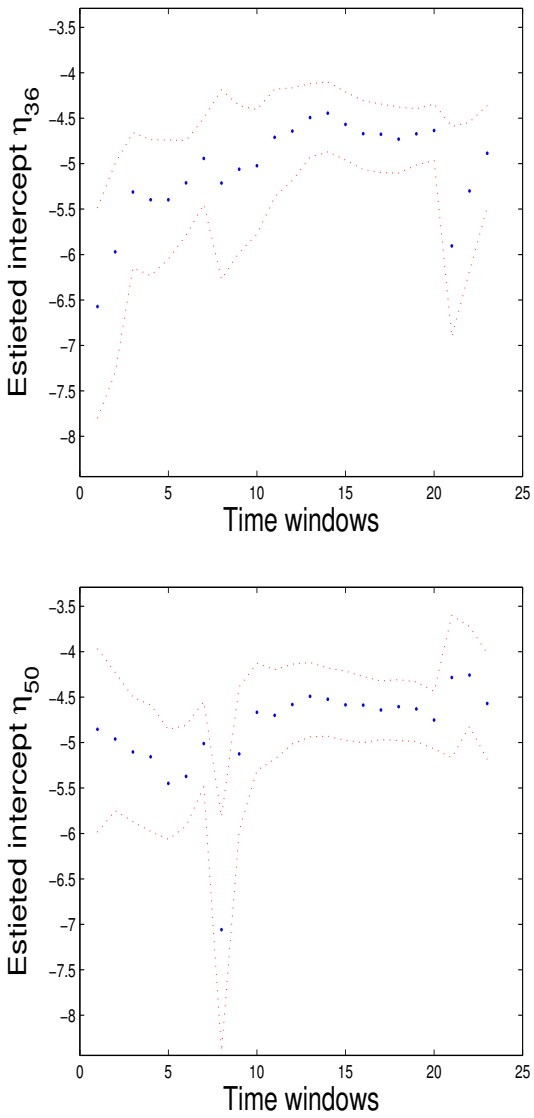

Figure 4: posterior estimates of the baselines firing coefficients for neurons 36 (top) and 50 (bottom). The former estimates decrease at the second change-point whereas the latter decrease in correspondence to the first change-point. None of the neuronal baseline firing coefficients explain the occurrence of both change-points.

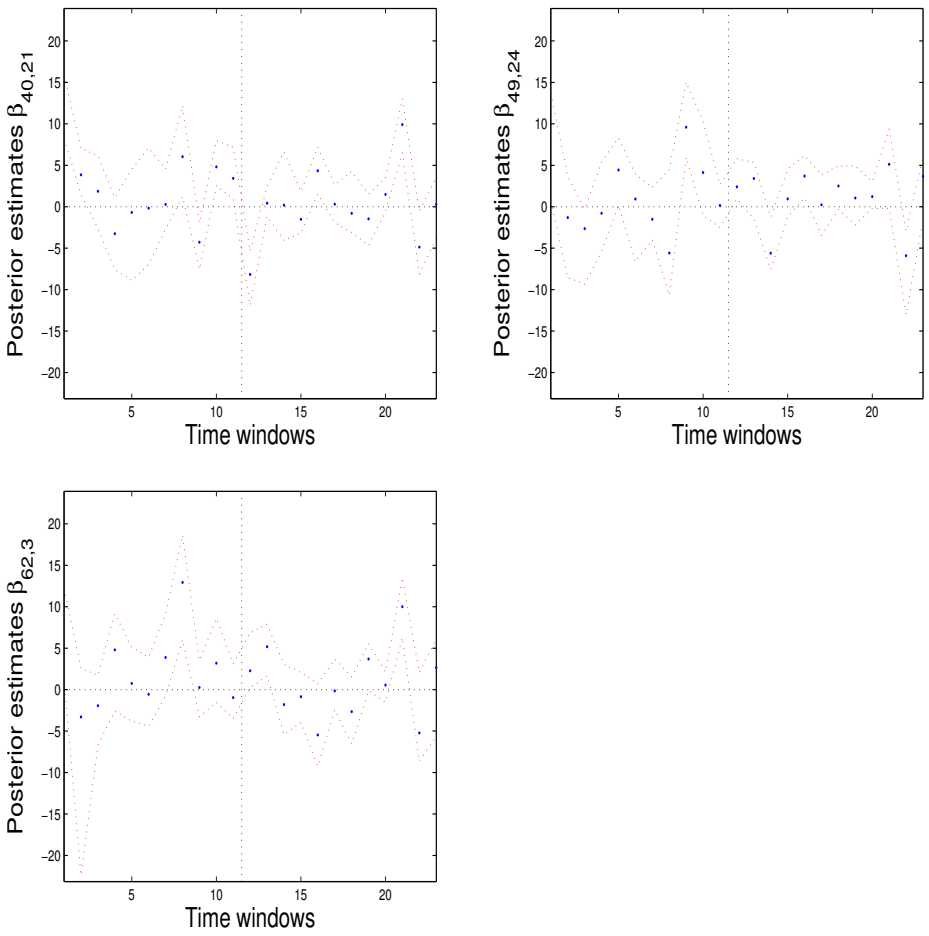

Figure 5: posterior estimates of the three pair-wise functional connectivities affected by both change-points. In each plot, a vertical dotted line separates the two contiguous experimental phases and a horizontal dotted line separates the negative and positive abscissae, which respectively identify inhibitory and excitatory functional connections. At both change-points cases the three functional connections become excitatory, which partially explains the increased network activity corresponding to the change-points.

pothesis is defined over the sample space of the last set of data observed. From this perspective, the methods developed in this paper follow the lines of Box [1980], Gelman et al. [1996] and Carota et al. [1996]. The former author defines the integration of the Bayesian and frequentist paradigms as follows: "sampling theory is needed for exploration and ultimate criticism of entertained models in the light of data, while Bayes' theory is needed for estimation". The integration between the two inferential paradigms proposed in this paper is slightly different from that defined by Box. Sampling theory is in fact not used for model criticism but for evaluating whether sequential Bayesian learning is supported by the data accrued over time.

A first remarkable consequence of the definition of change-point adopted in Section 1 is that the sequential test proposed in this work is not reversible in time. Lack of reversibility is due to the fact that the evidence in favour of the null hypothesis provided by the likelihood is weighted by the Kullback-Leibler statistic (3) using all the information about the model parameters accumulated since the occurrence of the last change-point. A second consequence of our definition of change is that arising of multimodality, a larger than expected concentration of the posterior density and the accrual of scarcely informative data are all detected as changes. Since the Kullback-Leibler statistic does not distinguish between these cases, in all cases a violation of the null hypothesis is detected and the learning process is reset. This procedure is justified if one accepts a definition of change based only on the expected information gain provided by the new data, as proposed in Section 1. More detailed definitions need to be developed to derive test statistics able to discriminate different types of departures from a null hypothesis of no change.

The alternative hypothesis formulated in Section 1 may not be the only

course of action available when a change-point is detected. For instance, if the first change occurs upon observing the ith+1 set of data, a family of i alternative hypothses obtains by letting

$$f(\theta_{i+1} \mid y_0, ..., y_{i+1}) \propto f(\theta_{i+1}) \prod_{j=w}^{i+1} P(y_j \mid \theta_{i+1}, y_0, ..., y_{j-1}), \tag{11}$$

with $w = 2, ..., i+1$. The alternative hypothesis used in this paper corresponds to letting $w = i+1$ in (11). Its interpretation is that a change-point causes the loss of all information accrued in the past about all model parameters. Since a change typically affects some but not all model parameters, this alternative hypothesis is not optimal because it discontinues the Bayesian sequential learning for the coefficients not affected by a change. This optimality can be achieved by specifying the null and the alternative hypotheses in terms of the marginal posterior distributions rather than of the joint posterior, leading to a multivariate generalisation of the change-point test proposed in this paper. Although our methodology can be extended in this direction, its computational cost would be high. For instance, for the example presented in Section 3, upon observing a new set of data one would have to compute 211 Kullback-Leibler statistics and to approximate their distributions under the null hypothesis in order to test for the occurrence of the corresponding marginal change-points.

Instead of selecting one alternative hypothesis, the null hypothesis of no change could be compared pair-wise with each of its i alternatives using their Bayes factor (Kass and Raftery [1995]). This approach was not considered in this work because for non-conjugate models computing the marginal likelihood of the ith+1 data under each of the i alternatives is rather cumbersome.

The simulation study reported in Section 2.1 shows empirically that for

a simple model the power of the Kullback-Leibler change-point test does not depend on the sample sizes of the data being compared. Proving this property of the statistic (3) in general requires the development of an approximation of its distribution under mild conditions on the form of the data sampling distribution. When both densities in (3) can be approximated as Gaussian, the closed form of the Kullback-Leibler statistic reported in example 2.1 provides a first step in this direction.

The analysis of the sheep data reported in Section 3 detects two change-points taking place respectively during the first experimental phase and at about three seconds after the onset of the visual stimulus. The occurrence of both changes corresponds to an increased level of overall network activity, which is partially explained by the activation of three excitatory functional connections. Although the two changes may be due to motor activity rather than to information processing, the small proportion of network coefficients significantly affected in both instances suggests that different mechanisms might be operating.

Acknowledgements

The author is grateful for numerous inspiring conversations with the members of CRiSM Jim Smith, Mark Steel and Jim Griffin, with Mathisca de Gunst and with Jaap van Pelt. Permission to use the multiple spike data analysed in Section 3 was kindly granted by Jiangfeng Feng.

References

H. Akaike. Likelihood of a model and information criteria. *Journal of Econometrics*, 16:3–14, 1981.

H. Akaike. On the likelihood of a time series model. *The Statistician*, 27:217–235, 1978.

L. Bélisle, P. abd Joseph, B. MacGibbon, D.B. Wolfson, and R. du Berger. Change-point analysis of neuron spike train data. *Biometrics*, 54:113 – 123, 1998.

R. Bellman. *Dynamic programming*. Dover Publications, first edition, 1957.

J. Bernardo. Expected information as expected utility. *The Annals of Statistics*, 7: 686–690, 1979.

J.M. Bernardo and A.F.M. Smith. *Bayesian theory*. Wiley, 2007.

G. E. P. Box. Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the American Statistical Association A*, 143:383–430, 1980.

D.R. Brillinger. Some statistical methods for random processes data from seismology and neurophysiology. *The Annals of Statistics*, 16:1–54, 1988.

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

E.N. Brown and R. Barbieri. Dynamic analyses of neural representations using the state-space modeling paradigm. *in: Madras, B., Von Zastrow, M. Colvis, C., Rutter, J. Shurtleff, D. and Pollock, J. eds. "The Cell Biology of Addiction", New York, Cold Spring Harbor Laboratory press*, 2006.

R.L. Brown, J. Dubin, and J.M. Ewans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society B*, 37: 149–192, 1975.

G. Buzsáki. Large scale recording of neuronal ensembles. *Nature Neuroscience*, 7: 446–451, 2004.

B.P. Carlin, A.E. Gelfand, and A.F.M. Smith. Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics*, 41:389–405, 1992.

C. Carota, G. Parmigiani, and N. Polson. Diagnostic measures for model criticism. *Journal of the American Statistical Association*, 91:753–762, 1996.

D. Crisan and A. Doucet. A survey of convergence results on particle filters for pratictioners. *IEEE transactions on signal processing*, 50:736–746, 2002.

F. Critchley, P. Marriott, and M. Salmon. Preferred point geometry and the local differential geometry of the Kullback-Leibler divergence. *The Annals of Statistics*, 22:1587–1602, 1994.

P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society B*, 68:411–436, 2006.

P. Diaconis and D. Ylvisaker. Conjugate priors for exponential families. *The Annals of Statistics*, 7:269–281, 1979.

A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10:197–208, 2000.

A. Doucet, J.F.G. Freitas, and N.J. Gordon. *Sequential Monte Carlo Methods in practice*. Springer; New York, 2001.

U.T. Eden, L.M. Frank, R. Barbieri, V. Solo, and E.N. Brown. Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Computation*, 16: 971–998, 2004.

S. E. Fienberg. Stochastic models for single neuron firing trains: a survey. *Biometrics*, 30:399–427, 1974.

A.E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

A. Gelman, X.L. Meng, and H.S. Stern. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6:733–807, 1996.

W.R. Gilks and C. Berzuini. Following a moving target - Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society B*, 63: 127–146, 2001.

N. Gordon, D. Salmond, and A.F.M. Smith. Novel approach to non-linear and non-Gaussian Bayesian state estimation. *Proceeding of the Institute of Electrical Engineering*, 140:107–113, 1993.

C. Goutis and C. Robert. Model choice in generalised linear models: A Bayesian approach via Kullback-Leibler projections. *Biometrika*, 85:29–37, 1998.

E. Gutiérrez-Peña. Moments for the canonical parameter of an exponential family under a conjugate distribution. *Biometrika*, 84:727–732, 1997.

P. Hall. On Kullback-Leibler loss and density estimation. *The Annals of Statistics*, 15:1491–1519, 1987.

P.J. Harrison and C.F. Stevens. Bayesian forecasting. *Journal of the Royal Statistical Society B*, 38:205–247, 1976.

T. Hastie. A closer look at the deviance. *The American Statistician*, 41:16–20, 1987.

S. Iyengar. The analysis of multiple neural spike trains. *In: Advances in Methodological and Applied Aspects of probability and Statistics; N. Bolakrishnan editor; Gordon and Breack*, pages 507–524, 2001.

M. Jain, M. Elhilali, N. Vaswami, J. Fritz, and S. Shamma. A particle filter for tracking adaptive neural responses in auditory cortex. *Submitted for publication*, 2007.

R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45, 1960.

R. E. Kass and A.E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995.

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

K.M. Kendrick, A.P. da Costa, A.E. Leigh, M.R. Hinton, and J.W. Peirce. Sheep don't forget a face. *Nature*, 414:165–166, 2001.

S. Kullback. *Information theory and Statistics*. Dover; New York, 1997.

S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951.

P.E. Latham, B.J. Richmond, P.G. Nelson, and S. Niremberg. Intrinsic dynamics in neuronal networks. I. Theory. *Journal of the Neurophysiology*, 83:808–827, 2000.

D. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27:986–1005, 1956.

C.R. Loader. Change point estimation using nonparametric regression. *The Annals of Statistics*, 24:1667–1678, 1996.

R.E. McCulloch. Information and the likelihood function in exponential families. *The American Statistician*, 42:73–75, 1988.

A. Mira and S. Petrone. Bayesian hierarchical nonparametric inference for change point problems. *Bayesian Statistics*, 5:693–703, 1996.

H.G. Muller. Change-points in nonparametric regression analysis. *The Annals of Statistics*, 20:737–761, 1992.

M. Okatan, M.A. Wilson, and E.N. Brown. Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural computation*, 17:1927–1961, 2005.

H.E. Plesser and W. Gerstner. Niose in integrate-and-fire neurons: from stochastic input to escape rates. *Neural Computation*, 12:367–384, 2000.

R.E. Quandt. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53: 873–880, 1958.

R.E. Quandt. Test of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, 55:324–330, 1960.

R.P.N. Rao. Hierarchical Bayesian inference in networks of spiking neurons. *Advances in NIPS; MIT press*, 17, 2005.

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

A. Smith, L. Frank, S. Wirth, M. Yanike, D. Hu, Y. Kubota, A.M. Graybiel, W. Suzuki, and E.N. Brown. Dynamic analysis of learning in behavioral experiments. *Journal of Neuroscience*, 24:447–461, 2004.

A. Smith, S. Wirth, W. Suzuki, and E.N. Brown. Bayesian analysis of interleaved learning and response bias in behavioral experiments. *Journal of Neurophysiology*, 97:2516–2524, 2006.

A. F.M. Smith and G.O. Roberts. Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, 55:3–23, 1993.

L. Srinivansan, U.T. Eden, A.S. Willsky, and E.N. Brown. A state-space analysis for reconstruction of goal-directed movements using neural signals. *Neural Computation*, 18:2465–2494, 2006.

M. Stone. Application of a measure of information to the design and comparison of regression experiments. *The Annals of Mathematical Statistics*, 30:55–70, 1959.

L. Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22:1701–1762, 1994.

W. Truccolo, U.T. Eden, M.R. Fellows, J.P. Donoghue, and E.N. Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *Journal of Neurophysiology*, 93:1074–1089, 2005.

N. Vaswami. Change detection in partially observed nonlinear dynamic systems with unknown change parameters. *Proceedings of the American Control Conference*, 6:5387 – 5393, 2004.

M. West and P.J. Harrison. *Bayesian forecasting and dynamic models*. Springer; New York, second edition, 1997.

M. West and P.J Harrison. Monitoring and adaptation in Bayesian forecasting models. *Journal of the American Statistical Association*, 81:741–750, 1986.

M. West, P.J. Harrison, and H.S. Migon. Dynamic generalised linear models and Bayesian forecasting. *Journal of the American Statistical Association*, 80:73–83, 1985.

A. Zellner. Information processing and Bayesian analysis. *Journal of Econometrics*, 107:41 – 50, 2002.