Bayesian analysis of a Poisson process with a change-point

By A. E. RAFTERY and V. E. AKMAN

Department of Statistics, Trinity College, Dublin 2, Republic of Ireland

Summary

A Bayesian approach to estimation and hypothesis testing for a Poisson process with a change-point is developed, and an example given.

Some key words: Bayes factor; Coal-mining disasters; Imaginary observation; Improper prior; Log linear intensity function.

1. Introduction

The purpose of this paper is to develop a Bayesian approach to the retrospective analysis of a Poisson process with a single change-point at an unknown time. The rate of occurrence at time s, $\lambda(s)$, is equal to λ_1 if $0 \leqslant s \leqslant \tau$, and to λ_2 if $s > \tau$. The analysis is based on the observation period $[0, T]$, during which n events occur at times $t = (t_1, \ldots, t_n)$.

In §2 we derive the posterior distribution of the model parameters λ_1, λ_2 and τ. In §3 we derive tests of the existence of a change-point both when the priors are proper, and when vague prior information about the model parameters is represented by limiting improper prior forms. In §4 we give an illustrative example.

Leonard (1978) proposed estimating the parameters of the Poisson process change-point model by minimizing the integrated squared difference between the estimated rate and a different, nonparametric, estimate of the rate function. If, of course, the change is known to have occurred at an event time, the problem reduces to that of a change-point in a sequence of independent exponential variables and may be solved using the methods of Smith (1975) and Diaz (1982). This assumption underlies the work of Commenges & Seal (1985).

Different, but related, problems have been worked on by Kutoyants (1984) who studied the statistical analysis of a Poisson process with a periodic discontinuous rate function, Kalbfleisch & Struthers (1982) who considered a Poisson process analogue of intervention analysis, and Matthews & Farewell (1982) and Nguyen, Rogers & Walker (1984) who analysed hazard rates with change-points.

2. Estimation

We assume that λ_1, λ_2 and τ are independent a priori, and that the prior densities of λ_1 and λ_2 have the conjugate form

$$p(\lambda_k) \propto \lambda_k^{b_k} e^{-a_k \lambda_k} \quad (k = 1, 2). \tag{2.1}$$

Prior beliefs about dependencies between λ_1 and λ_2 may quite easily be incorporated into the analysis (Smith, 1975, §3).

The likelihood is

$$p(t \mid \lambda_1, \lambda_2, \tau) = \lambda_1^{N(\tau)} e^{-\lambda_1 \tau} \lambda_2^{N(T) - N(\tau)} e^{-\lambda_2(T - \tau)},$$

where $N(s)$ is the number of events that occurred in the interval $[0, s]$. Thus the posterior density is such that

$$p(\lambda_1, \lambda_2, \tau \,|\, t) \propto p(\tau) \prod_{k=1}^{2} \lambda_k^{r_k(\tau)-1} e^{-\lambda_k S_k(\tau)}, \qquad (2\cdot2)$$

where $r_1(\tau) = N(\tau) + b_1 + 1$, $r_2(\tau) = N(T) - N(\tau) + b_2 + 1$, $S_1(\tau) = \tau + a_1$, $S_2(\tau) = (T - \tau) + a_2$. The posterior density of τ is thus

$$p(\tau \,|\, t) \propto p(\tau) \prod_{k=1}^{2} S_k(\tau)^{-r_k(\tau)} \Gamma\{r_k(\tau)\}. \qquad (2\cdot3)$$

The posterior density of λ_1 is obtained by integrating $(2\cdot2)$ over λ_2 and τ. This does not yield a simple analytic form, and the function in $(2\cdot2)$ is discontinuous in τ, so that the most convenient form for numerical integration is as a sum of integrals of continuous functions, as follows:

$$p(\lambda_1 \,|\, t) \propto \sum_{i=0}^{n} \Gamma(r_{2i}) \lambda_1^{r_{1i}-1} I_i^{(1)},$$

where

$$I_i^{(1)} = \int_{t_i}^{t_{i+1}} e^{-\lambda_1 S_1(\tau)} S_2(\tau)^{-r_{2i}} p(\tau) \, d\tau, \quad r_{1i} = i + b_1 + 1,$$

$$r_{2i} = n - i + b_2 + 1, \quad t_0 = 0, \quad t_{n+1} = T.$$

The result for λ_2 is similar.

The posterior distribution of the magnitude of the change, $\beta = \lambda_1/\lambda_2$, follows from the fact that by Lindley (1965, Th. 7.3.2), conditional on τ

$$\{S_1(\tau) \, r_2(\tau)\} \{S_2(\tau) \, r_1(\tau)\}^{-1} \beta \sim F_{\alpha_1, \alpha_2}, \qquad (2\cdot4)$$

where $\alpha_k = 2r_k(\tau)$ $(k = 1, 2)$. Thus

$$p(\beta \,|\, t) \propto \sum_{i=0}^{n} \Gamma(r_{1i}) \, \Gamma(r_{2i}) \, \beta^{r_{1i}-1} I_i^{(2)}, \qquad (2\cdot5)$$

where

$$I_i^{(2)} = \int_{t_i}^{t_{i+1}} \{S_2(\tau)/S_1(\tau)\}^{r_{1i}} \{S_2(\tau) + \beta S_1(\tau)\}^{-(n+b_1+b_2+2)} p(\tau) \, d\tau.$$

3. Hypothesis testing

Here we view the problem of testing for a change as one of comparing the change-point model with one which assumes a constant rate. We denote the constant rate model by M_0, its parameter by λ_0, and the change-point model by M_1. We assume that the prior density of λ_0 has the form $(2\cdot1)$ with $k = 0$. Then the Bayes factor for M_0 against M_1 is defined by

$$B_{01}^{(n)}(t, T) = p(t \,|\, M_0)/p(t \,|\, M_1), \qquad (3\cdot1)$$

where

$$p(t \,|\, M_0) = \int_0^\infty p(t \,|\, \lambda_0) \, p(\lambda_0) \, d\lambda_0$$

$$= \Gamma(n + b_0 + 1) \{\Gamma(b_0 + 1)\}^{-1} a_0^{b_0+1} (a_0 + T)^{-(n+b_0+1)},$$

$$p(t \,|\, M_1) = \int_0^T \int_0^\infty \int_0^\infty p(t \,|\, \lambda_1, \lambda_2, \tau) \, p(\lambda_1, \lambda_2, \tau) \, d\lambda_1 \, d\lambda_2 \, d\tau$$

$$= \left[\prod_{k=1}^{2} a_k^{b_k+1} \{\Gamma(b_k + 1)\}^{-1} \right] \sum_{i=0}^{n} \Gamma(r_{1i}) \, \Gamma(r_{2i}) \, I_i^{(3)},$$

where

$$I_i^{(3)} = \int_{t_i}^{t_{i+1}} S_1(\tau)^{-r_{1i}} S_2(\tau)^{-r_{2i}} p(\tau) \, d\tau.$$

We now consider the situation where vague prior information about the parameters λ_0, λ_1 and λ_2 is represented by limiting improper prior forms. This is specified by $a_k \to 0$ $(k = 0, 1, 2)$ in (2·1), so that $p(\lambda_0) = c_0 \lambda_0^{b_0}$ and $p(\lambda_1, \lambda_2, \tau) = c_1 \lambda_1^{b_1} \lambda_2^{b_2} p(\tau)$. The resulting Bayes factor involves an arbitrary, unspecified, ratio of constants, c_0/c_1, and is thus not well-defined. We shall allow c_0/c_1 to depend on T, denoting it by $c_{01}(T)$.

We adopt the method for assigning $c_{01}(T)$ proposed by Spiegelhalter & Smith (1982). This consists of imposing a boundary condition on the Bayes factor as a function of the data. The boundary condition is that a data set which involves the smallest possible sample size permitting a comparison of M_0 and M_1 and provides maximum possible support for M_0 should yield a Bayes factor which is only very slightly greater than one.

This method does not provide a solution to the problem for all improper prior specifications in the case of nonhomogeneous Poisson processes. We have shown in an unpublished paper that it does yield an operational solution if the Bayes factor in (3·1) is bounded as a function of t and invariant to scale changes in the time variable. Under mild conditions on $p(\tau)$, for this to be the case here it is necessary that $c_{01}(T)$ be proportional to $T^{-(1+b_1+b_2-b_0)}$.

When $p(\tau) = T^{-1}$ and $b_1 = b_2$, the appropriate boundary condition is that a data set consisting of a single event occurring half-way through the observation period should yield a Bayes factor approximately equal to one. If $b_0 = b_1 = b_2 = -\tfrac{1}{2}$, corresponding to Jeffreys's prior for λ_0 and also, conditionally on τ, for λ_1 and λ_2, we have $c_{01}(T) = 4\pi^{\frac{1}{2}} T^{-\frac{1}{2}}$, so that

$$B_{01}^{(n)}(t,T) = 4\pi^{\frac{1}{2}} T^{-n} \Gamma(n+\tfrac{1}{2}) \left\{ \sum_{i=0}^{n} \Gamma(i+\tfrac{1}{2}) \Gamma(n-1+\tfrac{1}{2}) I_i^{(4)} \right\}^{-1}, \qquad (3·2)$$

where

$$I_i^{(4)} = \int_{t_i}^{t_{i+1}} \tau^{-(i+\frac{1}{2})} (T-\tau)^{-(n-i+\frac{1}{2})} d\tau.$$

Equation (3·2) can provide the basis for an informal sequential procedure (Smith, 1975, § 5).

4. An illustrative example

We now apply the results of the previous sections to the data set consisting of intervals between coal-mining disasters given by Jarrett (1979), who corrected and extended the data set given by Maguire, Pearson & Wynn (1952). We extend the data set further again by noting from the *Colliery Year Book and Coal Trades Directory 1963–64* that no disasters appear to have occurred between 1 January and 14 March 1851, or between 23 March and 31 December 1962.

Previous authors, including Barnard (1953), Cox & Lewis (1966, p. 42), Jarrett (1979) and Berman (1981), have fitted smoothly changing rates of occurrence to the data. However, both the plot of the cumulative number of disasters against time of Jarrett (1979, Fig. 1) and the histogram with the smallest risk function estimate given by Rudemo (1982, Fig. 12A) suggest that a change-point model may be appropriate.
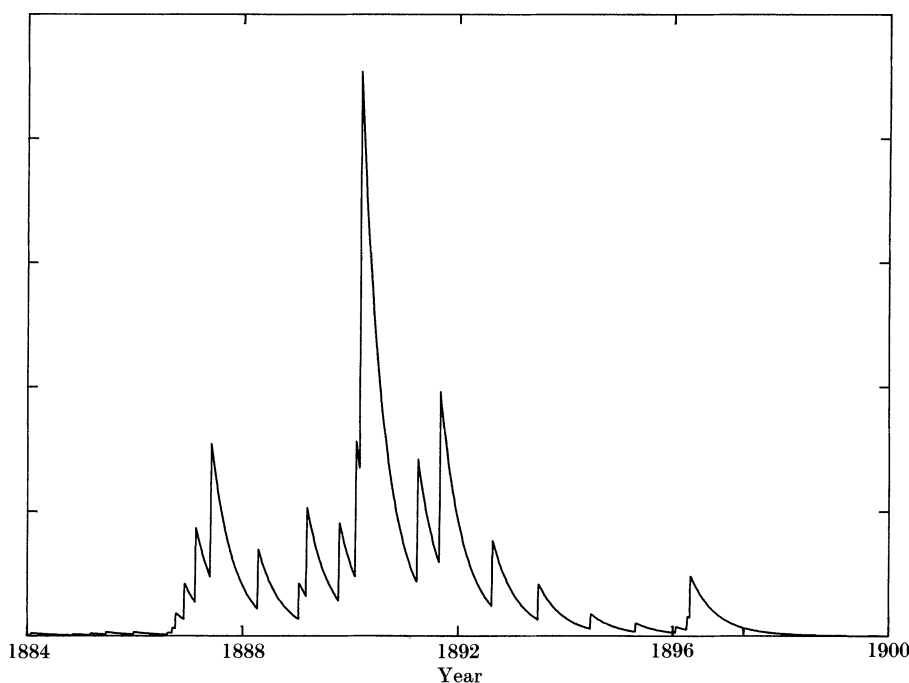
Fig. 1. Posterior density of the change-point for the coal-mining data.

We use a vague prior distribution, specified by setting $a_k = 0$ and $b_k = -\frac{1}{2}$ $(k = 1, 2)$ in (2·1), and $p(\tau) = T^{-1}$. Fig. 1 shows the posterior distribution of τ obtained from (2·3). The posterior mode is 10 March 1890 and the posterior median is 27 August 1890. A 95% Bayesian estimation interval for τ, with limits at the $2\frac{1}{2}\%$ and $97\frac{1}{2}\%$ points of the posterior distribution is [15 May 1887, 3 August 1895].

By (2·5), the magnitude of the change β has posterior mean 3·41 with 95% highest posterior density region [2·48, 4·46]. By comparison, using (2·4) conditionally on τ being equal to its posterior mode gives posterior mean 3·52 and 95% highest posterior density region [2·56, 4·68] for β, which is a fairly reasonable approximation.

Using (3·2), the Bayes factor for no change against a change-point is $1·58 \times 10^{-14}$, which constitutes strong evidence for a change.

We now compare the change-point model with the log linear rate model favoured by previous authors. Using methods similar to those developed in § 3, it can be shown that the Bayes factor for a constant rate against a decreasing log linear rate is $6·28 \times 10^{-11}$. Thus, assuming prior indifference between the two models, the posterior odds in favour of the change-point model as against the log linear rate model are over 4000 : 1.

This suggests that the observed decrease in the rate of occurrence between 1851 and 1962 may consist mainly of a fairly abrupt decrease around the 1887–1895 period, rather than a sustained gradual decrease. This may be associated with changes in the coal industry around that time, namely a severe decline in labour productivity starting at the end of the 1880s, and the emergence of the Miners' Federation at the end of 1889; see Taylor (1968). A possible causal link is suggested by Dwyer (1983) who argues that industrial accident rates are closely tied to the state of functioning of social relations in the workplace, so that trade unionization may decrease accident rates, while factors associated with higher labour productivity, such as overtime, may cause an increase in accident rates.

REFERENCES

BARNARD, G. A. (1953). Time intervals between accidents—a note on Maguire, Pearson and Wynn's paper. *Biometrika* **40**, 212–3.

BERMAN, M. (1981). Inhomogeneous and modulated gamma processes. *Biometrika* **68**, 143–52.

Colliery Year Book and Coal Trades Directory 1963–64. London: Colliery Guardian.

COMMENGES, D. & SEAL, J. (1985). The analysis of neuronal discharge sequences: Change-point estimation and comparison of variances. *Statist. Medicine* **4**, 91–104.

COX, D. R. & LEWIS, P. A. W. (1966). *The Statistical Analysis of Series of Events.* London: Methuen.

DIAZ, J. (1982). Bayesian detection of a change of scale parameter in sequences of independent gamma random variables. *J. Econometrics* **19**, 23–9.

DWYER, T. (1983). A new concept of the production of industrial accidents: A sociological approach. *N.Z. J. Ind. Rel.* **8**, 147–60.

JARRETT, R. G. (1979). A note on the intervals between coal-mining disasters. *Biometrika* **66**, 191–3.

KALBFLEISCH, J. D. & STRUTHERS, C. A. (1982). An analysis of the Reynolds Metal Company data. *Can. J. Statist.* **10**, 240–7.

KUTOYANTS, Y. A. (1984). *Parameter Estimation for Stochastic Processes.* Berlin: Heldermann-Verlag.

LEONARD, T. (1978). Density estimation, stochastic processes and prior information (with discussion). *J. R. Statist. Soc.* B **40**, 113–46.

LINDLEY, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*, **2**. *Inference.* Cambridge University Press.

MAGUIRE, B. A., PEARSON, E. S. & WYNN, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika* **39**, 168–80.

MATTHEWS, D. E. & FAREWELL, V. T. (1982). On testing for a constant hazard against a change-point alternative. *Biometrics* **38**, 463–8.

NGUYEN, H. T., ROGERS, G. S. & WALKER, E. A. (1984). Estimation in change-point hazard rate models. *Biometrika* **71**, 299–304.

RUDEMO, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.* **9**, 65–78.

SMITH, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika* **62**, 407–16.

SPIEGELHALTER, D. J. & SMITH, A. F. M. (1982). Bayes factors for linear and log-linear models with vague prior information. *J. R. Statist. Soc.* B **44**, 377–87.

TAYLOR, A. J. (1968). The coal industry. In *The Development of British Industry and Foreign Competition 1875–14*, Ed. D. H. Aldcroft, pp. 37–70. London: Allen & Unwin.