

An application of Markov chain Monte Carlo (MCMC) to continuous-time incurred but not yet reported (IBNYR) events

Garfield O. Brown

Statistical Laboratory, Centre for Mathematical Sciences, Cambridge CB3 0WB, UK

Winston S. Buckley*

Mathematical Sciences, Bentley University, Waltham, MA 02452, USA

Abstract

We develop a Bayesian model for continuous-time incurred but not yet reported (IBNYR) events under four types of secondary data, and show that unreported events, such as claims, have a Poisson distribution with a reduced arrival parameter if event arrivals are Poisson distributed. Using insurance claims as an example of an IBNYR event, we apply Markov chain Monte Carlo (MCMC) to the continuous-time IBNYR claims model of Jewell using Type I and Type IV data. We illustrate the relative stability of the MCMC method versus the Gammoid approximation of Jewell by showing that the MCMC estimates approach their prior parameters, while the Gammoid approximations grow without bound for Type IV data. Moreover, this holds for any distribution that the delay parameter is assumed to follow. Our framework also allows for the computation of posterior confidence intervals for the parameters.

Keywords

MCMC; Delay parameter; IBNYR events; Jewell model; Insurance claims

1. Introduction

We closely follow Jewell (1989) to develop a basic Bayesian model of delayed reporting of events that occur in a given exposure interval and analysed in any observation interval, when only an incomplete number of events have been reported. We use an exact, full-distributional Bayesian approach in our modelling. Using insurance claims as delayed reported events, we then apply Markov chain Monte Carlo (MCMC) to the continuous-time incurred but not yet reported (IBNYR) claims model of Jewell and illustrate the relative stability of MCMC methods versus the Gammoid approximation of Jewell (1989) when the data are claims count only. An IBNYR event is one that occurs randomly during a fixed exposure interval and incurs a random delay before it is reported. Both the rate at which such events occur and the parameters of the delay distribution are unknown random quantities. Conditional on the number of events that have been reported during some observation interval, along with secondary data on the dates of events, the problem is to estimate the true values of the events generating and delay parameters, and thereafter, to predict the number of events that are unreported. Typical examples of IBNYR events are

*Correspondence to: Winston S. Buckley, Centre for Mathematical Sciences, Bentley University, Waltham, MA 02452, USA. Tel: +1 203-550-9765; E-mail: winb365@hotmail.com

insurance claims, number of persons infected by an epidemic/disease, the impact of a drug on mortality rate, products under warranty, murders committed by a serial killer or gang, rapes, financial frauds, survey sampling by mail and undetected bugs in computer software. In this paper, we focus only on the number of incurred but unreported events, and not any associated costs. Thus, for example, severity or sizes of insurance claims are not considered in this paper, and is left for future research.

Reserving for future claims is an important aspect of an insurance company's operations. However, in order to plan effectively, there must be some way of obtaining estimates of the number of outstanding claims and their sizes. Jewell (1989) studies a continuous-time version of this problem for claims that have IBNYR. These claims occur at random times, but there is also a delay in the reporting of each occurrence relative to a fixed time horizon, which is usually 1 year. He discusses parameter estimation for four different scenarios determined by the amount of available data. Type I data have both known reporting and occurrence dates. Type II data have only reporting date known. Type III data have only occurrence date known, while Type IV data have neither occurrence nor reporting dates known, and are therefore very non-informative. Thus, for type IV data, only counts are available. Jewell (1990) also studies the same problem in discrete time.

As mentioned above, there are many forms of continuous-time IBNYR events. However, we apply our model only to IBNYR insurance claims as developed by Jewell (1989). We use MCMC methods to analyse the model when the secondary data are Type I and Type IV. A Gibbs sampling procedure is used to draw MCMC samples. We then compare our results with those of Jewell (1989) and discuss how an important limitation of his computational method can be overcome. We first develop the model for Type I data in detail, and then implement it using MCMC.

Assuming exponential delays, we also demonstrate that our computational methods have a significant advantage over the Gammoid approximation of Jewell (1989) for Type IV data, which is count data only. In a Bayesian setting where the likelihood function does not carry much information about the parameters of interest, the posterior distribution of such a parameter should revert to the prior distribution and should not become unboundedly large. In particular, we show that when MCMC methods are used, the likelihood function does not become unbounded when the data are Type IV, and hence, also for Type III data.

We illustrate the relative stability of the MCMC method versus the Gammoid approximation of Jewell (1989) by showing that the MCMC estimates approach the prior parameters, while the Gammoid approximations grow without bound. Moreover, this holds for any distribution that the delay parameter follows. However, unlike Jewell (1989), our framework allows for the computation of posterior confidence intervals for our parameters for both types of secondary data. We also show that delayed claims have a Poisson distribution with a reduced parameter that nests that of Jewell.

Our primary conclusion is that in the absence of an informative likelihood, the prior distribution dominates the likelihood function, and that the prior information presents our best available knowledge about the parameters of interest. Another interesting fact is that, based on the computation method that we use, these conclusions hold not only for the exponential distribution, but for any delay distribution including heavy-tailed distributions. Thus, delays following log-logistic, Weibull, Gamma, Pareto, log-normal, etc., could be appropriate in some applications.

The rest of the paper proceeds as follows. In section 2, we present the model and derive the likelihood function. The Jewell (1989) model is reviewed in section 3. We implement our model on insurance claims using MCMC in section 4. A summary and discussion are presented in section 5,

where we offer suggestions for a possible extension to the model to include a cost (e.g. claim size/severity) associated with the occurrence of each event.

2. The Model

We adopt the same model as in Jewell (1989), with the assumption that events are generated by a homogeneous Poisson process with arrival rate λ over some fixed interval $(0, T]$ (Figure 1).

Thus, there are unknown number, $N(T)$, of events occurring at unknown occurrence times s_1, s_2, \dots, s_n , given $N(T) = n$. We exclude the possibility of two or more events occurring at the same time, thus $s_i \neq s_j, i \neq j$ almost surely (a.s.). Each event j is linked with a positive random waiting time (reporting delay), $w_j > 0$, such that its observation epoch (reporting date) is $t_j = s_j + w_j$. We shall assume a fully parametric model and that the $\{w_j\}$ are independent and identically distributed random variables with common distribution function $F(w|\theta)$. We also assume F is absolutely continuous with density $f(w|\theta) = F'(w|\theta)$ and θ is a delay parameter, possibly vector valued.

Let τ denote the point in time at which we observe the IBNYR event process. For any generic event, the support of the joint density of the occurrence and reporting times would be the section of the (s, t) plane for which

$$0 \leq s \leq T \text{ and } s \leq t \leq \tau$$

since all events occur in the interval $(0, T]$ and are then reported after they occur. Conceptually, this means an event is a point in the plane; the first co-ordinate $s(\leq T)$, is the time when the event occurs and the second co-ordinate $t(\geq s)$ is the time when it is first reported.

We make the assumption that, conditional on the value of θ known, every pair (s, t) is statistically independent of every other pair, and has common joint density

$$p(s, t | \theta) = \begin{cases} \frac{1}{T} f(t-s | \theta) & 0 < s \leq T, s \leq t < \infty \\ 0 & \text{otherwise} \end{cases}$$

where f is the density of the waiting time random variable W . The scaling factor $\frac{1}{T}$ being necessary since

$$\begin{aligned} \int_0^T \int_s^\infty f(t-s | \theta) dt ds &= \int_0^T [F(t-s | \theta)|_{t=s}^{t=\infty}] ds \\ &= \int_0^T ds \\ &= T \end{aligned}$$

At time τ , assume that $r = r(\tau)$ events have been observed. We shall denote the ordered pair of occurrence time and reporting time by D_j , that is, $D_j = (s_j, t_j)$ where $j = 1, \dots, r(\tau)$. Also, we shall



Figure 1. Incurred events reporting process. Events occur at some time s and are reported at some later time t . IBNYR, incurred but not yet reported.

denote by $D(\tau)$ the collective data on all events which are reported at time τ , that is, $D(\tau) = \{D_1, \dots, D_{r(\tau)}\}$. For the degenerate case $r(\tau) = 0$, $D(\tau)$ is then the empty set. Clearly, $D(\tau)$ are non-decreasing sets, so that $D(\tau_1) \subset D(\tau_2)$ for $\tau_1 \leq \tau_2$, which means our knowledge of the process cannot decrease as time progresses. One of our aims is to predict the number of unreported events $u(\tau) = N(\tau) - r(\tau)$ at time τ , based on the information in $D(\tau)$.

2.1. The likelihood function

Recall that the IBNYR event process occurs in continuous time. We derive the likelihood of the observed process at time τ by conditioning on there being $N(\tau) = n$ events, and that $r(\tau)$ of these are known. Conditional on $N(\tau) = n$, the likelihood for the observed data at time τ , denoted L_τ , is

$$L_\tau(D(\tau) | \theta, n) \propto \frac{n!}{(n-r)!} \left[\prod_{j=1}^r p(D_j | \theta) \right] [1 - \Pi(\tau | \theta)]^{n-r} \tag{1}$$

with

$$p(D_j | \theta) = p(s_j, t_j | \theta)$$

and

$$\Pi(\tau | \theta) = \begin{cases} \frac{1}{T} \int_0^\tau F(w | \theta) dw & \tau \leq T \\ \frac{1}{T} \int_{\tau-T}^\tau F(w | \theta) dw & \tau > T \end{cases} \tag{2}$$

Note that $\Pi(\tau | \theta)$ represents the probability that an event will be reported before its observation time τ . Thus, $1 - \Pi(\tau | \theta)$ is the probability that an event is unreported at time τ , and therefore, the probability that $u = n - r$ events are outstanding/unreported at time τ is proportional to $[1 - \Pi(\tau | \theta)]^{n-r}$.

Let $M = \min\{\tau, T\}$. Given λ , $N(\tau)$ is $\text{Poisson}(\lambda M)$ by assumption that events are generated by a homogeneous Poisson process with rate λ over some fixed interval $(0, T]$. To see this, note that the exposure interval is $(0, T]$, and so for $\tau \leq T$, $N(\tau) \sim \text{Poisson}(\lambda \tau)$; while for $\tau > T$, $N(\tau) \sim \text{Poisson}(\lambda T)$, since no events can occur after T .

We can therefore multiply (1) by a $\text{Poisson}(\lambda M)$ density to arrive at

$$L_\tau(D(\tau) | \theta, n) \times p(n | \lambda) \propto \frac{n!}{(n-r)!} \prod_{j=1}^r [p(D_j | \theta)] [1 - \Pi(\tau | \theta)]^{n-r} \times \frac{1}{n!} (\lambda M)^n e^{-\lambda M}$$

After simplification, we have

$$\begin{aligned} L_\tau(D(\tau) | \lambda, n, \theta) &\propto \frac{1}{(n-r)!} \prod_{j=1}^r [p(D_j | \theta)] [1 - \Pi(\tau | \theta)]^{n-r} e^{-\lambda M} (\lambda M)^n \\ &= \frac{1}{(n-r)!} \prod_{j=1}^r [p(D_j | \theta)] [1 - \Pi(\tau | \theta)]^{n-r} e^{-\lambda M} (\lambda M)^{n-r} (\lambda M)^r \end{aligned}$$

and substituting $u = n - r$, yields

$$L_\tau(D(\tau) | \lambda, u, \theta) \propto \frac{1}{u!} \prod_{j=1}^r [p(D_j | \theta)] [\lambda M (1 - \Pi(\tau | \theta))]^u e^{-\lambda M} (\lambda M)^r$$

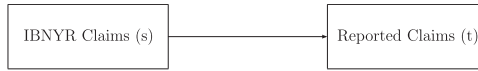


Figure 2. Claims reporting process. Claims occur at some time s and are reported at some later time t . IBNYR, incurred but not yet reported.

By rearranging terms, we see that the last part is a Poisson density in terms of u . Thus

$$L_\tau(D(\tau) | \lambda, u, \theta) \propto e^{-\lambda M} \prod_{j=1}^r [p(D_j | \theta)] (\lambda M)^r \frac{1}{u!} [\lambda M(1 - \Pi(\tau | \theta))]^u \tag{3}$$

Integrating (3) with respect to u and simplifying, yields

$$L_\tau(D(\tau) | \lambda, \theta) \propto e^{-\lambda M \Pi(\tau | \theta)} \prod_{j=1}^r [p(D_j | \theta)] (\lambda M)^r \tag{4}$$

Summarising, $u(\tau)$, the number of unreported events at observation time τ , has a Poisson distribution with parameter $\lambda M(1 - \Pi(\tau | \theta))$. This is not entirely surprising as we could also have derived this result using the colouring theorem for Poisson processes (see e.g. Kingman, 1993: 53). Note that, as is expected, the Poisson parameter for unreported events will tend to 0 as the observation time τ increases. The rate at which the parameter tends to 0, however, will be affected by the size of the right tail of the delay distribution (Figure 2).

Remark 1 Although some events can remain unreported for very long periods, the delay distribution is not usually long tailed. For example, asbestos claims can remain latent for decades and clearly, in this case, the delays are long tailed. However, it is generally accepted in the insurance industry that claims usually have smaller reporting delays.

The presentation given up to this point makes no restrictions on the delay parameter θ . However, for illustration purposes, we shall assume exponential delays. Thus, for Type I data

$$p(D_j | \theta) = \frac{1}{T} \theta e^{-\theta(t_j - s_j)} = \frac{1}{T} \theta e^{-\theta w_j}$$

while for Type IV data

$$p(D_j | \theta) = \Pi(\tau | \theta) = \frac{1}{T} \int_{(\tau - T)^+}^{\tau} F(w | \theta) dw$$

The reported events will be those for which $t_j \leq \tau$. For a Bayesian analysis, we place prior distributions on the parameters λ and θ . At time τ , the posterior density of the parameters λ and θ takes the form

$$\pi_\tau(\theta, \lambda | D(\tau)) \propto L_\tau(D(\tau) | \lambda, \theta) p(\lambda) p(\theta)$$

where $L_\tau(D(\tau) | \lambda, \theta)$ will be specified given the assumed data type.

3. The Jewell model

Jewell (1989) formulates a basic, continuous-time Bayesian model for predicting the total number of unreported IBNYR claims arising in a given exposure interval, when only an incomplete number of such claims have been reported by some point in time, called the observation time/date. He assumes that events are generated by a homogeneous Poisson process with rate parameter λ (events/year) over some fixed exposure interval $(0, T]$. There is also a delay in the reporting of these events, and the

Table 1. Secondary data type.

Type	Occurrence dates	Reporting dates
I	Known	Known
II	Unknown	Known
III	Known	Unknown
IV	Unknown	Unknown

waiting time is driven by a continuous distribution with parameter θ , which could be vector valued. The delay distribution is common to all claims.

In Jewell’s Bayesian formulation, the prior densities of $\tilde{\lambda}$ and $\tilde{\theta}$ are assumed to be independent, and the posterior parameter density, $p(\lambda, \theta | D)$, is not very revealing for any choice of priors. Information about these parameters are obtained through an experiment that observes all reported events $r(t)$ in some observation interval $(0, t]$, where t is continuous. Jewell therefore focusses on the prediction of the unreported claim count, $u(t) = n(T) - r(t)$, conditioned on D , the amount of secondary data that is available, of which there are four types (I–IV) as shown in Table 1.

The unreported claim u has posterior density $p(u | D)$, that takes the form

$$p(u | D) \propto p_\lambda(u | D)p_\theta(u | D) \tag{5}$$

where $p_\lambda(u | D)$ and $p_\theta(u | D)$ are occurrence/arrival and delay integrals, respectively, given by

$$p_\lambda(u | D) = \frac{T^u}{u!} \int \lambda^{r+u} e^{-\lambda T} p(\lambda) d\lambda$$

and

$$p_\theta(u | D) = \int L(\theta | D)[K(\theta)]^u p(\theta) d\theta$$

where $L(\theta | D) = \prod_{i=1}^r [p(D_i | \theta)]$, the Kernel $K(\theta) = 1 - \frac{\tau}{T} \Pi(t | \theta)$ and $\tau = \min\{t, T\}$.

With the choice of the Gamma (a, b) prior for $\tilde{\lambda}$, the occurrence integral and predictive density reduce, respectively, to

$$h_\lambda(u | D) = \frac{\Gamma(a+r+u)}{u!} \left(\frac{T}{b+T}\right)^u$$

$$\frac{p(u+1 | D)}{p(u | D)} = \left(\frac{a+r+u}{u+1}\right) \left(\frac{T}{b+T}\right) \left(\frac{h_\theta(u+1 | D)}{h_\theta(u | D)}\right)$$

which is the predictive density in recursive form. If θ is known exactly, then

$$p(u | D) = \frac{\Gamma(a+r+u)}{u!} \left(\frac{T-\tau\Pi(t|\theta)}{b+T}\right)^u$$

which is a Pascal predictive density, with respective conditional mean and variance

$$E(u | D) = \frac{(a+r)T}{b} \left(\frac{1-\frac{\tau}{T}\Pi(t|\theta)}{1+\frac{\tau}{T}\Pi(t|\theta)}\right)$$

$$Var(u | D) = E(u | D) \left(\frac{b+T}{b+\tau\Pi(t|\theta)}\right)$$

3.1. Gammoid approximation and Type III and IV data

When θ is unknown, the computation of the delay integral becomes more complex. Assuming that delays follow the familiar exponential distribution $f(w|\theta) = \theta e^{-\theta w}$, Jewell computes different likelihood depending on the type of data.

For example, Type I data have

$$p(D_j|\theta) = \frac{1}{T} \theta e^{-\theta w}$$

while for Type IV data

$$p(D_j|\theta) = \Pi(t|\theta) = \frac{\tau}{T} - \frac{1}{\theta T} \left(e^{-\theta(t-T)} - e^{-\theta t} \right)$$

Jewell estimates the predictive density

$$p_\theta(u|D) = \int L(\theta|D)[K(\theta)]^u p(\theta) d\theta$$

using the Gammoid function $g(\theta) = (A\theta)^\Gamma e^{-\Delta\theta}$ to approximate the first two factors in the integral for $p_\theta(u|D)$ above. This is achieved by assuming that the prior on the unknown delay parameter θ is Gamma (c_0, d_0), which has a mode at $\theta_0 = \frac{c_0-1}{d_0}$. Approximations are centred around the mode of the integrand. This procedure converts the integral into a Gamma integral that is dependent on u . Jewell obtains excellent results with $c_0 = 3$ or 4 for Type I or II data. For Type III or IV data, he obtains the recursive formula for the predictive density as

$$\frac{p(u+1|D)}{p(u|D)} = \left(\frac{a+r+u}{u+1} \right) \left(\frac{T}{b+T} \right) \left(\frac{d_0 + \Delta + \delta_K u}{d_0 + \Delta + \delta_K + \delta_K u} \right)^{c_0 + \Gamma}$$

where Γ and Δ are computed from the number of reported events, and $\delta_K \approx \frac{t^3}{2T^2}, t \leq T$ and $\delta_K \approx t - \frac{T}{2}, t \geq T$. Jewell’s model works very well for Types I and II. However, using Type IV data (count data only), a numerical example shows that the model gives a steady and dramatic increase in all the predictors as r increases (or t increases beyond the exposure interval), and the point estimators (mean and variance) grow without bound. Jewell posits that this happens because there is less and less information in Type IV data as t increases. This is also true for Type III data.

Jewell concludes that if the reporting dates of IBNYR claims are not available then Bayesian predictions, though mathematically correct, are operationally useless. This is because Type IV (or III) data are uninformative when the priors on $\tilde{\lambda}$ and $\tilde{\theta}$ are not sufficiently precise. Satisfactory stability in estimating the parameters and predicting the unreported claims require the observation of at least the reporting dates, that is, the time history of the reported claims, $r(t)$.

We will show that when MCMC is applied, no such problems arise for Type III or IV data, which has the least information. Thus, the problems encountered by Jewell’s Gammoid approximation are completely circumvented when MCMC methods are used. This affirms Jewell’s claim that the Bayesian formulation is not the problem. However, we disagree with his conclusion that it is the lack of information that creates explosions in the estimators, but rather the approximation method that was employed.

4. MCMC Implementation

We now apply MCMC to the continuous-time IBNYR claims model of Jewell (1989). A Gibbs sampling procedure is used to draw MCMC samples. We first develop the model for Type I data in detail, and implement it using MCMC. We then consider claims count only using Type IV data and demonstrate that our computational methods have a significant advantage over the Gammoid approximation of Jewell (1989) for Type IV (or III) data. In a Bayesian setting where the likelihood function does not carry much information about the parameters of interest, the posterior distribution of such a parameter should revert to the prior distribution and should not become unboundedly large. In particular, we show that when MCMC methods are used, the likelihood function does not become unbounded when the data are Type IV. We illustrate the relative stability of the MCMC method versus the Gammoid approximation of Jewell (1989) by showing that the MCMC estimates approach their prior parameters, while the Gammoid approximations grow without bound. Moreover, this holds for any distribution that the delay parameter follows. Our framework also allows for the computation of posterior confidence intervals for our parameters.

4.1. Simulations

We implement the model using simulated data. Our data set have known parameters $\theta = 0.5$, $\lambda = 100$ and $T = 1$, for which the expected number of claims is $\lambda T = 100$. For this particular simulated data set, the observed number of claims is 107 and the mean inter-arrival time is $0.432^{-1} = 2.315$ years (Figures 3 and 4).

Before presenting detailed results for Types I and IV data, note that for $0 < \tau \leq T$, we focus on the number of claims in $[0, \tau]$ which are unreported at time τ , while for $\tau > T$, we focus on claims in $[0, T]$ which are unreported at time τ . This is different from Jewell (1989) who at any observation time $\tau > 0$ considers only unreported claims in the interval $[0, T]$.

4.2. Claims with full reporting dates

For the Bayesian analysis, we use a Gamma (a_1, b_1) prior on λ and a Gamma (a_2, b_2) prior on θ . At time τ , given $r = r(\tau)$ reported claims, the joint posterior of λ and θ is

$$\pi(\lambda, \theta | D(\tau)) \propto \prod_{\{j: t_j \leq \tau\}} \theta e^{-\theta w_j} (\lambda M)^r e^{-\lambda M \Pi(\tau|\theta)} p(\lambda) p(\theta)$$

where we have removed the uninformative terms involving T , and

$$\Pi(\tau | \theta) = \begin{cases} \frac{\tau}{T} - \frac{1}{\theta T} [1 - e^{-\theta \tau}] & \tau \leq T \\ 1 - \frac{1}{\theta T} [e^{-\theta(\tau-T)} - e^{-\theta \tau}] & \tau > T \end{cases}$$

follows from (2) above, given exponential waiting times from occurrence to notification. The posterior conditional of λ is

$$\begin{aligned} \pi_\tau(\lambda | \theta, D(\tau)) &\propto L_\tau(D(\tau) | \lambda, \theta) p(\lambda) \\ &\propto \lambda^{a_1 + r - 1} \exp\{-\lambda(b_1 + M\Pi(\tau|\theta))\} \end{aligned} \tag{6}$$

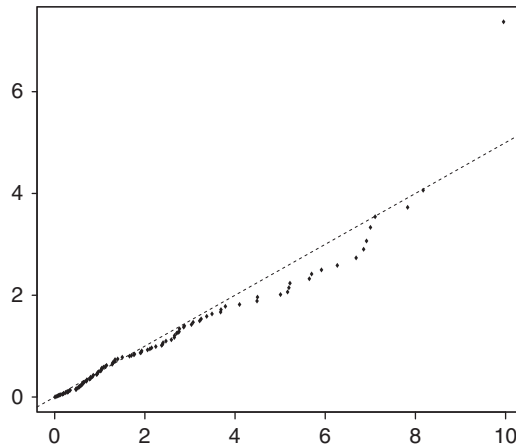


Figure 3. QQ plot of the simulated inter-arrival times against the standard exponential. The dashed line has slope 1/2.

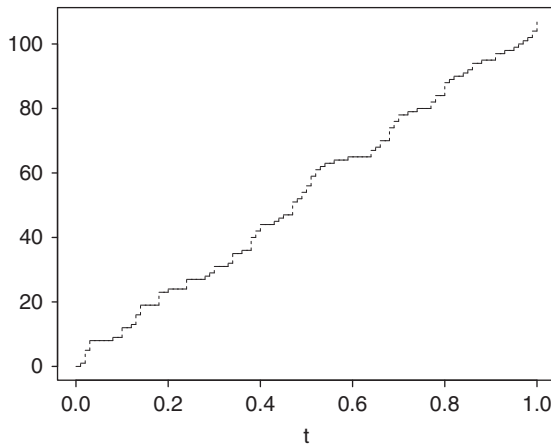


Figure 4. Step function plot of $N(t)$ against t for our simulated data set.

which is a Gamma $(a_1 + r, b_1 + M\Pi(\tau|\theta))$ density. We also determine the posterior of θ at time τ

$$\begin{aligned} \pi_\tau(\theta|D(\tau), \lambda) &\propto L_\tau(D(\tau)|\lambda, \theta)p(\theta) \\ &\propto \theta^{a_2+r-1} \exp\left\{-\theta\left(\sum_{j=1}^r w_j + b_2\right) - \lambda M\Pi(\tau|\theta)\right\} \end{aligned}$$

which is non-standard and so cannot be sampled from directly.

Remark 2 We note that this resembles a Gumbel distribution which has a long tail; the long-tailed nature of this distribution may contribute to the failure of the Gammoid approximation used by Jewel.

We propose a Gibbs updating scheme where each parameter is updated at each iteration. The updating is done via a Gibbs updating scheme with a single update random walk Metropolis sampler

used for θ . For inference on $u(\tau)$, the number of unreported claims at time τ , we note that it is a Poisson random variable with parameter $\lambda M(1 - \Pi(\tau|\theta))$ (see equation (3)). For any time τ , the updating proceeds as follows, given $\lambda^{(i)}$ and $\theta^{(i)}$:

- $\lambda^{(i+1)} \sim \text{Gamma}(a_1 + r(\tau), b_1 + \Pi(\tau|\theta^{(i)}))$
- $\theta^{(i+1)} \sim \pi_\tau(\theta|\lambda^{(i+1)})$
- $u^{(i+1)}(\tau) \sim \text{Poisson}(\lambda^{(i+1)}M(1 - \Pi(\tau|\theta^{(i+1)})))$.

We follow Jewell (1989) and choose $a_1 = 2, b_1 = 0.02, a_2 = 4$ and $b_2 = 6$. For our analysis, we do not estimate the parameters at time $\tau = 0$ since the number of claims at time 0 is 0 a.s.

Remark 3 Generally, we cannot obtain useful posterior estimates until at least one claim is known, since, in this case, no data are available, and we would simply be sampling from the prior distributions. If $D(\tau)$ is empty the likelihood should be defined to be equal to 1 since we would then be conditioning on an empty set. In this case, the posterior for λ and θ are their respective priors. So even though non-observation might have some information we cannot really update our beliefs about λ , the intensity of claims occurrence, which is not the same as claims observed.

Table 2 shows how the posterior means of the parameters λ, θ and the unreported claims $u(\tau)$, change with time. Confidence intervals are shown in parentheses. The fourth column shows the actual number of unreported claims. Note that the number of unreported claims can increase or decrease with time since not all claims occur at time 0.

Table 2. Posterior means of the model parameters as the observation time τ varies.

τ	$\hat{\lambda}_\tau$		$\hat{\theta}_\tau$		$\widehat{u(\tau)}$		$u(\tau)$	Metropolis rate
	Mean	95% HPDI	Mean	95% HPDI	Mean	95% HPDI		
0.5	150.3	48.4, 272.8	0.791	0.285, 1.378	69.0	18, 129	49	0.560
1.0	91.2	34.1, 166.3	0.614	0.192, 1.118	70.5	16, 144	87	0.379
1.5	98.6	46.8, 167.6	0.575	0.202, 0.997	59.3	10, 125	69	0.306
2.0	112.0	61.4, 180.4	0.510	0.203, 0.838	57.2	13, 123	52	0.254
2.5	95.2	60.3, 139.5	0.573	0.249, 0.892	33.6	6, 73	46	0.288
3.0	108.1	71.1, 152.8	0.480	0.237, 0.734	36.0	6, 75	35	0.237
3.5	101.1	72.2, 133.7	0.515	0.287, 0.746	23.8	5, 50	30	0.251
4.0	106.2	77.9, 137.7	0.475	0.276, 0.675	22.4	4, 46	23	0.231
4.5	99.7	76.1, 123.9	0.523	0.341, 0.713	13.7	2, 28	21	0.240
5.0	101.6	79.0, 125.3	0.505	0.335, 0.671	11.5	1, 24	17	0.225
5.5	102.8	81.0, 125.5	0.488	0.337, 0.641	9.9	0, 20	14	0.213
6.0	104.8	83.1, 127.4	0.467	0.329, 0.609	8.8	0, 18	11	0.201
6.5	103.4	82.6, 125.0	0.478	0.351, 0.615	6.5	0, 14	10	0.194
7.0	102.8	82.8, 124.0	0.486	0.360, 0.612	4.8	0, 11	9	0.186
7.5	105.9	85.0, 126.9	0.449	0.337, 0.568	5.0	0, 11	6	0.176
8.0	109.0	88.0, 130.6	0.419	0.308, 0.522	5.1	0, 11	3	0.166
8.5	109.1	88.1, 130.2	0.416	0.315, 0.520	4.3	0, 10	2	0.161
9.0	109.4	88.8, 130.6	0.413	0.315, 0.512	3.6	0, 8	1	0.156
9.5	108.5	88.1, 129.4	0.423	0.330, 0.523	2.7	0, 7	1	0.152
10.0	107.9	87.6, 128.5	0.430	0.333, 0.526	2.0	0, 5	1	0.149

Note: The subscript τ , on λ and θ , indicates that the estimate was obtained using data reported up to that time. HPDI, highest posterior density interval.

Based on the posterior analysis of the results in Table 2, the estimate of θ is decreasing with time. In particular, the posterior estimates of θ decreases from 0.791 to 0.430 as time increases. This is intuitive because claims with smaller delays tend to be reported first, and hence, our initial estimates of the delay time θ^{-1} will be small at first and will increase as time increases and more claims are reported. The estimate of the parameter θ , at time $\tau = 10$, is 0.430(0.048). This yields an average delay time of ~ 2.33 years, which is very close to 2.26 years, the average of the observed delay times.

Figure 5 compares the posterior mean of the simulated values of μ with the actual values at the times shown. Both quantities converge as time increases.

4.3. Claims with no reporting dates

In this section, we illustrate the relative stability of the MCMC method versus the Gammoid approximation of Jewell (1989). For type IV data, we have a count of the number of claims but no reporting dates. Thus for $1 \leq j \leq r(\tau)$, the density of the j th claim is

$$\begin{aligned}
 p(D_j | \theta) &= \Pi(\tau | \theta) \\
 &= \frac{1}{T} \int_{(\tau-T)^+}^{\tau} F(w | \theta) dw
 \end{aligned}$$

which is just the probability that a claim is reported without knowing the actual dates. Substituting the above expression into (4) we get the joint posterior density of λ and θ

$$\pi(\lambda, \theta | D(\tau)) \propto p(\lambda)p(\theta)e^{-\lambda M\Pi(\tau|\theta)} (\lambda M)^r \prod_{j=1}^r \Pi(\tau | \theta)$$

The posterior conditional density of λ is

$$\begin{aligned}
 \pi(\lambda | \theta, r) &\propto p(\lambda)L(D | \lambda, \theta) \\
 &\propto \lambda^{a_1-1} e^{-b_1\lambda} \lambda^r \exp\{-\lambda M\Pi(\tau | \theta)\} \\
 &= \lambda^{a_1+r-1} \exp\{-\lambda(b_1 + M\Pi(\tau | \theta))\}
 \end{aligned} \tag{7}$$

which we identify as the density of a random variable with a Gamma distribution. Likewise, θ has posterior density

$$\begin{aligned}
 \pi(\theta | \lambda, r) &\propto p(\theta)L(D | \lambda, \theta) \\
 &\propto \theta^{a_2-1} e^{-b_2\theta} [\Pi(\tau | \theta)]^r \exp\{-\lambda M\Pi(\tau | \theta)\} \\
 &\propto \theta^{a_2-1} [\Pi(\tau | \theta)]^r \exp\{-(b_2\theta + \lambda M\Pi(\tau | \theta))\}
 \end{aligned} \tag{8}$$

Remark 4 A representation of equation (8) is displayed in Figure 5 which shows the joint posterior distribution of (θ, τ) . We highlight the posterior density of θ versus τ . As τ increases, the dispersion of θ decreases markedly, while its mean increases steadily. Moreover, the density of θ is long tailed for each fixed τ . However, its peakedness increases while the thickness of its tail and skewness decrease as τ increases.

We agree with Jewell (1989) that there is less and less information in Type IV data as τ increases (Figure 6). To see this, note that as τ increases $\Pi(\tau|\theta) \rightarrow 1$ and from (8) we see that $\pi(\theta|\lambda) \rightarrow p(\theta)$, the prior density of θ , for any λ . In the next section, we show that whereas the MCMC estimates

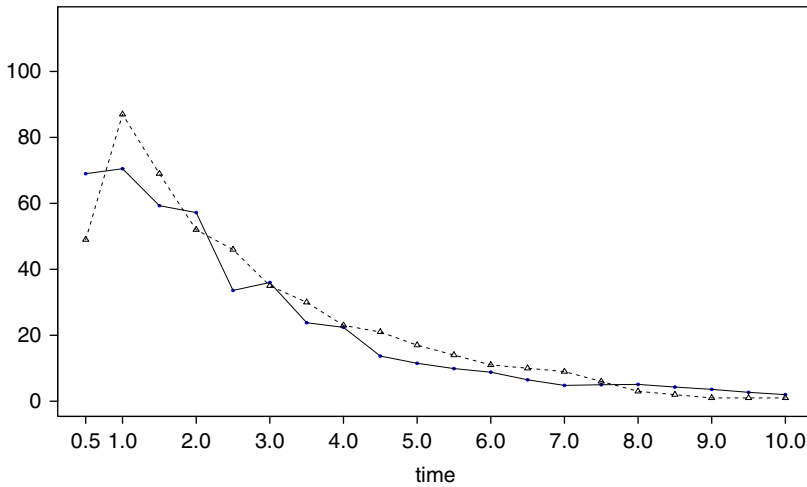


Figure 5. Plot of posterior mean of $\widehat{u(\tau)}$ against time (solid line). The figure also shows the actual number of unreported at the times given (dashed line).

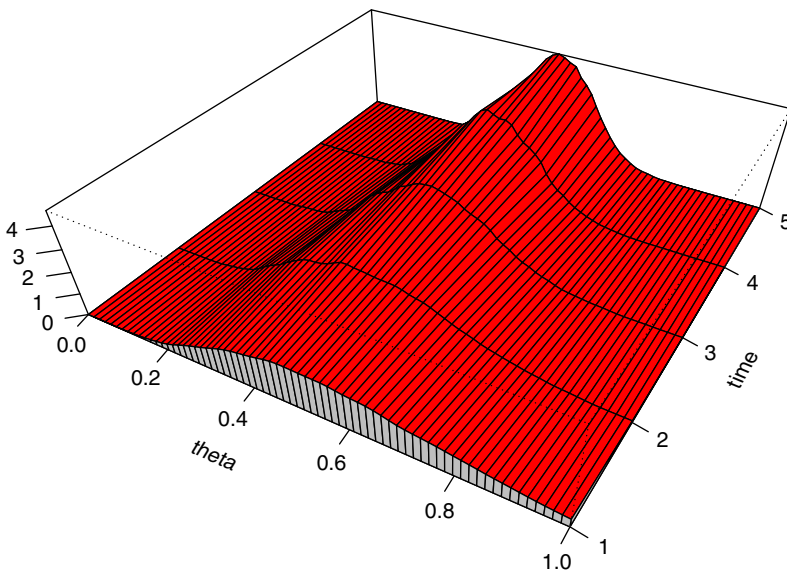


Figure 6. Plot showing the posterior density of θ as τ increases from $\tau = 1$ to $\tau = 5$. As τ increases, and more information is contained in the process $D(\tau)$, the estimate of θ shows less dispersion around its mean value.

approach the prior parameters, the Gammoid approximations grow without bound! It is interesting to note that the above argument holds for any inter-arrival distribution that θ is assumed to follow.

4.4. Comparison of results

Setting $a_1 = 2$, $b_1 = 0.02$, $a_2 = 4$, $b_2 = 6$, we run the MCMC algorithm and then compare our results with those of Jewell (1989). The results for Type IV data are summarised in Tables 3 and 4.

Table 3. τ versus θ , $a_1 = 2$, $b_1 = 0.01$, $a_2 = 4$, $b_2 = 6$.

τ	$r(\tau)$	$\hat{\theta}$	95% HPDI	\hat{u}	95% HPDI	Metropolis rate
0.5	7	0.807	0.274, 1.447	68.6	17, 129	0.833
1.0	20	0.642	0.177, 1.213	69.6	12, 144	0.709
1.5	38	0.661	0.175, 1.244	54.7	6, 123	0.672
2.0	55	0.681	0.187, 1.297	43.7	1, 105	0.670
2.5	61	0.678	0.175, 1.320	32.8	0, 87	0.683
3.0	72	0.680	0.173, 1.309	26.6	0, 77	0.698
3.5	77	0.681	0.173, 1.329	20.9	0, 64	0.715
4.0	84	0.678	0.161, 1.318	17.3	0, 57	0.731
4.5	86	0.668	0.153, 1.291	14.1	0, 49	0.747
5.0	90	0.675	0.152, 1.309	11.5	0, 42	0.761
5.5	93	0.675	0.143, 1.311	9.5	0, 37	0.774
6.0	96	0.664	0.147, 1.290	8.2	0, 33	0.785
6.5	97	0.669	0.153, 1.292	6.6	0, 28	0.796
7.0	98	0.663	0.155, 1.289	5.7	0, 25	0.805
7.5	101	0.668	0.156, 1.315	4.9	0, 23	0.813
8.0	104	0.666	0.145, 1.302	4.3	0, 20	0.821
8.5	105	0.666	0.140, 1.304	3.8	0, 18	0.827
9.0	106	0.668	0.160, 1.327	3.1	0, 16	0.833
9.5	106	0.667	0.153, 1.307	2.7	0, 14	0.839
10.0	106	0.667	0.142, 1.311	2.4	0, 13	0.843

Note: HPDI, highest posterior density interval.

Table 4. Prior sensitivity, $\tau = 4T$.

a_2	b_2	$\hat{\theta}$	HPDI	\hat{u}	HPDI	Metropolis rate
4	6	0.678	0.161, 1.31	17.3	0, 57	0.731
8	14	0.580	0.224, 0.975	17.9	0, 47	0.677
16	30	0.536	0.286, 0.796	18.1	1, 39	0.620
32	62	0.516	0.340, 0.693	18.1	4, 34	0.542
40	78	0.513	0.363, 0.678	18.0	4, 32	0.511
50	98	0.511	0.371, 0.653	17.9	4, 31	0.480
64	126	0.508	0.383, 0.631	17.9	5, 30	0.442
128	258	0.496	0.411, 0.581	18.4	7, 29	0.330

Note: HPDI, highest posterior density interval.

Here we focus on θ , since from (7), we see that the posterior conditional of λ is the same as for Type I data given in equation (6).

Our Table 3 is analogous to table 2 in Jewell (1989). Table 3 shows the changes in the parameter estimates as τ increases. As our previous discussions suggest, as τ increases there is less information contained in the likelihood function. In this case, the prior dominates and the posterior density of θ should be asymptotically equal to the prior density. While this is intuitive, the estimates based on Gammoid approximations grow without bound as τ increases.

Our results are also sensitive to the prior information assumed on the delay parameter θ . Moreover, they show that with even relatively uninformative priors, our algorithm converges quite easily in

some cases where the Gammoid approximations give poor estimations. This is contrary to what is reported in Jewell (1989). In fact, our estimates stabilise as τ increases, whereas those based on the Gammoid approximations grow without bound. Thus, we see that the MCMC method can offer computational advantages over the poor approximation of Jewell. Table 4 shows the effect of prior assumptions on the parameter estimates as a_1 , b_1 , a_2 and b_2 change.

Our results also show that stability is more sensitive to choices of a_2 and b_2 than on values of a_1 and b_1 ; that is, if $p(\theta)$ is informative then $p(\lambda)$ can be diffuse without affecting the results.

The values of a_2 and b_2 were chosen so that the mode $(a_2 - 1)/(b_2)$, of the Gamma prior is the same as the known value θ , hence the (maximum likelihood) estimates would approach the true value of θ . In Table 4, we exclude the case where both a_2 and b_2 are infinitely large.

Another important point to note is that for any choice of inter-arrival distribution time as $\tau \rightarrow \infty$ $\Pi(\tau\theta) \rightarrow 1$ and thus we are sampling from the prior on θ . This is also the case for Type I and Type II data.

5. Summary and Discussion

5.1. Motivating the Gammoid approximations

Jewell (1989) shows that there is no maximum likelihood estimator of λ and θ in the case of Type IV data, and hence no estimator of the number of unreported claims. By assuming a Bayesian formulation however, a solution in terms of Gammoid functions was shown to be possible. The total number of unreported claims at time τ has a Poisson distribution with rate parameter $\lambda[T - M\Pi(\tau\theta)]$, where $M = \min\{\tau, T\}$. This rate is bounded above by $\lambda T[1 - \Pi(T\theta)]$ which is the rate used in Jewell (1989) when the observation time $\tau \geq T$. This includes two types of unreported claims; the first being all which have already occurred and are unreported, and all which are yet to occur. Assuming Gamma priors on θ and λ , it can then be shown that the posterior marginal distribution of u can be approximated by a Gamma-type integral. However, some approximation must be made to obtain the parameters of this Gamma integral. Full details can be found in Jewell (1989).

5.2. MCMC

A natural method for the computation of the posterior quantities λ and θ , and hence also the prediction of unreported claims u , is the MCMC algorithm. In addition to posterior estimates, we are able to compute posterior confidence intervals and determine how useful the Bayesian formulation is. We implemented the model assuming all of our data are of type IV (i.e. no reporting dates or occurrence dates are known) and examined how the choice of the parameters a_2 and b_2 that generate the prior θ affects the outcome. When no information is contained in the likelihood function, the prior should dominate the likelihood. This is clear from our analysis. However, with the Gammoid approximations this is not the case, except where the priors are very informative. Our primary conclusion is that in any Bayesian analysis with the absence of an informative likelihood, the prior distribution dominates the likelihood function; also the prior information presents our best available knowledge about the parameters of interest. Another interesting fact is that these conclusions hold for any waiting time/delay distribution. The type of information available does not affect the estimate of λ , the arrival parameter.

5.3. Assessing convergence of the MCMC algorithm

All our results are based on running the Gibbs algorithm for 100,000 iterations after a burn in of 100,000. To assess convergence we use the method of Gelman (1996). Since we are looking at a continuous-time model, we have to assess convergence for each t that we choose to estimate the model parameters. The last column in each of Tables 2–4 shows the Metropolis acceptance rates for θ .

5.4. Possible extensions to the model

The basic result presented here is that IBNYR events, such as claims, have a Poisson distribution with a reduced arrival parameter if event arrivals are Poisson distributed. This is not entirely surprising and we could also have derived the result using the colouring theorem for Poisson processes (see e.g. Kingman, 1993: 53). There are a plethora of settings where the model can be applied to estimate the true values of the events generating and delay parameters, and then to predict the number of events that are unreported. Typical examples are insurance claims, number of persons infected by an epidemic/disease, the impact of a drug on mortality rate, products under warranty, murders committed by a serial killer or gang, rapes, financial frauds, survey sampling by mail and undetected bugs in computer software.

The model could be extended to incorporate event costs such as claim size/severity, and also letting the parameters θ and λ be time dependent. In this context, let us now introduce a random variable X for the cost/claim size/severity. Thus, each event is now an ordered triple (s, t, X) , where s is the occurrence time, t the reporting date and X the unreported costs (e.g. claim size). One possible extension is to model the joint density of (s, t, X) as $p(s, t, X) \propto f\left(\frac{t-s}{\kappa X^\alpha} \mid \lambda, \theta, \xi\right)$, where λ , θ and ξ are unknown model parameters. The development then proceeds analogously: conditional on time and the model parameters, we derive the likelihood which we can then use for inference. Other forms might allow for more general inhomogeneous Poisson processes such as Markov-modulated Poisson processes.

Acknowledgements

Garfield O. Brown would like to acknowledge the financial support of the Cambridge Commonwealth Trust. Winston S. Buckley would like to thank Nathan Carter, Charles Hadlock and Lucy Kimball for their support and encouragement.

References

- Gelman, A. (1996). Inference and monitoring convergence. In W.R. Gilks, S. Richardson & D. J. Spiegelhalter, (Eds.) *Markov Chain Monte Carlo in Practice* (pp. 131–144). Chapman and Hall, London.
- Jewell, W.S. (1989). Predicting IBNYR events and delays I. Continuous time. *ASTIN Bulletin*, **19**(1), 25–55.
- Jewell, W.S. (1990). Predicting IBNYR events and delays II. Discrete time. *ASTIN Bulletin*, **20**(1), 93–111.
- Kingman, J.F.C. (1993). *Poisson Processes*. Oxford University Press, New York.