

Change-Point Analysis of Neuron Spike Train Data

Patrick Bélisle,¹ Lawrence Joseph,^{1,2,*} Brenda MacGibbon,³
David B. Wolfson,⁴ and Roxane du Berger¹

¹Division of Clinical Epidemiology, Montreal General Hospital, Department of Medicine,
1650 Cedar Avenue, Montreal, Quebec H3G 1A4, Canada

²Department of Epidemiology and Biostatistics, McGill University,
1020 Pine Avenue West, Montreal, Quebec H3A 1A2, Canada

³Département de mathématiques, Université du Québec à Montréal,
C.P. 8888, Succursale Centre-Ville, Montréal, Québec H3C 3P8, Canada

⁴Department of Mathematics and Statistics, McGill University, Burnside Hall,
805 Sherbrooke Street West, Montreal, Quebec H3A 2K6, Canada

SUMMARY

In many medical experiments, data are collected across time, over a number of similar trials, or over a number of experimental units. As is the case of neuron spike train studies, these data may be in the form of counts of events per unit of time. These counts may be correlated within each trial. It is often of interest to know if the introduction of an intervention, such as the application of a stimulus, affects the distribution of the counts over the course of the experiment. In such investigations, each trial generates a sequence of data that may or may not contain a change in distribution at some point in time. Each sequence of integer counts can be viewed as arising from a Poisson process and are therefore independently distributed or as an integer-valued time series that allows for correlations between these counts. The main aim of this paper is to show how the ensemble of sample paths may be used to make inference about the distribution of the instantaneous times of change in a given population. This will be accomplished using a Bayesian hierarchical model for these change-points in time. A bonus of these models is they also allow for inference about the probability of a change in each unit and the magnitude of the effects, if any. The use of such change-point models on integer-valued time series is illustrated on neuron spike train data, although the methods can be applied to other situations where integer-valued processes arise.

1. Introduction

Neurophysiology has become a field of increasing attention as scientists attempt to explain system response by describing the reaction of individual neurons or groups of neurons (Arieli et al., 1996; Hanes and Schall, 1996). In this and other research, the time it takes for a single neuron to react is taken into account as part of a cascade of events following application of a stimulus (Koch, 1997).

Even in a resting or nonstimulated state, neurons continue to show spontaneous electrical activity that is stochastic in intensity. This activity is observed as a sequence of discrete electrical discharges, called spike trains or action potentials. While the duration and magnitude of the discharges vary little and carry almost no information, the counts of spike trains in equal intervals are informative and may be modelled as random variables that follow a point process. The usual response to a stimulus may then be manifested as a sudden change in the parameters of this process. The time of increase in electrical activity lags behind the time that the stimulus is applied. It is the distribution

* *Corresponding author's email address:* joseph@binky.ri.mgh.mcgill.ca

Key words: Autoregressive time series; Bayesian inference; Change-point; Gibbs sampler; Integer-valued autoregressive model; Integer-valued time series; Longitudinal data; Markov-chain Monte Carlo.

of this time lag, or reaction time, that is of chief interest here, although we also examine the before and after rates.

Commenges and Seal (1985) describe an experiment designed to study the time-to-reaction following the stimulation of a neuron located in the posterior parietal cortex of a monkey. A stimulus was applied to the neuron. The electrical discharge of the neuron was recorded for a time period both before and after the stimulus. This procedure was repeated on the same neuron 35 times, resulting in 35 data sequences; since the sequence of runs was terminated before the cell tired, it is reasonable to assume that the reaction times remained identically distributed (but not identical) through these 35 trials.

Commenges and Seal (1985) considered a parametric renewal process with intensity that changes at one or possibly two time points, although change-point estimation is only discussed for each single trial separately, after which the change-points are combined in an *ad hoc* fashion (Commenges, Seal, and Pinatel, 1986). Wegman and Habib (1992) and Brillinger (1992) provide reviews of stochastic modelling for nerve cell spike train data, the latter emphasizing the influence of networks of nerve cells on the action potential process of a single designated cell.

This paper illustrates how a hierarchical Bayesian change-point model can be employed to combine data across trials. A preliminary autocorrelation analysis of the data suggested that there was no reason to reject the assumption of independence of the counts of spike trains in disjoint intervals. Consequently, we initially modelled each sequence of counts as independent Poisson random variables that arise from a Poisson process with a change-point. Then, because of some concern about the ability of the autocorrelation tests to detect dependence, as well as possible physiological justification, we modelled counts as an integer-valued autoregressive (INAR) time series. Such a model, which is a discrete analog of the standard autoregressive process for continuous data, allows for correlations of the counts. As far as we are aware, this is the first time a Bayesian approach is taken for inference about INAR parameters and the first time INAR models with a change-point have been considered.

Analysis of the above data, which were provided by Commenges (personal communication), is deferred to Section 4, after details of the models are provided in Section 2. A Bayesian approach to estimating the parameters of both models is described in Section 3. The final section contains further discussion.

2. The Models

Assume that there are data in the form of an $M \times N$ array

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1T_1} & X_{1(T_1+1)} & \cdots & X_{1N} \\ X_{21} & X_{22} & \cdots & X_{2T_2} & X_{2(T_2+1)} & \cdots & X_{2N} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ X_{M1} & X_{M2} & \cdots & X_{MT_M} & X_{M(T_M+1)} & \cdots & X_{MN} \end{pmatrix}. \quad (1)$$

Each vector $X_{i1}, X_{i2}, \dots, X_{iN}$ represents observations over time from the i th row or sequence, $i = 1, 2, \dots, M$. A change-point is said to have occurred at $T_i = \tau_i$ in sequence i , $i = 1, 2, \dots, M$ and $1 \leq \tau_i \leq N - 1$, if the random vector $X_{i1}, X_{i2}, \dots, X_{i\tau_i}$, has distribution F_{i1} , which is different from the distribution F_{i2} of the random vector $X_{i\tau_i+1}, X_{i\tau_i+2}, \dots, X_{iN}$. If $T_i = N$, then no change has occurred in row i . The distribution of the time points of change T_i and unknown parameters of the distributions F_{ik} , $i = 1, \dots, M$, $k = 1, 2$, are to be estimated from a realization of the data matrix (1). When there are multiple paths or sequences, we shall refer to a multipath change-point problem to distinguish it from the classical single-path problem, when $M=1$.

It is assumed that the times of change T_i in each row or sequence are themselves independent and identically distributed in a given population, following a distribution $g(t) = Pr\{T_i = t\}$, $i = 1, \dots, M$, $t = 1, \dots, N$, which is to be estimated. If $g(N) > 0$, then it is possible that there is no change in some rows. The introduction of $g(\cdot)$ does not necessarily mean that each subject in the population has exactly the same change-point. It is emphasized that it represents the probabilities for the location of the change-point for a randomly selected unit in the population. While it is well known (Hinkley, 1970) that the single-path maximum likelihood estimator of the change-point is not consistent as N increases, the nonparametric estimator of $g(\cdot)$ in the multipath case has been shown in Joseph and Wolfson (1993) and Joseph, Vandal, and Wolfson (1996b) to be consistent as M increases under certain regularity conditions.

The case where $M = 1$ has received considerable attention in the literature, including maximum likelihood approaches of Hinkley (1970), Hinkley and Hinkley (1970), and Yao (1987, 1990), nonparametric methods of Pettitt (1979), and the Bayesian approach of Smith (1975). Carlin, Gelfand,

and Smith (1992) used the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990) to find marginal posterior distributions in a hierarchical single-path change-point model. Introducing $M \geq 2$ considerably broadens the applicability of change-point models (see Lange, Carlin, and Gelfand, 1992; Joseph et al., 1996a).

As a first approximation, we assume that the process of spike trains is a Poisson process. It then follows that X_{ij} , $j = 1, 2, \dots, N$, are independent Poisson random variables for each fixed $i = 1, 2, \dots, M$. Next, we investigated the possibility of correlations between counts, where it is important to recognize that the data are integer-valued. Al Osh and Alzaid (1987) and Jin-Guan and Yuan (1991), based on an operator of Steutel and Van Harn (1979), have devised nonnegative integer-valued time series, called INAR processes, that have analogous correlation structure to an autoregressive real-valued time series.

The Steutel–Van Harn operator (Steutel and Van Harn, 1979; Al Osh and Alzaid, 1987) is defined by the sum $\sum_{i=1}^X Y_i$, where Y_i is a sequence of independent and identically distributed (i.i.d.) Bernoulli(α) random variables, $\alpha \in [0, 1]$, and X is a Poisson(λ) random variable independent of Y_i , $i = 1, 2, \dots$. A useful modification of this operator, which Gauthier and Latour (1994) call a generalized INAR or GINAR model, is to define α and X as above, but generalize Y_i to be arbitrary i.i.d. integer-valued random variables with finite mean α .

Let ϵ_t , $t = 1, 2, \dots$, represent a sequence of Poisson random variables with parameter δ , independent of X , Y , or α . The sequence ϵ_t , $t = 1, 2, \dots$, is usually called the innovation process. Then either of the above series can be used to generate an integer-valued time series of order p (p a nonnegative integer), with $\{\alpha_k\}_{k=1, \dots, p}$ satisfying $\alpha_k \in [0, 1]$, $k = 1, 2, \dots, p-1$, and $\alpha_p \in (0, 1]$, by specifying

$$X_t = \sum_{k=1}^p \sum_{i=1}^{X_{t-k}} Y_i + \epsilon_t$$

for $t \geq 2$ and $X_1 \sim \text{Poisson}(\lambda)$. If the Steutel–Van Harn operator is used and ϵ_t is a sequence of Poisson(λ) random variables, this is called a Bernoulli–Poisson model of order p , and if the operator defined by Gauthier and Latour (1994) is used, then this is termed a completely Poisson process of order p . Here we will discuss in detail only the Bernoulli–Poisson model of order $p = 1$, although similar techniques to those we present can be extended to higher order models and completely Poisson processes. Note that if $\delta = \lambda(1 - \alpha)$, then the Bernoulli–Poisson process is stationary.

3. Estimation of Parameters via the Gibbs Sampler

3.1 Independent Counts Model with a Change-Point

The likelihood for the independent counts model with one change-point described in Section 2 is given by

$$f(x | \theta_1, \theta_2, \pi) = \prod_{i=1}^M \sum_{h=1}^N \left\{ \prod_{j=1}^h f_1(x_{ij} | \theta_1) \right\} \left\{ \prod_{j=h+1}^N f_2(x_{ij} | \theta_2) \right\} \pi_h,$$

where θ_1 and θ_2 , possibly vector valued, are the parameters of the densities f_1 and f_2 , respectively, and $\pi = (\pi_1, \pi_2, \dots, \pi_N)$, where $\pi_h = \text{Pr}\{T_i = h\}$, $h = 1, 2, \dots, N$, and $i = 1, 2, \dots, M$. Inference using this likelihood is difficult since it takes the form of a mixture. However, conditional on knowledge of the latent data (Tanner and Wong, 1987) τ_i , $i = 1, 2, \dots, M$, the change-points in each data sequence, the likelihood simplifies to

$$f(x | \theta_1, \theta_2, \tau_1, \dots, \tau_M) = \prod_{i=1}^M \left\{ \prod_{j=1}^{\tau_i} f_1(x_{ij} | \theta_1) \right\} \left\{ \prod_{j=\tau_i+1}^N f_2(x_{ij} | \theta_2) \right\}. \quad (2)$$

When each x_{ij} , $i = 1, \dots, M$, $j = 1, \dots, N$, follows a Poisson distribution, the parameters in the model are $\theta_1 = \lambda_1 = (\lambda_{11}, \dots, \lambda_{M1})$ and $\theta_2 = \lambda_2 = (\lambda_{12}, \dots, \lambda_{M2})$, vectors of the means of the Poisson distributions before and after the change-point in each row, $\pi = (\pi_1, \dots, \pi_N)$, the multinomial probabilities that a change occurs at position i in each row, $i = 1, \dots, M$, and $\tau = (\tau_1, \dots, \tau_M)$, the unobserved latent data representing the change-points in each row.

For simplicity, one may choose conjugate prior distributions, although nonconjugate priors can also be accommodated, as in Müller (1991). Since the gamma distribution is conjugate to a Poisson

likelihood function and the Dirichlet distributions form a conjugate family for the parameters of a multinomial random variable, the prior distributions in the case that the x_{ij} are Poisson could be given as

$$f(\lambda_{ik}) = \frac{1}{\Gamma(a_{ik})b_{ik}^{a_{ik}}} \lambda_{ik}^{a_{ik}-1} \exp\left(-\frac{\lambda_{ik}}{b_{ik}}\right), \quad i = 1, \dots, M, \quad k = 1, 2, \tag{3}$$

and

$$f(\pi_1, \dots, \pi_N) = \frac{\Gamma(\alpha_0)}{N} \prod_{i=1}^N \pi_i^{\alpha_i-1}, \tag{4}$$

$$\prod_{l=1}^N \Gamma(\alpha_l)$$

where $\sum_{i=1}^N \pi_i = 1$, $\alpha_0 = \sum_{i=1}^N \alpha_i$, $\alpha_i > 0$, $i = 1, \dots, N$, and the a_{ik} 's, b_{ik} 's, and α_i 's are chosen according to the available prior information.

Implementation of the Gibbs sampler to find the marginal posterior distributions requires the specification of the full conditional distribution of the parameters, i.e., the conditional distribution of each parameter given the values of all of the other parameters. These are specified following standard procedures for conjugate analyses (see DeGroot, 1970) as

$$f(\lambda_{i1} | x, \tau_i) \sim \text{gamma}\left(a_{i1} + \sum_{j=1}^{\tau_i} x_{ij}, \left(\tau_i + \frac{1}{b_{i1}}\right)^{-1}\right), \tag{5}$$

$$f(\lambda_{i2} | x, \tau_i) \sim \text{gamma}\left(a_{i2} + \sum_{j=\tau_{i+1}}^N x_{ij}, \left(N - \tau_i + \frac{1}{b_{i2}}\right)^{-1}\right), \tag{6}$$

$$Pr\{\tau_i = t | \lambda_1, \lambda_2, \pi, x\} = \frac{\left\{ \prod_{j=1}^t \frac{(\lambda_{i1})^{x_{ij}} \exp(-\lambda_{i1})}{x_{ij}!} \right\} \left\{ \prod_{j=t+1}^N \frac{(\lambda_{i2})^{x_{ij}} \exp(-\lambda_{i2})}{x_{ij}!} \right\} \pi_t}{\sum_{k=1}^N \left\{ \prod_{j=1}^k \frac{(\lambda_{i1})^{x_{ij}} \exp(-\lambda_{i1})}{x_{ij}!} \right\} \left\{ \prod_{j=k+1}^N \frac{(\lambda_{i2})^{x_{ij}} \exp(-\lambda_{i2})}{x_{ij}!} \right\} \pi_k}, \tag{7}$$

and

$$f(\pi | \tau) \sim \text{Dirichlet}(\beta'), \tag{8}$$

where β'_k , the k th element of β' , is given by $\beta_k + \sum_{i=1}^M I_{\{\tau_i=k\}}$, where $I_{\{y\}}$ is the indicator function for the set $\{y\}$.

The Gibbs sampler algorithm proceeds by drawing a random sample from each full conditional distribution (5)–(8) in turn. The parameters sampled from the immediately preceding random draw are used in the conditional distribution for subsequent draws. A large number of iterations are run, and after discarding iterates from an initial burn-in period to allow for convergence of the algorithm, the remaining random vectors, whose components are the samples drawn from (5) through (8) in each iteration, can be regarded as samples from the joint posterior distribution of the parameters, from which inferences can be made.

We used the method of Raftery and Lewis (1992) to estimate the required number of iterations and burn-in for each parameter based on the output from a preliminary run of the sampler. Marginal posterior density estimates were generated by what has become known as the Rao–Blackwell method (see Gelfand and Smith, 1990). Four independent runs of the Gibbs sampler were carried out as an additional convergence check, and convergence was assumed only if all marginal densities were identical for all practical purposes.

3.2 Integer-Valued Autoregressive Model with a Change-Point

In order to illustrate the methods for correlated count data, we considered the Bernoulli–Poisson model of order $p = 1$, suitably modified to accommodate the possibility of a change-point. This model was chosen because it represents a natural discrete counterpart to the well-known versatile

autoregressive models of continuous-time series. Conditional on knowledge of the latent data τ_i , $i = 1, \dots, M$, the likelihood again simplifies to a product of terms rather than a mixture, although the before and after change-point terms are now more complex, as discussed below. In the case of a first-order Bernoulli–Poisson INAR model, the parameters to be estimated are $\theta_1 = (\lambda_1, \delta_1, \alpha_1) = (\lambda_{11}, \dots, \lambda_{M1}, \delta_{11}, \dots, \delta_{M1}, \alpha_{11}, \dots, \alpha_{M1})$ and $\theta_2 = (\lambda_2, \delta_2, \alpha_2) = (\lambda_{12}, \dots, \lambda_{M2}, \delta_{12}, \dots, \delta_{M2}, \alpha_{12}, \dots, \alpha_{M2})$, where λ_1 and λ_2 represent, respectively, the vectors of Poisson means before and after the change-point, δ_1 and δ_2 the mean vectors, respectively, of the Poisson innovation processes before and after the change-point, and α_1 and α_2 , respectively, the Bernoulli parameters before and after the change-point. The stationarity condition for the Bernoulli–Poisson process, given in Section 2, applies separately to each of the before and after change-point variables in each sequence. Hence, δ_k is a function of α_k and λ_k , $k = 1, 2$. The vectors π and τ remain as previously defined following equation (2). Note that if $\alpha_1 = \alpha_2 = \mathbf{0}$, then the model reduces to the independent model described in Section 3.1. This can be seen by substituting $\alpha_{1i} = 0$ in the likelihood equations below.

The likelihood function can now be derived as follows. Let l_{1i} denote the first segment of the likelihood function in sequence i , that is, the segment from the first observation up to and including the observation at the change-point τ_i . Then

$$\begin{aligned} l_{1i} &= P[X_{i,1} = x_{i,1}] \prod_{j=2}^{\tau_i} P[X_{i,j} = x_{i,j} \mid X_{i,j-1} = x_{i,j-1}] \\ &= P[X_{i,1} = x_{i,1}] \prod_{j=2}^{\tau_i} \sum_{k=0}^{x_{i,j}} P[\text{bin}(x_{i,j-1}, \alpha_{1i}) = k \cap \text{Poisson}(\delta_{1i}) = x_{i,j} - k] \\ &= \frac{e^{-\lambda_{1i}} \lambda_{1i}^{x_{i,1}}}{x_{i,1}!} \prod_{j=2}^{\tau_i} \sum_{k=0}^{\min(x_{i,j}, x_{i,j-1})} \binom{x_{i,j-1}}{k} \alpha_{1i}^k (1 - \alpha_{1i})^{x_{i,j-1} - k} \frac{e^{-\delta_{1i}} \delta_{1i}^{x_{i,j} - k}}{(x_{i,j} - k)!}, \end{aligned}$$

where $\text{bin}(x, \alpha)$ denotes the binomial probability function with x trials and parameter α and $\text{Poisson}(\delta)$ denotes a Poisson probability function with mean δ . The stationarity condition $\delta_{1i} = \lambda_{1i}(1 - \alpha_{1i})$ implies

$$\begin{aligned} l_{1i} &= \frac{e^{-\lambda_{1i}} \lambda_{1i}^{x_{i,1}}}{x_{i,1}!} \prod_{j=2}^{\tau_i} e^{-\lambda_{1i}(1 - \alpha_{1i})} \lambda_{1i}^{x_{i,j}} (1 - \alpha_{1i})^{x_{i,j-1} + x_{i,j}} \\ &\quad \times \sum_{k=0}^{\min(x_{i,j}, x_{i,j-1})} \binom{x_{i,j-1}}{k} \frac{\alpha_{1i}^k}{(1 - \alpha_{1i})^{2k} \lambda_{1i}^k (x_{i,j} - k)!}. \end{aligned}$$

Defining $N_{ab}^i(\tau_i) = \#\{j : j \leq \tau_i, x_{i,j-1} = a \text{ and } x_{i,j} = b\}$ and applying the binomial theorem, it can be shown that

$$\begin{aligned} l_{1i} &= \frac{1}{x_{i,1}!} \sum_{k=0}^{N_{11}^i(\tau_i)} \sum_{l=0}^{N_{12}^i(\tau_i)} \sum_{m=0}^{N_{21}^i(\tau_i)} \binom{N_{11}^i(\tau_i)}{k} \binom{N_{12}^i(\tau_i)}{l} \binom{N_{21}^i(\tau_i)}{m} e^{-\lambda_{1i}[1 + (1 - \alpha_{1i})(\tau_i - 1)]} \\ &\quad \times \lambda_{1i}^{\sum_{j=1}^{\tau_i} x_{i,j} - k - l - m} \frac{\alpha_{1i}^{k+l+m} (1 - \alpha_{1i})^{x_{i,1} + x_{i,\tau_i} + 2 \sum_{j=2}^{\tau_i} x_{i,j} - 2(k+l+m)}}{2^{N_{02}^i(\tau_i) + N_{12}^i(\tau_i) - l - m}}. \end{aligned}$$

The second half of the likelihood, l_{2i} , has an analogous form, and the entire likelihood given (τ_1, \dots, τ_M) can then be written as $l(x \mid \theta_1, \theta_2, \tau_1, \dots, \tau_M) = \prod_{k=1}^2 \prod_{i=1}^M l_{ki}$. The same prior distributions as described in Section 3.1 can be used for π , λ_1 , and λ_2 [see equations (3) and (4)]. Since the components of α_1 and α_2 are restricted to the range $[0, 1]$ and represent binomial parameters, beta prior densities were used.

The full conditional distributions for λ_{1i} and λ_{2i} are in the form of a mixture of gamma densities, while the full conditional distributions of the α_{1i} 's are also in the form of a mixture, although the mixture components are not from a standard distribution. We used a sampling importance resamp-

ling algorithm (Rubin, 1987) to draw random samples from the latter mixture. We first compared each component of this mixture to a beta density matched for mean and variance. Since each fit was very close, we used this mixture of beta densities as the proposal density for the sampling importance resampling algorithm. The full conditional densities are

$$f(\lambda_{1i} \mid \mathbf{x}, \tau_i, \alpha_{1i}, \alpha_{2i}, \lambda_{2i}) \propto l_{1i} \times f(\lambda_{1i}) \\ \propto \sum_{k=0}^{N_{11}^i(\tau_i)} \sum_{l=0}^{N_{12}^i(\tau_i)} \sum_{m=0}^{N_{21}^i(\tau_i)} w_{1iklm}^* \text{gamma}(A_{1iklm}^*, (B_{1i}^*)^{-1}),$$

where $A_{1iklm}^* = \sum_{j=1}^{\tau_i} x_{i,j} - k - l - m + a^{\lambda_1}$, $B_{1i}^* = 1 + (1 - \alpha_{1i})(\tau_i - 1) + b^{\lambda_1}$, and

$$w_{1iklm}^* = \binom{N_{11}^i(\tau_i)}{k} \binom{N_{12}^i(\tau_i)}{l} \binom{N_{21}^i(\tau_i)}{m} \frac{2^{l+m} \alpha_{1i}^{k+l+m}}{(1 - \alpha_{1i})^{2(k+l+m)}} \frac{\Gamma(A_{1iklm}^*)}{(B_{1i}^*)^{A_{1iklm}^*}}.$$

The full conditional distribution of λ_{2i} is similar. For the binomial parameters,

$$f(\alpha_{1i} \mid \mathbf{x}, \tau_i, \lambda_{1i}, \alpha_{2i}, \lambda_{2i}) \\ = \sum_{k=0}^{N_{11}^i(\tau_i)} \sum_{l=0}^{N_{12}^i(\tau_i)} \sum_{m=0}^{N_{21}^i(\tau_i)} w_{1iklm} \frac{1}{G(A_{1i}, B_{1iklm}, C_{1iklm})} \times e^{-A_{1i}(1-\alpha_{1i})} \alpha_{1i}^{B_{1iklm}} (1 - \alpha_{1i})^{C_{1iklm}},$$

where

$$G(A_{1i}, B_{1iklm}, C_{1iklm}) = \int_{\alpha_{1i}=0}^1 e^{-A_{1i}(1-\alpha_{1i})} \alpha_{1i}^{B_{1iklm}} (1 - \alpha_{1i})^{C_{1iklm}} d\alpha_{1i}, \\ w_{1iklm} = \binom{N_{11}^i(\tau_i)}{k} \binom{N_{12}^i(\tau_i)}{l} \binom{N_{21}^i(\tau_i)}{m} \frac{2^{l+m}}{\lambda_{1i}^{k+l+m}} G(A_{1i}, B_{1iklm}, C_{1iklm}), \\ A_{1i} = \lambda_{1i}(\tau_i - 1), \\ B_{1iklm} = k + l + m + a^{\alpha_1} - 1,$$

and

$$C_{1iklm} = x_{i,1} + x_{i,\tau_i} + 2 \sum_{j=2}^{\tau_i-1} x_{i,j} - 2(k + l + m) + b^{\alpha_1} - 1.$$

A similar full conditional distribution is obtained for α_{2i} , and the full conditional distributions for τ_i , $i = 1, \dots, M$, and π have similar forms as in the independent case of Section 3.1, with the likelihood function adjusted as indicated above. The quantities $(a^{\alpha_1}, b^{\alpha_1})$ represent the beta parameters for the prior distributions for α_{1i} . The subscript i is omitted from these prior parameters since a common prior density is assumed across all trials, $i = 1, \dots, M$. This notation was used in order to avoid confusion with the Dirichlet parameters introduced in Section 3.1. Detailed derivations of the likelihood function and all full conditional distributions are available in the technical report by Bélisle et al. (1997). Again, the Gibbs sampler proceeds by sampling each variable in turn from its full conditional distribution.

4. Neuron Spike Train Data Analysis

Data consisting of counts of electrical discharges in 20-millisecond (ms) intervals approximately one-half second before and after a stimulus was applied to the neuron at $t = 500$ ms were observed on $M = 35$ data sequences. Each time, the neuron was allowed to return to the resting state before the experiment was resumed. See Figure 1. All sequences had 25 observations before the stimulus was applied, but the number of observations after the stimulus varied between 11 and 24. This variation should not cause substantial bias in estimating π unless there is evidence that the change-point occurs after approximately 220 ms poststimulus, which was not the case in this data set.

The output produced by the Gibbs sampler for π is a sample from a Dirichlet distribution in N dimensions. Since this distribution is difficult to visualize, summary statistics are necessary. In particular, the means of the marginal Dirichlet posterior distributions can be calculated, and posterior marginal densities for selected change-point probabilities may be plotted. The latter display the variability about the Dirichlet means and are calculated here as a Rao–Blackwell mixture of beta densities over the set of random samples generated by the Gibbs algorithm. Within each iter-

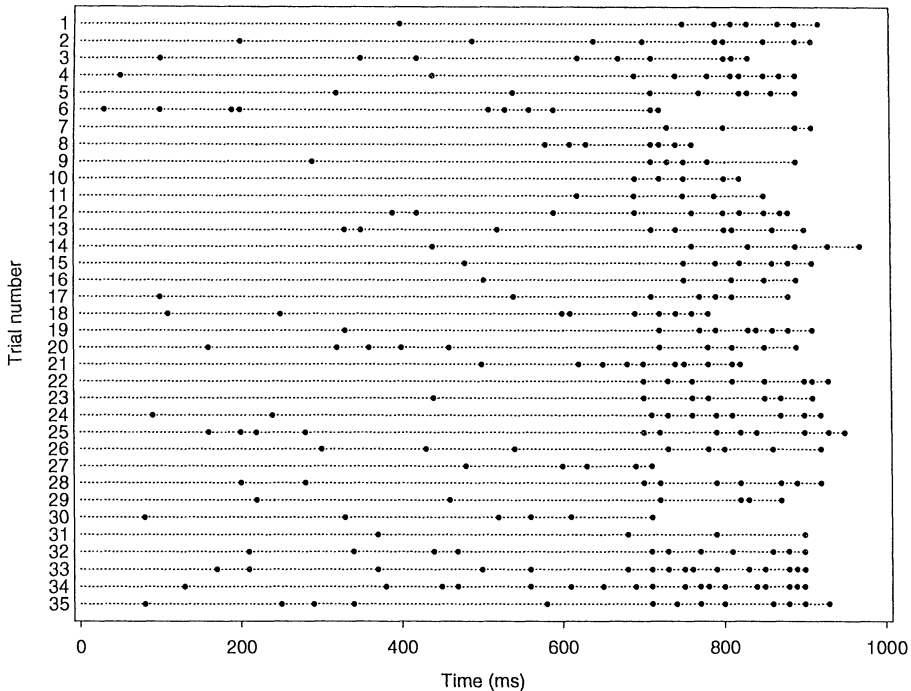


Figure 1. Data from the experiment with $M = 35$ trials. A stimulus was applied at 500 ms. Each \bullet represents a spike train at the indicated time. The total length of the broken lines indicate the follow-up times in each trial.

ation, each sequence may have $\tau_i < N$ or $\tau_i = N$, $i = 1, \dots, M$. A useful statistic is then $\{\# \text{ times } \tau_i < N\} / \text{number of iterations}$. This approximates the sequence or trial-specific probability of a change-point.

4.1 Results from the Independent Poisson Model with a Change-Point

Prior densities. The overall strategy was to create conservative prior distributions, in the sense that prior values for each parameter were selected to cover a somewhat wider interval than the available prior information might suggest. In this way, the entire ranges of most likely values for the parameters were covered by relatively flat portions of the prior densities so that the data themselves would contribute most of the information in the posterior densities.

Accordingly, a Dirichlet prior density with $\alpha_1 = \alpha_2 = \dots = \alpha_{24} = 0$ and $\alpha_{25} = \alpha_{26} = \dots = \alpha_{49} = 0.04$ was used. The sample size equivalent of this prior density is one observation ($\sum \alpha_i = 1$), so that $35/36 = 97\%$ of the information in the marginal posterior density on π would come from the data. We used prior gamma(4, 0.03) and gamma(8, 0.03) distributions for the before and after Poisson parameters, with mean rates of 0.12 and 0.24 firings per 20-ms interval and standard deviations of 0.06 and 0.085 firings per 20-ms interval, respectively. Hence, at least a doubling in the usual rate of discharges is anticipated after the stimulus, although the large standard deviations allow for a wide range of other values to be chosen by the data.

Results. The Gibbsit software made available by Raftery and Lewis indicated that 5100 iterations, including a burn-in of 100 iterations, is sufficient to estimate 95% credible intervals for each parameter that would have true coverage between 92.5% and 97.5% with high probability.

The mean marginal posterior change-point probability for τ_{34} , which corresponds to 680 ms, was equal to 0.85, indicating that there is indeed a change in electrical activity following the application of the stimulus, occurring roughly 180 ms after the stimulus. None of the other change-point mean marginal probabilities was greater than 0.025, and in particular, there was a negligible estimated probability of no change. The overall mean rate of discharges before the change-point is estimated to be 0.10 per 20-ms interval and 0.31 per 20-ms interval after the change, so that an approximate three-fold difference from the baseline rate occurs on average after the change-point. While a mean difference of 0.21 per 20 firings per 20 ms was observed, there were variations in the rates from se-

quence to sequence. The minimum difference was 0.11, while the maximum difference was 0.28. The standard deviation of the mean differences was 0.04 firings per 20-ms interval.

The marginal posterior distribution for τ_{34} is given in Figure 2. This figure indicates that at least 60% of similar trials will change at 680 ms and that the proportion could be almost as high as 100%.

The mean of the trial-specific probabilities of a change-point was 0.9999, so under this model, there is virtual certainty that there will be a change-point in each trial.

A sensitivity analysis was performed by allowing the change-point to occur at any time rather than being restricted to occurring after 500 ms and by moving the means of the prior densities of λ_{2i} 50% closer to the prior means of λ_{1i} . The results remained nearly identical to those reported above.

4.2 Results from the INAR Model with a Change-Point

Prior densities. We used the same prior distributions for the λ_1 , λ_2 , and π parameters as in the independent Poisson model. We initially considered using uniform(beta(1,1)) densities for the parameters α_{ki} , $k = 1, 2$, and $i = 1, \dots, M$. However, with this prior density, the INAR model would lead to very strong correlations ($\alpha_{ki} > 0.8$) between observations, which seems unlikely. Consequently, beta(1.5, 5) distributions were selected since they are concentrated between 0 and 0.6, the latter seeming to be a reasonable upper bound. Nevertheless, we performed a sensitivity analysis using both beta(1, 1) and beta(1.7, 4) densities, both of which allow for higher α_{ki} values. Since the results from the latter two analyses were very similar, they are not discussed further here.

Results. The Gibbsit algorithm again indicated that 5100 iterations were sufficient, including a burn-in of 100 iterations.

The mean marginal posterior change-point probability for τ_{34} was 0.80. The second largest mean change-point probability was 0.036, on τ_{35} . When these two are summed, one arrives at a value close to the value on τ_{34} found in the independent model. Again, there was a negligible estimated probability of no change. The overall mean rate of discharges before the change-point is estimated to be 0.11 per 20-ms interval and 0.30 per 20-ms interval after the change. The mean difference is 0.19 per 20 ms, again similar to that estimated by the independent model. The standard deviation

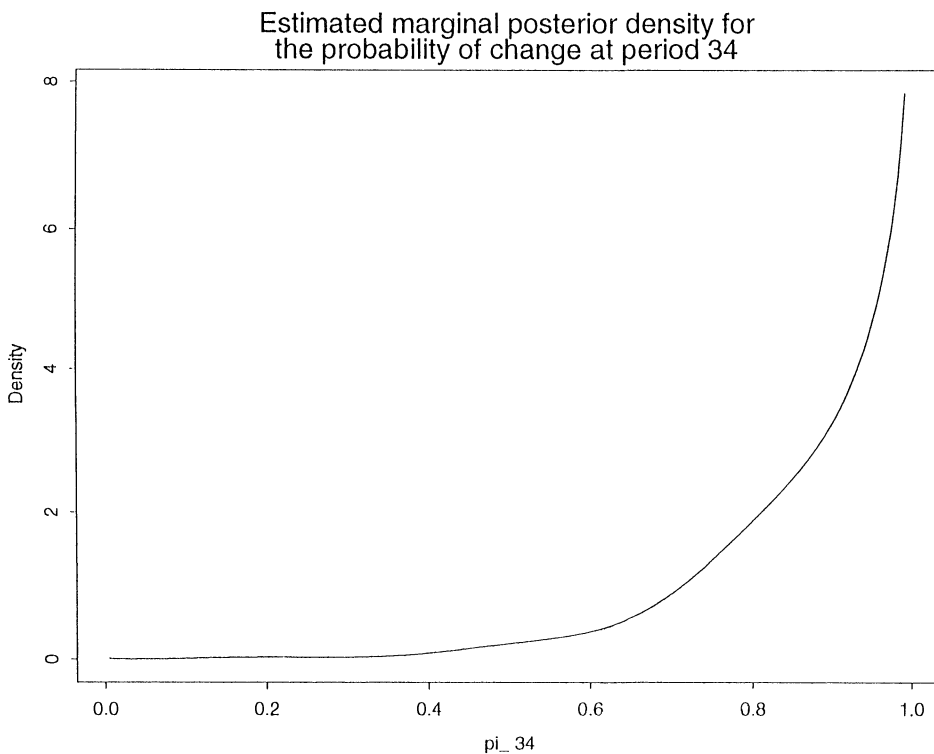


Figure 2. Marginal posterior densities for the probability of a change-point at τ_{34} , using the independent model of Section 3.1.

of the mean differences was again 0.04. The mean of the trial-specific probabilities of a change-point dropped slightly to 0.9624.

The posterior densities of the α_{ki} were investigated. The overall mean across trials was 0.25, close to the prior mean of 0.23. These means were similar both pre- and post-change-point. Even though there were up to 49 observations per sequence, there was little opportunity to update the prior densities on these parameters from these data. This is because most intervals prior to the change-point contained no spike trains, in which case the full conditional distribution for the α_{ki} 's closely resembles the prior density. This can be seen from the form of the full conditional density for α_{ki} when there are no spike trains (see Section 3.2). While there were many more spike trains after the change-point, the change-points occurred toward the end of a period of observation, again affording little data with which to update the prior distributions of the α_{ki} . While there was little evidence of correlation in this data set, one must conclude that much longer observation times would be required to definitively settle this issue, which remains open to further study. In any case, both models seem to be equivalent for estimating the distribution of the reaction times to the stimulus, at least in this experiment.

A Bayes factor (Kass and Raftery 1995) was estimated to compare the two models through the information contained in the data. The Bayes factor was calculated by computing the average likelihood values of each model across the Gibbs sampler iterations and was found to have a value of approximately 10^{14} in favour of the independent model. Given that virtually identical results were obtained from both models but that 70 fewer parameters are required for the independent model, this value is not surprising.

5. Discussion

In the data analysed here, the overall conclusions were not greatly affected by the choice of whether or not to incorporate correlations. In other cases, however, longer periods of observation and/or stronger correlations between observations may exist, making it important to account for them via an INAR or similar model. While we used a low-order binomial-Poisson model, the methodology described here can be extended to higher order and Poisson-Poisson GINAR models. This would allow for models incorporating seasonal variations, for example. Here we did not allow for correlations between sequences. The low correlation within sequences, as well as the small likelihood of a tiring effect across trials, made independence between sequences most likely.

The methods may be extended in many directions. For example, one may have data from several experiments, each using a different neuron. In this case, one may analyse each experiment in a similar fashion to the analyses presented here, perhaps adding hierarchical terms in the model to describe the distributions of neuron firing rates among a population of neurons.

ACKNOWLEDGEMENTS

The data analysed here were provided by Daniel Commenges of INSERM, France. This work was supported in part by the Natural Science and Engineering Council of Canada and the Fonds pour la Formation de chercheurs et l'aide à la Recherche, Gouvernement du Québec. Lawrence Joseph is a research scholar supported by the Fonds de la Recherche en Santé du Québec.

RÉSUMÉ

Dans de nombreuses expériences médicales, les données sont recueillies chronologiquement, par plusieurs essais similaires, ou plusieurs unités expérimentales. Comme dans le cas des études sur les trains de décharges neuronales. Ces données peuvent être sous la forme de nombres d'événements par unité de temps. Ces décomptes peuvent être corrélés au sein de chaque essai. Il est souvent intéressant de savoir si une intervention telle que l'application d'un stimulus, modifie la distribution des décomptes pour l'expérience. Dans de telles recherches, chaque essai engendre une suite de données qui peut ou non inclure un changement de distribution à une certaine date. Chaque suite de décomptes de nombres entiers peut être vue comme provenant d'un processus de Poisson et alors indépendantes, ou bien comme série chronologique à valeurs entières prenant en compte les corrélations entre ces décomptes. Le but principal de cet article est de montrer comment on peut utiliser un ensemble de trajectoires pour développer une inférence sur la distribution des temps de changement pour une population donnée. Ceci est réalisé par un modèle bayésien hiérarchique pour des temps de changement. Un intérêt supplémentaire de ces modèles est qu'ils permettent un développement inférentiel sur la probabilité de changement en chaque point, et sur l'amplitude éventuelle des effets. L'utilisation de tels modèles à point de changement pour des séries temporelles à valeurs entières est illustrée ici par des données sur les trains de décharge neuronale bien que ces méthodes s'appliquent dès que des processus à valeurs entières sont impliqués.

REFERENCES

- Al Osh, M. and Alzaid, A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* **8**, 261–275.
- Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: Explanation of the large variability in evoked cortical response. *Science* **273**, 1868–1871.
- Bélisle, P., Joseph, L., MacGibbon, B., Wolfson, D., and du Berger, R. (1997). *Change-point analysis of neuron spike train data*. GERAD Technical Report G-97-31, L'École des Hautes Études Commerciales de Montréal, Montreal.
- Brillinger, D. R. (1992). Nerve cell spike train data analysis: A progression of technique. *Journal of the American Statistical Association* **87**, 260–271.
- Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics* **41**, 389–405.
- Commenges, D. and Seal, J. (1985). The analysis of neuronal discharge sequences: Change-point estimation and comparison of variances. *Statistics in Medicine* **4**, 91–104.
- Commenges, D., Seal, J., and Pinatel, F. (1986). Inference about a change point in experimental neurophysiology. *Mathematical Biosciences* **80**, 81–108.
- DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.
- Gauthier, G. and Latour, A. (1994). Convergence fort des estimateurs des paramètres d'un processus GENAR(p). *Annals des Mathématiques du Québec* **18**, 49–71.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.
- Hanes, D. P. and Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science* **274**, 427–429.
- Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1–16.
- Hinkley, D. V. and Hinkley, E. A. (1970). Inference about the change-point in a sequence of binomial random variables. *Biometrika* **57**, 477–488.
- Jin-Guan, D. and Yuan, L. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis* **12**, 129–142.
- Joseph, L. and Wolfson, D. B. (1993). Maximum likelihood estimation in the multi-path change-point problem. *Annals of the Institute of Statistical Mathematics* **45**, 511–530.
- Joseph, L., Wolfson, D. B., du Berger, R., and Lyle, R. M. (1996a). A change-point analysis of a randomized trial on the effects of calcium supplementation on blood pressure. In *Bayesian Biostatistics*, D. Berry and K. Stangl (eds), 617–649. Marcel Dekker: New York.
- Joseph, L., Vandal, A., and Wolfson, D. B. (1996b). Estimation in the multi-path change-point problem for correlated data. *Canadian Journal of Statistics* **24**, 37–53.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.
- Koch, C. (1997). Computation and the single neuron. *Nature* **385**, 207–210.
- Lange, N., Carlin, B. P., and Gelfand, A. E. (1992). Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers (with discussion). *Journal of the American Statistical Association* **87**, 615–632.
- Müller, P. (1991). *A generic approach to posterior integration and Gibbs sampling*. Technical Report 91-09, Department of Statistics, Purdue University, West Lafayette, Indiana.
- Pettitt, A. N. (1979). A nonparametric approach to the change-point problem. *Applied Statistics* **28**, 126–135.
- Raftery, A. and Lewis, S. (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics*, Volume 4, J. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds), 763–773. Oxford: Oxford University Press.
- Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Smith, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika* **62**, 407–416.
- Steutel, F. W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability* **7**, 893–899.

- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior densities by data augmentation (with discussion.) *Journal of the American Statistical Association* **82**, 528–550.
- Wegman, E. J. and Habib, M. K. (1992). Stochastic methods for neural systems. *Journal of Statistical Planning and Inference* **33**, 5–25.
- Yao, Y.-C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Annals of Statistics* **15**, 1321–1328.
- Yao, Y.-C. (1990). On the asymptotic behavior of a class of nonparametric tests for a change-point problem. *Statistics and Probability Letters* **9**, 173–177.

Received March 1994; revised February and June 1997; accepted June 1997.