Statistica Sinica **7**(1997), 687-703

ANALYSIS OF PANEL DATA WITH CHANGE-POINTS

Lawrence Joseph*+, David B. Wolfson+, Roxane du Berger* and Roseann M. Lyle

Montreal General Hospital, +McGill University and Purdue University

Abstract: In many experiments data are collected over time or space, on a number of subjects or sites. In medical experiments, for example, it is often of interest to know if the introduction of an intervention, such as the administration of a drug, affects the distribution of a certain variable recorded several times over the course of the trial. In such investigations, each patient generates a sequence of data which may or may not contain a change in distribution at some point in time. Different subjects within a given population may react differently to the intervention. Each sequence can be viewed as a sample path from a stochastic process. 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 times or locations of change.

Key words and phrases: Bayesian inference, change-point, clinical trials, Gibbs sampler, panel data.

1. Introduction

Often in a medical experiment, data on a number of subjects are collected on several occasions during the trial. If a drug is introduced during the trial that may affect the distribution of the variable under study, then the following questions are of main interest:

1. What proportion of the target population would respond to the intervention?
2. What is the distribution of the magnitude of the effect among those who do respond?
3. Among responders, what is the distribution of time to response?

The latter question is especially relevant if a delay of unknown or variable duration is expected from the time the drug is administered to the time it takes effect, such as may be the case in a trial of a new drug designed to lower cholesterol or blood pressure. Standard methods, such as a repeated measures analysis of variance, are not suited to answering these questions. For example, in a clinical trial, there is rarely an attempt to classify individual patients into "responders" or "non-responders"; instead one looks for significant mean differences between treatment groups. In addition, it is usually required that the time of change be specified in advance, coinciding with the commencement of the intervention and assumed identical for all subjects. Often, little is known about the reaction time which may vary from subject to subject. By not taking into account a delay to

response, the mean treatment effect may be diluted, thus leading to a considerable loss of power. Valid interpretations of reported mean effects are only possible after estimating the proportion of positive responders in the population as well as the distribution of these delays. This paper presents a multi-path change-point model and illustrates how it can be used to address these issues. The model also facilitates the comparison of baseline measurements to post-intervention values within each individual, eliminating the need to explicitly model the effects of baseline means.

The model below is also applicable to different areas such as soil science and geology. Soil profiles change with increasing depth and frequently these changes occur quite abruptly because of the way sedimentary beds were laid down. Soil profiles are not only important in providing a geological history of a region but also directly affect the rate at which water is absorbed and retained, and are a crucial component of petrology and mineral exploration. Data collected at different sites and at several depths per site may be used to fit the type of model described in Section 2 and hence to address the questions of interest briefly referred to above.

We motivate the model by outlining two examples where the methodology can be effectively employed. One example consists of Poisson data and the other of Normal data. After details of the model are provided in Section 2, these examples are revisited and analysed in Section 4. Section 3 describes the inferential procedures and the final section contains further discussion.

1.1. Health effects of urea formaldehyde foam insulation

Urea-formaldehyde foam insulation (UFFI) was installed in many homes in Canada until it was banned by the Federal Government on December 18, 1980. The decision to ban UFFI was largely based on preliminary results mostly from studies on rats (Albert et al. (1982)) that alluded to its toxic effects. Following reports in the press, residents began to complain to their physicians of symptoms purportedly related to UFFI, resulting in heated debate over whether or not these concerns were justified. One possible objective indicator of the danger posed by UFFI would be an increase in the rate at which household occupants visit a doctor after installation compared with before. Of course, even if there is an increase, one would not necessarily expect an instant reaction to the substance, as the vapor slowly seeps out from between the walls of a home and into the ambient air. Furthermore, it is suspected that not everyone would react, with certain people being hypersensitive, others having no reaction, and the rest being only mildly reactive.

In Section 4, tri-monthly data on the number of visits to a doctor for one year before and after installation of the foam in the homes of 285 persons in Quebec,

Canada (L'Abbé (1984)) are analyzed via the change-point model introduced in Section 2. Of interest here is whether there is evidence for an increase in the visit rate after installation, and if so, the proportion of exposed subjects who would experience an increase. Also of importance are the magnitude of the increase and the length of time after installation that it occurs.

1.2. The effect of calcium on blood pressure measurement

A double-blinded placebo controlled randomized trial evaluating the effects of calcium supplementation on blood pressure measurements of normotensive men aged 19 to 52 is reported by Lyle et al. (1987). The drug (tablets containing 500 mg of elemental calcium) group consisted of 37 subjects, with 38 subjects randomized to a control group (placebo tablets). Blood pressure readings on all patients were taken a total of 10 times, 4 weekly readings during a baseline period, and 6 readings, one every two weeks, after randomization. A repeated measures analysis of variance showed statistically significant mean decreases in blood pressure in the calcium group of 2 to 3 mm Hg, which are less than the 5 mm Hg differences usually required for clinical significance. However, delays in the onset of treatment effects could have diluted the mean decreases. This is because the mechanism by which calcium may lower blood pressure is not known, so that it is plausible that blood pressure decline does not occur immediately but rather the concentration of calcium in the body slowly increases, and the metabolism adjusts to the new levels. Also, the mean responses are difficult to interpret, since it is possible that nonresponders may have contributed to the overall means, and that the effects in some subjects may be higher and in the clinically interesting range.

A multi-path change-point analysis is pertinent here, since not all patients may respond, and the reaction times may be different among those who do react. As will be shown below, the analysis proposed allows for comparisons of the proportions who experience a decrease in blood pressure in each group (which will likely be small in the placebo group), as well as the time to reaction and magnitude of the decrease in those who respond. Information concerning minimally clinically interesting differences can also be incorporated into the analysis in the form of a prior distribution, so that individuals changing less than this amount are counted as nonresponders.

2. The Model

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

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1\tau_1} & X_{1\tau_{1+1}} & \cdots & X_{1N} \\ X_{21} & X_{22} & \cdots & X_{2\tau_2} & X_{2\tau_{2+1}} & \cdots & X_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ X_{M1} & X_{M2} & \cdots & X_{M\tau_M} & X_{M\tau_{M+1}} & \cdots & X_{MN} \end{pmatrix}. \tag{1}$$

Each sequence, $X_{i1}, \ldots, X_{iN}$, represents observations over time from the ith subject, $i = 1, \ldots, M$. A change-point is said to have occurred at $T_i = \tau_i$ in sequence or row i,$i = 1, \ldots, M$, and $1 \leq \tau_i \leq N-1$, if $X_{i1}, \ldots, X_{i\tau_i}$, are identically distributed with common distribution F_{i_1}, which is different from the common distribution, F_{i_2} of $X_{i\tau_i+1}, \ldots, X_{iN}$. If $T_i = N$, then no change has occurred in row i. The distribution of the points of change, τ_i, and unknown parameters of the distributions $F_{i_k}, i = 1, \ldots, M, k = 1, 2$, are to be estimated from the data matrix (1). When there are multiple paths or sequences, we shall refer to a "multi-path"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 $g(t) = Pr\{T_i = t\}, i = 1, \ldots, M, t = 1, \ldots, N$, 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 individual in the population.

The case $M = 1$ has received considerable attention in the literature. Among early approaches, Hinkley (1970) and Hinkley and Hinkley (1970) found the asymptotic distribution (as $N \longrightarrow \infty$ and $(N - \tau) \longrightarrow \infty$) of the maximum likelihood estimator in the case of Normal and binomial data, respectively. Yao (1987, 1990) and many others have also more recently investigated maximum likelihood change-point inference. Pettitt (1979) offered a non-parametric solution to test for a change-point, while Smith (1975) takes a Bayesian approach. More recently, Carlin, Gelfand and Smith (1992) have used the Gibbs sampler to find marginal posterior distributions in a hierarchical single-path change-point model. As we have seen above, introducing $M \geq 2$ considerably broadens the applicability of change-point models.

The multi-path extension introduces several interesting theoretical questions. While it has been shown by Hinkley (1970) that the single-path maximum likelihood estimator of the change-point is not consistent, the nonparametric estimator of $g(\cdot)$ has been shown in Joseph and Wolfson (1993) and Joseph, Vandal and Wolfson (1996) to be consistent under certain conditions in the multi-path case. Further, two powerful statistical techniques that have not been effectively employed (although, see Hinkley and Schechtman (1987) for a conditional bootstrap approach) in the single-path context are the bootstrap (Efron (1979, 1982)) and empirical Bayes methods (Robbins (1964)). It has been shown in Joseph and Wolfson (1992) that both of these techniques may be utilized in the multi-path context.

From a practical point of view, the implementation of a multi-path change-point model offers substantial challenges. Assuming only one unknown parameter from each F_{i_k}, a Bayesian approach involves the calculation of posterior distributions of $2 \times M + N$ inter-related parameters. Similarly, finding the maximum likelihood estimators even in problems of moderate dimension requires a considerable and often infeasible number of computations. However, recent advances in statistical computing algorithms such as the EM algorithm (Dempster, Laird and Rubin (1977)) or data augmentation (Tanner and Wong (1987)) and the closely related Gibbs sampler (Geman and Geman (1984), Gelfand and Smith (1990) Gelfand et al. (1990)) have, to a considerable extent, alleviated these difficulties.

2.1. Alternative modelling strategies

In the sequel, it is assumed that $X_{ij},\ j = 1, \ldots, N$ are independent for each fixed $i = 1, \ldots, M$. In certain examples, including some of those described in Section 1, it can be argued that within sequence observations may be correlated, and should be modelled accordingly (Joseph, Vandal and Wolfson (1996)). However, if the measurements are sufficiently separated in time or space or represent long term averages or rates, the correlation may be weak. In such cases an assumption of independence may provide a much simpler framework within which to carry out a change-point analysis without seriously compromising the results. If a Bayesian time series approach is required, one can adapt the methods of Marriot et al. (1992) to the change-point situation, where again the Gibbs sampler may be employed to estimate the parameters of the model. Nevertheless, even the simplest such models would necessitate an increase in the number of parameters to estimate. With short data sequences, it may be difficult or impossible to reliably estimate all of the unknown parameters, particularly since there may be different time series parameters before and after any change-points, and since these time series parameters may differ from patient to patient. For the same reason, it would be difficult to test whether a time series model is supported by the data, or whether a simpler model that assumes independence is sufficient. This view is supported by Henderson (1986) who suggests that ignoring weak correlations may have little effect on estimating change-points.

Change-point models assume abrupt changes in parameter values, while it may be more realistic in some situations to assume gradual changes. For example, one may investigate a tri-linear regression model with two change-points, indicating the beginning and end of a transition period. A single change-point model is simpler to implement, and again, it may not be feasible to reliably estimate all parameters of a tri-linear regression model for data sets with a limited number of observations per sequence. The simpler approach taken in this paper is conservative, in the sense that estimated pre- and post-treatment mean differences may be attenuated if the true change is gradual. This phenomenon will

occur whenever measurements are taken during the period of gradual change, but falsely included in either the pre-or post-treatment mean estimate. The change-point will be estimated to occur at the time point that best divides the data into two sets each with apparently identically distributed variables. Whether such a change will be found will depend on specific features of the data, such as the magnitude and the rate of change, and how many observations there are for each subject. Growth curves could also be used, but in order to answer the specific questions listed in Section 1, one would need to estimate where the derivative of the curve changes. This is a complex problem, and again reliable estimates may require much more data than does our simpler model.

On the other hand, more parsimonious models could also be proposed. For example, rather than allowing subject specific parameters for the before and after change-point distributions F_1 and F_2, one could assume a random effects model with a common distribution for the parameters associated with different subjects. In cases with a high degree of homogeneity so that the specification of this distribution can be reasonably made, this approach may be preferable; as is well known a reduction in mean square error is the benefit from being able to "borrow strength" from all the paths. However, the great between subject variability with respect to the before-and-after-change parameters in the two data sets under discussion in Section 4, suggest that these parameters are not identically distributed. Even if they are identically distributed the benefit of a reduction in mean-square error, when the between subject variance is large, could hide within subject changes in the parameters. This would happen if these changes are smaller than the between subject differences and do not occur for all subjects; shrinkage towards the prior mean averages out those paths with changes and those without. Consequently, random effects models were not considered for these parameters.

It is more reasonable, nevertheless, to use a random effects model for the change-point, since the concern here is not with before and after effects that could be hidden by "shrinkage". One may initially choose to model the change-point distribution g. For example, a "tent-shape" for g may be assumed if it is suspected that there is a change-point location that is most likely, with points further from this location having progressively decreasing probability. However, if there is a strong possibility that some sequences have no change-points, no simple shape may be sufficient to adequately represent g. We have therefore chosen to modify the random effects model by imposing a hierarchical model on the change-point distribution; as is well known such models are more robust against misspecification of the prior. Another simplification that may be appropriate in some situations is to consider a single common change-point for all subjects. This model was investigated by Joseph and Wolfson (1992), where it was called

a "fixed-τ" change-point model. Again, if some subjects may change while other may not, fixed-τ models are ruled out.

Bayesian hypothesis testing for a change-point could also be performed. For example, a prior distribution on g that assigns half the probability to $g(N)$ and the rest uniformly distributed across $g(i)$, $i = 1, 2, \ldots, N-1$, seems appropriate, since half of the prior probability would be placed on the null hypothesis of no change. The posterior probabilities for $g(N)$ would then indicate whether the data support a change-point.

Throughout this paper, $f(\cdot)$ will be used to generically denote a probability density or probability function, and $F(\cdot)$ will denote a cumulative distribution function. The random variables to which these distributions refer will be clear from their arguments and the context in which they appear. Where concreteness is desirable, motivated by the first example in Section 4, the F_{i_k} are assumed to be Poisson. The techniques easily carry over to many other distributions including the Normal, which is used for the second example. Müller (1991) demonstrates that Gibbs sampling can be combined with a Metropolis algorithm to generalize applicability of the techniques to virtually any distribution.

3. Estimation of Parameters via the Gibbs Sampler

The likelihood for the model described in Section 2 is given by

$$f(x|\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, \underset{\sim}{\pi}) = \prod_{i=1}^{M} \sum_{h=1}^{N} \Big\{ \prod_{j=1}^{h} f_1(x_{ij}|\underset{\sim}{\theta}_1) \Big\} \Big\{ \prod_{j=h+1}^{N} f_2(x_{ij}|\underset{\sim}{\theta}_2) \Big\} \pi_h, \tag{2}$$

where $\underset{\sim}{\theta}_1$ and $\underset{\sim}{\theta}_2$, possibly vector valued, are the parameters of the densities f_1 and f_2 respectively, and $\underset{\sim}{\pi} = (\pi_1, \ldots, \pi_N)$, where $\pi_i = Pr\{T_i = \tau_i\}$, $i = 1, \ldots, N$. Inference using this likelihood is difficult since it takes the form of a mixture. However, conditional on knowledge of "latent data" (Tanner and Wong (1987)) τ_i, $i = 1, \ldots, M$, the change-points in each data sequence, the likelihood simplifies to

$$f(x|\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, \tau_1, \ldots, \tau_M) = \prod_{i=1}^{M} \Big\{ \prod_{j=1}^{\tau_i} f_1(x_{ij}|\underset{\sim}{\theta}_1) \Big\} \Big\{ \prod_{i=\tau_i+1}^{N} f_2(x_{ij}|\underset{\sim}{\theta}_2) \Big\}. \tag{3}$$

When x_{ij} follows a Poisson distribution, the parameters in the model are:

1. $\underset{\sim}{\theta}_1 = \underset{\sim}{\lambda}_1 = (\lambda_{11}, \ldots, \lambda_{M1})$ and $\underset{\sim}{\theta}_2 = \underset{\sim}{\lambda}_2 = (\lambda_{12}, \ldots, \lambda_{M2})$, vectors of the means of the Poisson distributions before and after the change-point in each row
2. $\underset{\sim}{\pi} = (\pi_1, \ldots, \pi_N)$, the multinomial probabilities that a change occurs at position i in each row, $i = 1, \ldots, M$
3. $\underset{\sim}{\tau} = (\tau_1, \ldots, \tau_M)$, the unobserved latent data representing the change-points in each row.

The uncertainty in these parameter values is reflected in the choice of their prior distributions. For simplicity, one may choose conjugate prior distributions, although non-conjugate priors can also be accommodated (Müller (1991)). Since the Gamma distributions are conjugate priors for a Poisson random variable, and the Dirichlet distributions form a conjugate family for the parameters of a multinomial random variable (see, for example, DeGroot (1970), Chapter 9), the priors in the case that the X_{ij} follow Poisson distributions could be given as

$$f(\lambda_{ik}) = \frac{1}{\Gamma(a_{ik})b_{ik}^{a_{ik}}}\lambda_{ik}^{a_{ik}-1}\exp\Big(-\frac{\lambda_{ik}}{b_{ik}}\Big), \quad i=1,\ldots,M, k=1,2 \tag{4}$$

and

$$f(\pi_1,\ldots,\pi_N) = \frac{\Gamma(\alpha_0)}{\prod_{l=1}^{N}\Gamma(\alpha_l)}\prod_{i=1}^{N}\pi_i^{\alpha_i-1}, \tag{5}$$

where $\alpha_0 = \sum_{i=1}^{N}\alpha_i,\ \alpha_i > 0$, $i = 1,\ldots,N$, and where 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 below, following standard procedures for conjugate analyses by DeGroot (1970). Note that the full conditional distribution of each parameter does not always depend on all of the other parameters, which leads to some further simplifications:

$$f(\lambda_{i1}|X,\tau_i) \sim Gamma\Big(a_{i1}+\sum_{j=1}^{\tau_i}x_{ij}, (\tau_i+\frac{1}{b_{i1}})^{-1}\Big) \tag{6}$$

$$f(\lambda_{i2}|X,\tau_i) \sim Gamma\Big(a_{i2}+\sum_{j=\tau_{i+1}}^{N}x_{ij}, (N-\tau_i+\frac{1}{b_{i2}})^{-1}\Big) \tag{7}$$

$$\begin{aligned} &Pr\{\tau_i = t|\underset{\sim}{\lambda}_1,\underset{\sim}{\lambda}_2,\underset{\sim}{\pi},x) \\ &= \frac{\Big\{\prod_{j=1}^{t}\frac{(\lambda_{i1})^{x_{ij}}\exp(-\lambda_{i1})}{x_{ij}!}\Big\}\Big\{\prod_{j=t+1}^{N}\frac{(\lambda_{i2})^{x_{ij}}\exp(-\lambda_{i2})}{x_{ij}!}\Big\}\pi_t}{\sum_{k=1}^{N}\Big\{\prod_{j=1}^{k}\frac{(\lambda_{i1})^{x_{ij}}\exp(-\lambda_{i1})}{x_{ij}!}\Big\}\Big\{\prod_{j=k+1}^{N}\frac{(\lambda_{i2})^{x_{ij}}\exp(-\lambda_{i2})}{x_{ij}!}\Big\}\pi_k} \end{aligned} \tag{8}$$

$$f(\underset{\sim}{\pi}|\underset{\sim}{\tau}) \sim Dirichlet(\underset{\sim}{\alpha}'), \tag{9}$$

where α_k', the kth element of $\underset{\sim}{\alpha}'$ is given by $\alpha_k + \sum_{i=1}^{M} I_{\{\tau_i=k\}}$, and where $I_{\{y\}}$ is the indicator function for the set $\{y\}$.

The Gibbs sampler algorithm proceeds as follows: Starting from arbitrary initial values, a random sample is drawn from each full conditional distribution

(6)-(9) in turn. The parameters drawn from previous iterations are used in the conditional distribution for subsequent iterations. A cycle is completed when each conditional distribution has been sampled from, and the cycle is repeated a large number of times. The random variables thus generated at the end of each full cycle can be regarded as a random sample from the correct joint posterior distribution, and hence any subset of them as a random sample from the corresponding marginal posterior distribution.

The approach taken here is summarized by the steps given below:

1. Four independent Gibbs sequences of identical length were generated from different starting values using different seeds for the random number generators.
2. Selected summary statistics (typically the means of all marginal posterior distributions for the change-point locations and selected percentiles for a random sample of parameters of the other posterior distributions) were plotted against iteration number to monitor convergence for all four runs.
3. If convergence seemed likely, as evidenced by stabilization of the above quantities after a certain number of iterations, marginal posterior density estimates were generated by the Rao-Blackwell method, as proposed by Gelfand and Smith (1990). Plots of these densities were constructed and overlaid to check for any differences, using the "thick felt-tip pen test" as suggested by Gelfand and Smith (1990). Convergence was assumed only if all marginal densities were identical for all practical purposes. In particular, each marginal posterior change-point location mean probability had to be within 0.01 of the overall mean probability from the four runs, to be declared convergent. If convergence had not been attained, steps 1 and 2 were repeated wherein the Gibbs sample was run for a larger number of iterations.

More details are provided in the next section when methods deviated from those given above.

4. Examples

In this section the methods are applied to the two data sets introduced in Section 1. For brevity we present only a portion of the available output from each example, highlighting the benefits that can be derived from a multi-path change-point analysis.

The output from the Gibbs sampler can be used in a variety of ways:

1. The marginal posterior density of any component of the parameters of F_{i1} or F_{i2} can be approximated. For example, in the case that F_{i1} is Poisson, the marginal posterior density of the Poisson parameter λ_{i1} can be approximated as a Rao-Blackwell average of Gamma densities of the form (6).

2. Summary statistics of the above posterior marginal densities such as the marginal means or medians of $\lambda_{i2} - \lambda_{i1}$ may be easily computed and graphically displayed.
3. The output produced from equation (9) for $\underset{\sim}{\pi}$ is a sample from a Dirichlet distribution in N dimensions. Since this distribution is difficult to visualize, summary statistics are important.
   (a) Bar graphs of the means of the marginal Dirichlet posterior distributions can be constructed.
   (b) Posterior marginal densities for selected change-point probabilities may be plotted. They display the variability about the above 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.
4. Each subject in each iteration may have $\tau_i < N$, or $\tau_i = N$. Another interesting statistic is then $\frac{\{\#\text{ times } \tau_i < N\}}{\#\text{ iterations}}$. This approximates the subject specific probability of change.

4.1. Urea formaldehyde foam insulation

Data consisting of the number of physician visits by $M = 285$ patients were collected in tri-monthly intervals over a two year period, resulting in $N = 8$ observations per patient. Four of these observations were taken before the installation of UFFI, the rest, afterwards.

A Poisson distribution was assumed for the counts of visits in each three month period. The Poisson parameters, which represent the visit rates before and after installation of UFFI, were not assumed equal from sequence to sequence. In other words, different persons were allowed to have different underlying visit rates, both prior to and after the installation.

Fifteen of these patients had at least one and up to three missing data values. A straightforward extension of the methods presented in Section 3 is to impute the missing values in addition to estimating the usual parameters of the model. This requires one additional step to the algorithm, where before calculation of (8), any missing data points x_{ij} are filled in by sampling from the current estimate of $f(x_{ij}|\tau_i, \lambda_{i1}, \lambda_{i2})$.

The parameters for the Dirichlet prior were $\alpha_1 = \alpha_2 = \alpha_3 = 0$, $\alpha_4 = \alpha_5 = \cdots = \alpha_8 = 1$, i.e., a uniform prior over all possible times of change after the installation of UFFI. The sum of Dirichlet prior parameters, here equal to five, can be viewed as a measure of certainty in the prior distribution, with higher values indicating more confidence. The ratio of this sum to the same sum plus the number of sequences, M, provides a guide to the proportion of information in the posterior distribution obtained from the prior. Here, $5/(285 + 5) = 0.017$

indicates that almost all of the posterior change-point distribution arises from the data. For the visit rates, Gamma($\alpha = 1, \beta = 15$) distributions were selected as approximately locally uniform priors, covering the range of average visit rates across all participants. This distribution in fact slopes down gently from its maximum at zero, which gives slightly higher weight to lower values of the visit rate, which matches what was expected.

Runs of $10,000$ random variables per unknown parameter were performed, the last $3,000$ of which were used for inference and density estimation. Figure 1 contains a bar graph of the posterior marginal means of the components of $\underset{\sim}{\pi}$. The mean probability of no change is estimated to be 0.53, and for those participants that do change, it occurs approximately equally at three, six, nine and twelve months after installation. Of course, an increase in the visit rate is not necessarily due to UFFI, and there may be minor effects that do not precipitate a visit to a doctor, or more serious effects that surface only after one year of exposure.

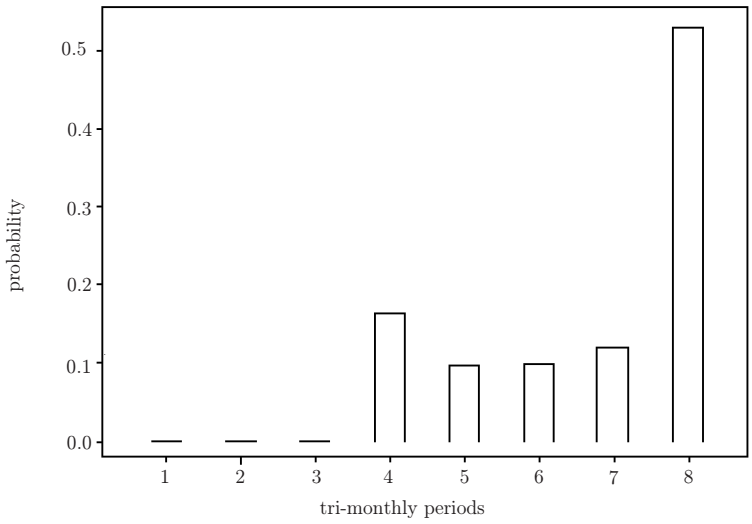

Figure 1. Means of the marginal Dirichlet posterior distributions for the location of the change-point for the UFFI data. Tri-monthly period number 8 indicates no change.

The marginal distribution for the probability of no change (8th tri-monthly period) and the probability of a change at the 4th tri-monthly period are given in Figure 2. They are useful for assessing the uncertainty in the point estimates given in Figure 1. In this case, for example, most of the mass of the probability of no change is concentrated in the interval from 0.43 to 0.63.

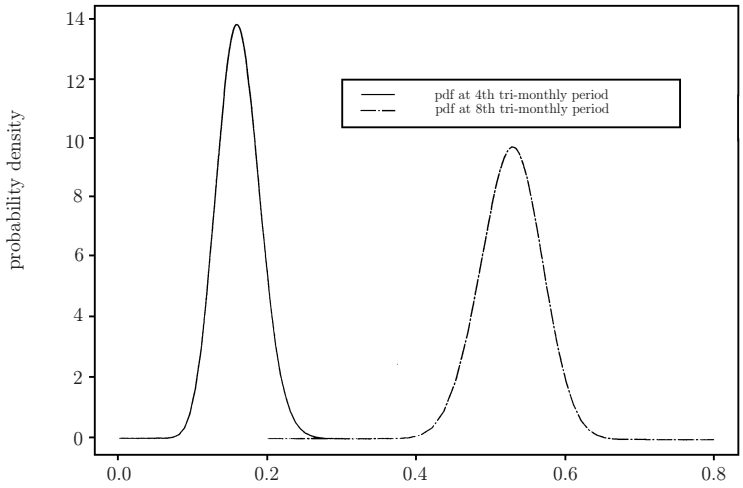

Figure 2. Posterior probability densities for the fourth and eighth (no change) tri-monthly periods, for the UFFI data.

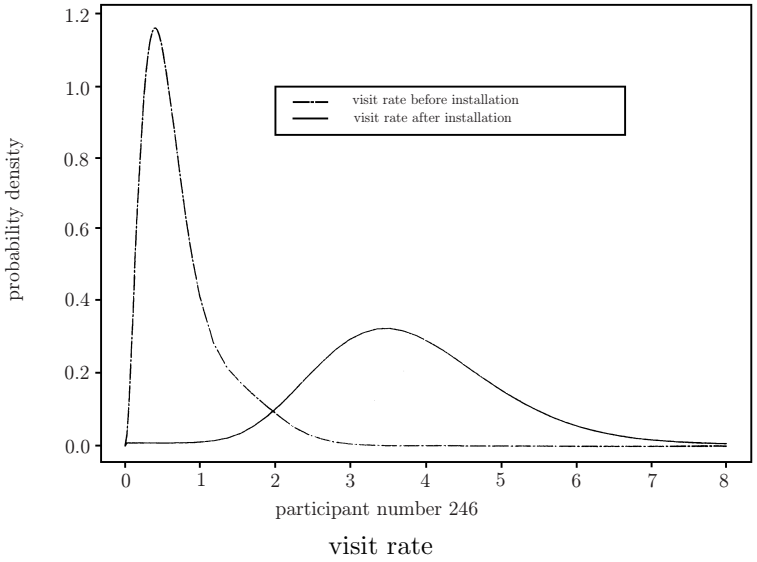

Figure 3. Posterior distribution of the pre and post change Poisson parameters for the visit rate for participant number 246.

A sample graph showing the difference in the posterior visit rates for participant #246, who does appear to experience a change in visit rate (patient specific probability of a change = 0.90), is shown in Figure 3. This picture was typical for participants with clear changes. The algorithm produces similar graphics for all participants, or, more concisely for large data sets, summary measures of the

differences across all M sequences, such as means, variances, histograms or boxplots. The above quantify the qualitative results of other research, which suggest that the effects of UFFI appear to be small, and may only occur in a proportion of occupants of homes with the material (L'Abbé 1984). The above analysis provides an estimate of this proportion, and the distribution of the magnitudes of the changes.

4.2. Calcium and blood pressure

Lyle et al. (1987) reported a double-blind randomized controlled trial comparing the effects of calcium supplementation with placebo on the blood pressures of men aged 19 to 52. Four baseline weekly blood pressure readings and six biweekly readings during treatment resulted in $N = 10$ measurements, on $M_1 = 37$ calcium patients and $M_2 = 38$ control patients.

Seated systolic blood pressure measurements are discussed here in detail, although supine and diastolic pressures were also recorded. A Dirichlet prior with $\alpha_1 = \alpha_2 = \alpha_3 = 0$ and $\alpha_4 = \alpha_6 = \alpha_8 = \cdots = \alpha_{16} = 0.1$ was used for the change-point distribution in each group. Normal/Gamma priors with equal pre- and post-change means equal to 115 $mm\ Hg$ were used for prior distributions on the blood pressures. The mean of the prior variance was taken to be 25.

Runs of length $10,000$ random variables were produced for each parameter, with $7,000$ retained for inference. The posterior marginal means of the probability of change at each time period after baseline is given in Figure 4, and boxplots of the mean differences in (after-before) blood pressures are presented in Figure 5. The posterior probability of no change in blood pressure is estimated to be about 0.71 and 0.87 in the calcium and placebo groups respectively. Of the 29% in the calcium group who do appear to experience an effect, 17% appear to change within the first two weeks of administration. Figure 5 suggests that virtually all calcium subjects who change, experience a decrease in blood pressure, with close to half in the clinical range. The figure also suggests a possible placebo effect of only slightly lesser magnitude. Nevertheless, the overall mean decrease was 2 $mm\ Hg$ larger in the calcium compared to placebo. It is also interesting to estimate $\mathcal{P} = Pr\{\pi_{N_C} < \pi_{N_P}\}$, which represents the probability that subjects on calcium (C) supplementation are more likely to change than placebo subjects (P). This probability can be estimated using a Mann-Whitney-type statistic, $\hat{\mathcal{P}} = (\sum_{i=1}^{m}\sum_{j=1}^{m} I_{\{\pi_{N_{C_i}} < \pi_{N_{P_j}}\}})/m^2$, where $m = 7,000$ is the total number of Gibbs iterates generated, $\pi_{N_{C_i}}$ and $\pi_{N_{P_j}}$ represent the generated values for π_N in the two groups, $i, j = 1, \ldots, m$, and $I\{y\}$ is the indicator function for the set y. Here $\hat{\mathcal{P}} = 0.82$ indicates a relatively high probability that a randomly selected individual on calcium will be more likely to change than a similarly chosen subject given a placebo.

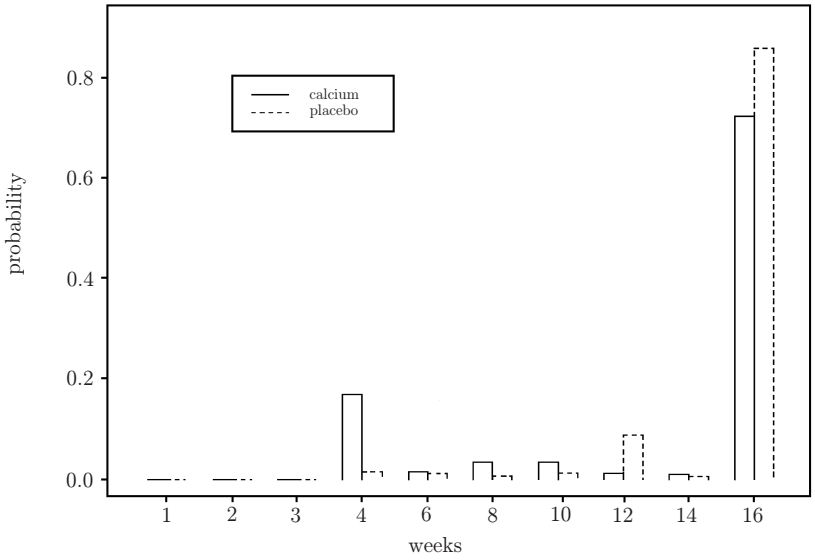

Figure 4. Means of the marginal Dirichlet posterior distributions for the location of the change-point for both calcium and placebo groups for the blood pressure data. Week 16 indicates no change.

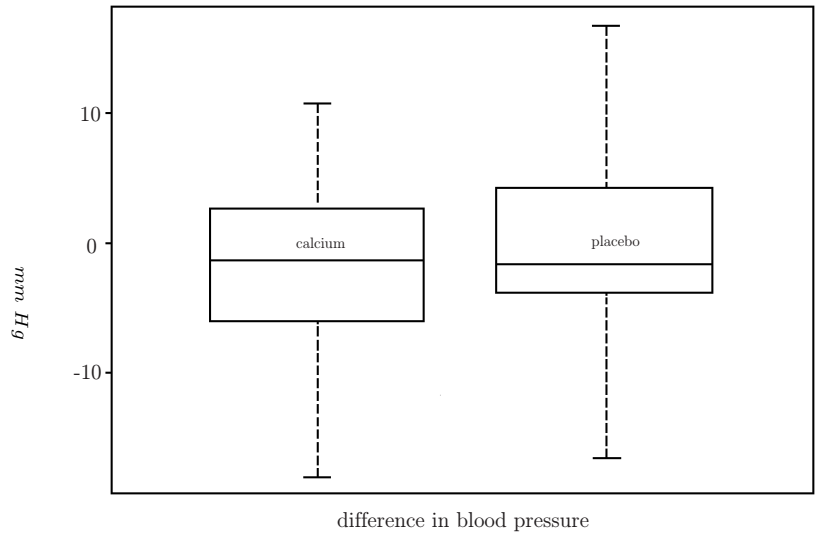

Figure 5. Boxplots of the mean blood pressure differences (after-before) for the calcium and placebo groups for the blood pressure data.

These results agree with the repeated measures analysis of Lyle et al. (1987), which found a small but statistically significant average decrease in seated systolic

blood pressure in the calcium group compared to the placebo. The multi-path change-point analysis presented here provides additional information concerning the proportion of responders, and the distribution of the before and after blood pressures in each group, as well as information concerning the timing of the effect.

An induced change of 5 $mm\ Hg$ is a commonly required minimum for a treatment to lower blood pressure to be considered clinically meaningful. An identical analysis to that described above was run, with the exception that the mean of the prior distribution for the post-change blood pressures for each patient was increased by 5 $mm\ Hg$. In this way, patients experiencing only small changes would be considered as non-responders in the analysis. The marginal posterior probability of no change increased to 0.79 in the calcium group, and 0.94 in the placebo group.

5. Discussion

Multi-path change-point methods could be used along with standard methods of analysing repeated measures panel data. They offer estimates of the proportion of subjects in a given population who respond to the treatment, as well as the magnitude of and time delay to the change.

Throughout the analyses presented here, non-informative prior distributions were used for the change-point parameters, and nearly non-informative prior distributions (in the sense that the distributions were approximately flat over the range of likely values) were used for all other parameters. Of course, one of the principle advantages of a Bayesian approach is the opportunity to formally incorporate prior information into the inferential procedure. In many situations, reporting posterior distributions arising from a range of prior distributions is desirable (Hughes 1993).

The methods may be extended in many directions. When the data from the ith subject are allowed to be dependent, it is possible to account for seasonal variation. Similar methods may also be applied to changes in hazard parameters in a survival analysis, extending the work of Zelterman et al. (1994) and Liang, Self and Liu (1990). Work on different structures for prior information, such as dependences between the multinomial probabilities and the Poisson rates would be interesting. Also in the present formulation, the data sequences need not arise from the same family of distributions, as only a common distribution for the change-point is assumed. If different subgroups of a population are expected to possibly have different change-point distributions, then they may be analysed separately.

Fortran software is available that implements the methods in this paper. Send the email message "send mpcpp from general" or "send mpcpn from general" to StatLib@lib.stat.cmu.edu to receive the software for Poisson and normal

data, respectively. Running this software on a Sun SPARCstation ELC, generating 10,000 iterations from the Gibbs sampler algorithm took approximately 10 minutes.

Acknowledgements

The authors wish to thank Marie-Pierre Aoun for her skills in preparing this document. The data for the urea formaldehyde foam insulation example were provided by Dr. John Hoey, of the Department of Epidemiology and Biostatistics, McGill University. 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.

References

Albert, R. E., Sellakumar, A. R., Laskin, S., Kuschner, M., Nelson, N. and Snyder, C. A. (1982). Gaseous formaldehyde and hydrogen chloride induction of nasal cancer in the rat. *J. Natl. Cancer. Inst.* **68**, 597-602.

L'Abbé, K. (1984). *Health Effects of Urea-Formaldehyde Foam Insulation.* Masters Thesis, Department of Epidemiology and Biostatistics, McGill University.

Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Appl. Statist.* **41**, 389-405.

DeGroot, M. H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. *Soc. Ind. Appl. Math. CBMS-Natl. Sci. Found. Monogr.* **38**.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data using Gibbs sampling. *J. Amer. Statist. Assoc.* **85**, 972-985.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **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.

Henderson, R. (1986). Change-point problem with correlated observations, with an application in material accountancy. *Technometrics* **28**, 381-389.

Hinkley, D. V. and Schechtman, E. (1987). Conditional bootstrap methods in the mean-shift model. *Biometrika* **74**, 85-93.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1-17.

Hinkley, D. V. and Hinkley, E. A. (1970). Inference about the change-point in a sequence of binomial variables. *Biometrika* **57**, 477-488.

Hughes, M. D. (1993). Reporting Bayesian analyses of clinical trials. *Statistics in Medicine* **12**, 1651-1663.

Joseph, L., Vandal, A. and Wolfson, D. B. (1996). Estimation in the multi-path change-point problem for correlated data. *Canadian Journal of Statistics* **24**, 37-53.

Joseph, L. and Wolfson, D. B. (1993). Maximum likelihood estimation in the multi-path change-point problem. *Ann. Inst. Statist. Math.* **45**, 511-530.

Josdph, L. and Wolfson, D. B. (1992). Estimation in multi-path change-point problems. *Comm. Statist. Theory Methods* **21**, 897-913.

Liang, K.-Y., Self, S. G. and Liu, X. (1990). The Cox proportional hazards model with change point: An epidemiologic application. *Biometrics* **46**, 783-793.

Lyle, R. M., Melby, C. L., Hyner, G. C., Edmonson, J. W., Miller, J. Z., and Weinberger, M. H. (1987). Blood pressure and metabolic effects of calcium supplementation in normotensive white and black men. *JAMA* **257**, 1772-1776.

Marriot, J., Ravishanker, N., Gelfand, A. and Pai, J. (1992). Bayesian analysis of ARMA processes: complete sampling based inference under full likelihoods. *Research Report*. Department of Statistics, University of Connecticut, Storrs.

Müller, P. (1991). A generic approach to posterior integration and Gibbs sampling. Purdue University, Department of statistics, Technical Report 91-09.

Pettitt, A. N. (1979). A nonparametric approach to the change-point problem. *Appl. Statist.* **28**, 126-135.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1-20.

Smith, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika* **62**, 407-416.

Tanner, M. A. and Wong W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82**, 528-550.

Yao, Y.-C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Ann. Statist.* **15**, 1321-1328.

Yao, Y.-C. (1990). On the asymptotic behavior of a class of nonparametric tests for a change-point problem. *Statist. Probab. Lett.* **9**, 173-177.

Zelterman, D., Grambsch, P. M., Le, C. T., Ma, J. Z. and Curtsinger, J. W. (1994). Piecewise exponential survival curves with smooth transitions. *Math. Biosci.* **120**, 233-250.

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

E-mail: joseph@binky.ri.mgh.mcgill.ca

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

E-mail: roxane@linda.epi.mcgill.ca

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

E-mail: david@zaphod.math.mcgill.ca

Health Promotion Section, Lambert 106, Purdue University, West Lafayette, Indiana, 47907, U.S.A.

E-mail: lyle@vm.cc.purdue.edu

(Received March 1995; accepted May 1996)