A Bayesian analysis of the multiple change-point problem for a Poisson process

Peter J. Green
Department of Mathematics
University of Bristol
Bristol BS8 1TW, UK*

July 18, 1994

1 Introduction

There are a number of interesting statistical problems, typically involving inference about curves, surfaces or images, where the dimension of the object of inference is not fixed. The example discussed here concerns the multiple change-point problem for Poisson processes, where it is assumed that the rate is piecewise constant, but changes an unknown number of times. The times of change and the different rates are unknown. The object of inference is therefore a step function.

The aim of this note is to demonstrate (a) that a Bayesian approach is attractive for such problems, and (b) that the computations for such inference can be handled by novel Markov chain Monte Carlo (MCMC) methods. These methods can probably be adapted for many varying-dimension problems.

2 Coal mining disasters

A classic data set for demonstrating methods for change-point analysis is the point process of dates of serious coal-mining disasters between 1851 and 1962, given by Raftery and Akman (1986). There will be few such disasters in the future (Heseltine, 1993). Figure 1 displays the dates of the 192 disasters in these 112 years = 40907 days as a jittered dot plot, together with the cumulative counting process, shown as a dotted line. For data points $\{y_i, i = 1, 2, \ldots, n\} \in [0, L]$ from a Poisson process with rate given by the function $x(t)$, the log-likelihood is

$$\sum_{i=1}^{n} \log(x(y_i)) - \int_0^L x(t)dt.$$

3 A prior model for step functions

Suppose that there are k steps, at positions $0 \leq p_1 \leq p_2 \leq \ldots \leq p_k \leq L$, and that the step function takes the value h_j on the subinterval (p_j, p_{j+1}), $j = 0, 1, 2, \ldots, k$ (writing $p_0 = 0, p_{k+1} = L$ for convenience). The prior model is specified by supposing that k is drawn from the Poisson distribution

$$\pi(k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Given k, $p_1, p_2, \ldots, p_k$ are distributed as the even-numbered order statistics from $2k+1$ points uniformly distributed on $[0, L]$, and $h_0, h_1, \ldots, h_k$ are independently drawn from the $\Gamma(s, \alpha)$ density

$$\alpha^s h^{s-1} e^{-\alpha h}/\Gamma(s)$$

*Email: P.J.Green@bristol.ac.uk Send an email message to me if you would like a formal written version of this paper, when it exists.

Presented at the 9th International Workshop on Statistical Modelling, Exeter, July 1994.

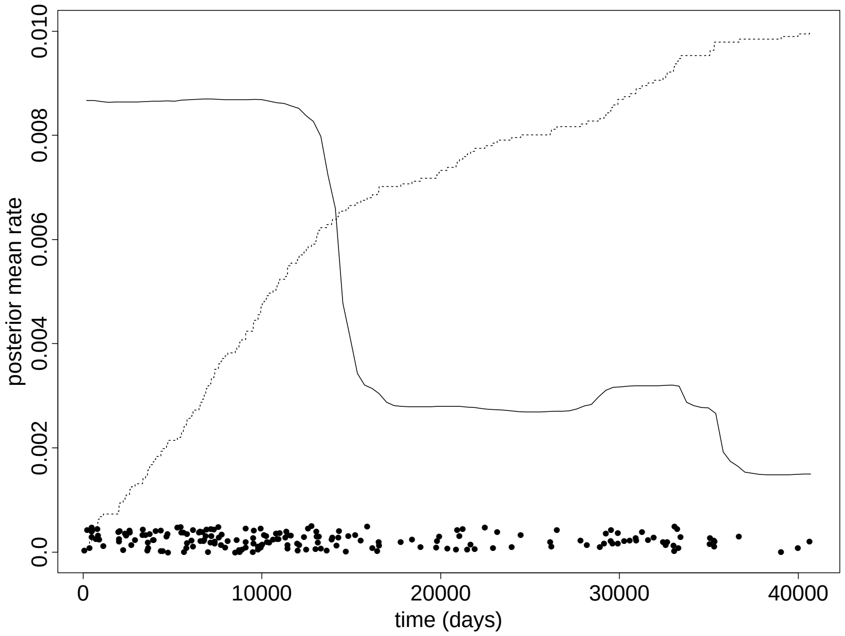

Figure 1: Coal mining disaster data: dates of disasters, cumulative counting process (dotted curve) and posterior mean rate of occurrence (solid curve)

4 MCMC using reversible jumps

Let $\pi(dx)$ denote the target distribution of interest. In MCMC computation (see Besag, Green *et al.* (1994) for a wide-ranging account) we construct a Markov transition kernel $P(x \rightarrow dx')$ satisfying *detailed balance*

$$\int_A \int_B \pi(dx) P(x \rightarrow dx') = \int_B \int_A \pi(dx') P(x' \rightarrow dx),$$

and then simulate this chain to obtain a (dependent, approximate) sample from $\pi(dx)$. Green (1994) described how to do this when there is no simple underlying measure. Briefly, for each of a range of move types indexed by $m = 1, 2, \ldots$, detailed balance is achieved by a generalisation of the method of Hastings (1970). At x, a move of type m to dx' is proposed with probability $q_m(x \rightarrow dx')$, and accepted with probability

$$\alpha(x, x') = \min\left(1, \frac{\pi(dx') q_m(x' \rightarrow dx)}{\pi(dx) q_m(x \rightarrow dx')}\right).$$

The only condition on $q_m()$ is that the ratio above is well-defined, which requires matching dimensions appropriately. (There is no space here to give details of proposals or calculations of α.)

When the object x is a step function on $[0, L]$, some possible moves are (a) birth of a new step at a randomly chosen location in $[0, L]$, (b) death of a randomly chosen step, (c) a change to the height of a randomly chosen step, and (d) a change to the position of a randomly chosen step. Note that (a) and (b) involve changing the dimension of x, so that standard MCMC theory does not apply. In some applications, the number of steps would be fixed in advance (often, change-point analysis assumes exactly one step). Nevertheless, there are clear advantages for efficient MCMC computation to allow k to vary, but to condition on k when drawing information from the realisation. This will allow much better mixing.

For Bayesian computation, the target distribution $\pi(dx)$ is the posterior distribution of x given y, obtained in the usual way from the prior and likelihood terms defined above.

5 Results

Figures 1 to 4 show different aspects of one particular analysis, in which the hyperparameters are set as $\lambda = 3$, $s = 1$ and $\alpha = 200$. The MCMC simulation was run for 40000 updates, after a burn-in period of 4000 updates. A pilot run established that one could have confidence that convergence had taken place by

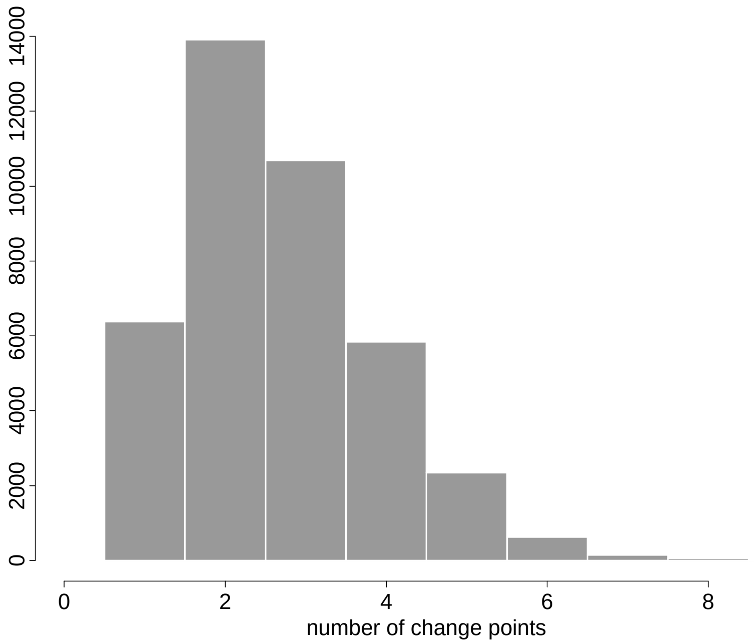

Figure 2: Coal mining disaster data: posterior distribution of k, the number of change points

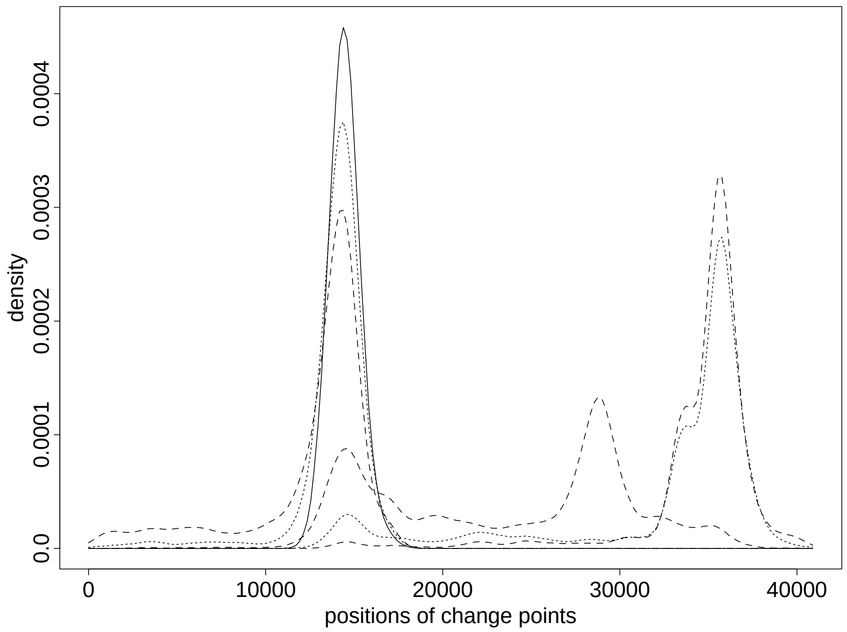

Figure 3: Coal mining disaster data: posterior density estimates of positions of change points, conditional on number of change points $k = 1$ (solid curve), $k = 2$ (dotted curves) and $k = 3$ (broken curves).

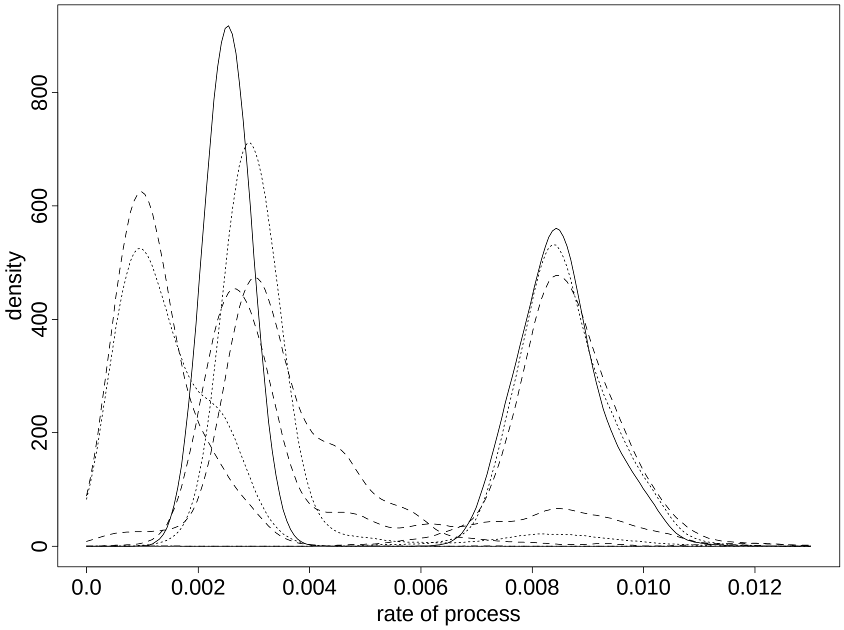

Figure 4: Coal mining disaster data: posterior density estimates of heights of segments of rate step function, conditional on number of change points $k = 1$ (solid curves), $k = 2$ (dotted curves) and $k = 3$ (broken curves).

this point. The computation took less than 1 minute on a Sun IPX workstation. In Figure 1, the solid curve shows the estimated posterior mean curve $E(x(t)|y)$ (not of course a step function). Figure 2 shows the posterior distribution of k, the number of steps. In Figure 3, we show the posterior densities of the step positions, conditional on values $k = 1, 2$ and 3 (the graphs become confusing to interpret with more than this many superimposed). The density estimates are obtained using a Gaussian kernel method with s.d. 625 days. Similarly, Figure 4 shows the corresponding conditional posterior density estimates of step height, using kernel s.d. 0.0003 days^{-1}.

6 Other potential uses for reversible jump MCMC

- factorial experiments (prior allowing factor effects to tie)
- Bayesian choice between models with different numbers of parameters
- image segmentation (2-D analogue of change point problem – pilot version uses Voronoi tessellation based step surfaces)
- object recognition (approached via marked spatial point processes)

References

Besag, J., Green, P. J., Higdon, D. and Mengersen, K. (1994). Bayesian computation and stochastic systems, *University of Bristol Mathematics Research Report S-93-04*, to appear in *Statistical Science*.

Green, P. J. (1994). Contribution to the discussion of paper by Grenander and Miller (1994). *Journal of the Royal Statistical Society* B, **56**, 589–590.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Raftery, A. E. and Akman, V. E. (1986) Bayesian analysis of a Poisson process with a change point. *Biometrika*, **73**, 85–89.