MA40198 Report

Ben Johnstone - bfj20, Ben Cox - bc270, Jack Silman – js561

Section 1 – Aim of the Project

The aim of the project is to fit a non-parametric Bayesian smoothing model to a time series of count data. In order to analyse the data a Metropolis Hastings sampler is constructed to simulate from the posterior of θ, our parameter, and $\boldsymbol{x}$ using a combination of random walk and approximate block-Gibbs sampling.

Section 2 – How The Sampler Was Constructed

A suitable hierarchical model for the data is:

$$y_i \sim Po(\lambda(m_i)),\ \lambda(t) = e^{x(t)}$$
$$x(t-1) - 2x(t) + x(t+1) = e(t)$$
$$e(t) \sim N(0, \tau^{-1})$$

A Metropolis Hastings sampler was constructed in order to sample from the posterior of θ and $\boldsymbol{x}$, where $\theta = \log(\tau)$. The process we wish to follow is outlined below:

1) Generate proposals $\theta' \sim q(\theta' | \theta^{j-1})$, $\mathbf{x}' \sim q(\mathbf{x}' | \theta')$

2) Calculate the acceptance probability α:

$$\alpha = \min\left(1, \frac{f(\theta')f(\mathbf{x}'|\theta')f(y|\theta',\mathbf{x}')q(\theta^{j-1}|\theta')q(\mathbf{x}^{j-1}|\theta^{j-1})}{f(\theta^{j-1})f(\mathbf{x}^{j-1}|\theta^{j-1})f(y|\theta^{j-1},\mathbf{x}^{j-1})q(\theta'|\theta^{j-1})q(\mathbf{x}'|\theta')}\right)$$

The prior for θ has an improper uniform density and the proposal for θ is symmetric so we can simplify α such that:

$$\alpha = \min\left(1, \frac{f(\mathbf{x}'|\theta')f(y|\theta',\mathbf{x}')q(\mathbf{x}^{j-1}|\theta^{j-1})}{f(\mathbf{x}^{j-1}|\theta^{j-1})f(y|\theta^{j-1},\mathbf{x}^{j-1})q(\mathbf{x}'|\theta')}\right)$$

3) Generate $u \sim Unif(0,1)$ letting $\theta^j, \mathbf{x}^j = \begin{cases} \theta', \mathbf{x}' & if\ u \le \alpha(\theta', \mathbf{x}' | \theta^{j-1}, \mathbf{x}^{j-1}), \\ \theta^{j-1}, \mathbf{x}^{j-1} & otherwise. \end{cases}$

Firstly an appropriate initial $\theta^{(0)}$ was found by examining the data overlaid with the exponential of 100 samples from the Gaussian approximation to the x-posterior using varying θ values and choosing an initial $\theta^{(0)}$ that looked to fit the data approximately. The initial value chosen was $\theta^{(0)} = 10$.

Now,

$$\alpha = \exp\Big(\min\Big(0, (l(y|\mathbf{x}',\theta') - \log(q(\mathbf{x}'|\theta'))) - (l(y|\mathbf{x}^{j-1},\theta^{j-1}) - \log(q(\mathbf{x}^{j-1}|\theta^{j-1})))\Big)\Big)$$

So we wish to find the log-likelihood function $l(y|\theta,\mathbf{x})$ for $f(\mathbf{x}|\theta)f(y|\theta,\mathbf{x})$, such that $l(y|\theta,\mathbf{x}) = l(\theta|\mathbf{x}) + l(\theta,\mathbf{x}|y)$. (This is what the function 'lf' computes.)

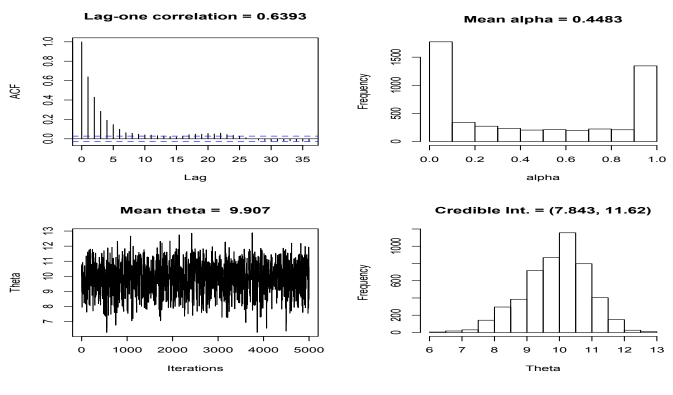
Note that $l(\theta,\mathbf{x} \mid y)$ is simply the sum of the log-density of the Poisson distribution from our hierarchical model for $\lambda = \exp(x(t))$. We derive $l(\theta \mid \mathbf{x})$ below:

$$
\begin{aligned}
& x(t-1) - 2x(t) + x(t+1) = e(t) \sim N(0,\tau^{-1}) \\
& \Rightarrow f(x(t) \mid \theta) = \left(\sqrt{2\pi}\right)^{-1} \sqrt{\tau} \exp(-\frac{\tau}{2}(x(t-1) - 2x(t) + x(t+1)^2) \\
& \Rightarrow f(\mathbf{x} \mid \theta) = \prod_{t=2}^{n-1} f(x \mid \theta) = \left(\sqrt{2\pi}\right)^{-(n-2)} \left(\sqrt{\tau}\right)^{n-2} \exp\left(-\frac{\tau}{2}\sum_{t=2}^{n-1}(x(t-1) - 2x(t) + x(t+1))^2\right) \\
& \Rightarrow f(\mathbf{x} \mid \theta) \propto \tau^{\frac{n-2}{2}} \exp(-\frac{\tau}{2}\mathbf{x}^T Q \mathbf{x}) \\
& \Rightarrow l(\mathbf{x} \mid \theta) = \frac{n-2}{2}\log(\tau) - \frac{\tau}{2}\mathbf{x}^T Q \mathbf{x}
\end{aligned}
$$

A normal centered on the current value of θ is used as the proposal for θ, tuning the variance, σ, to achieve an acceptance ratio $\approx$ 0.44. The joint proposal model given by the multivariate normal $N(\mathbf{E}, RQ^{-1})$ is used as the ***x*** proposal. The function 'qx' finds **E** the mode of ***x*** and the scaled precision matrix Q by optimising $l(y \mid \theta, \mathbf{x})$ using a Newton loop where RQ denotes the cholesky matrix of Q. Approximate block-Gibbs sampling is used by fixing θ and sampling from the joint proposal of ***x*** using the 'rqx' function.

Section 3 – Sampler Efficacy

Plots were produced to see how well the sampler was working. The first plot shows the acf of θ. As we can see from the graph there is a quick drop-off with the correlation at lag-one = 0.639. The second graph shows a histogram of the acceptance values with the mean acceptance value. In order to minimise the correlation we aimed for an acceptance rate $\approx$ 0.44 in accordance with existing literature (as stated in 'Optimal Proposal Distributions and Adaptive MCMC by Jeffrey S. Rosenthal) and we found that after trying a range of different values for the tuning variable, σ, this did indeed minimise the correlation of θ and thus provided better mixing of our chains.



The 3rd graph of the iterations of θ shows that it converges quickly and looks to have good mixing, not getting stuck at any points along the chain. The 4th plot gives a histogram of our θ values giving the 95% computed credible interval for our value of θ.

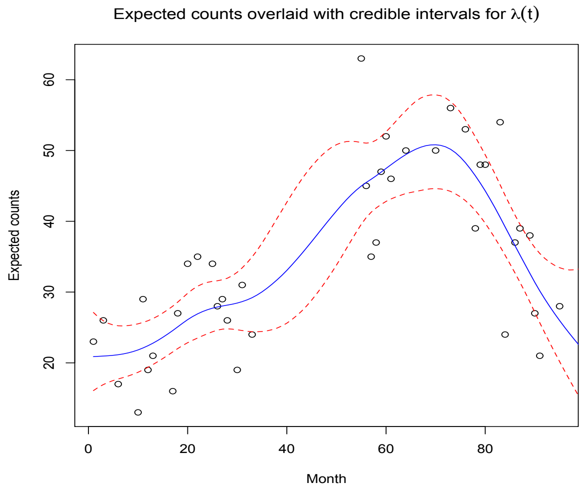
To test for convergence radically different starting points were chosen with $\theta^{(0)} = 1$, and $\theta^{(0)} = 20$ both of which showed convergence to $\theta \approx 9.91$. We also plotted iterations of x[10] and x[40] (not shown) and found these chains had good mixing as well.

Section 4 – The Model Parameter

The burn-in period for θ to converge was relatively small and we found this burn-in period by using the Augmented Dickey Fuller test on θ to test for stationarity. Over repeated runs of the sampler we found that n>150 lead to a stationary time series so we took 150 as our burn-in value.

Taking into account our burn-in value the model parameter θ was found to have mean θ = 9.91 with a 95% credible interval of (7.84, 11.62).

The credible intervals for $\lambda(t)$ were approximated by taking



$$\mu = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i,$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \mu)^2,$$

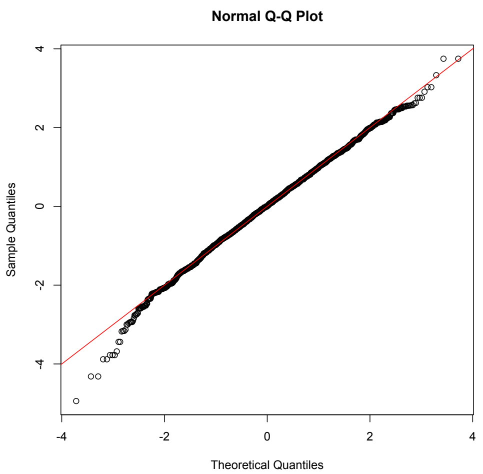
With 95% credible intervals:
$\exp(\mu - 1.96s,\ \mu + 1.96s)$

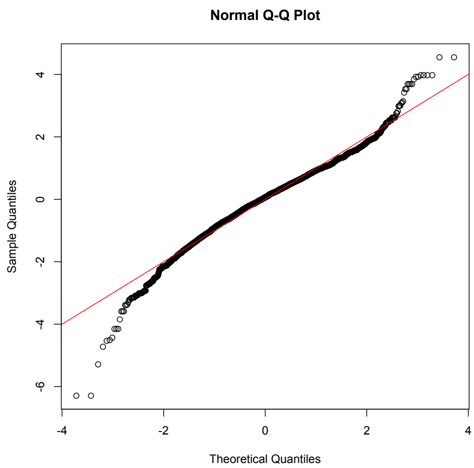
The diagram above shows the expected counts overlaid with upper and lower credible intervals for $\lambda(t)$.

The plot looks to follow the trend of the data fairly well noticing that there is an increase in the width of the credible intervals between months $\in$ [34,54] as would be expected where there is missing data.

Section 5 – Checking Model Assumptions

Normality was assumed for the distribution of the ***x*** vector, however, after analysis at x[10] and x[40] it was found that this was perhaps not valid.

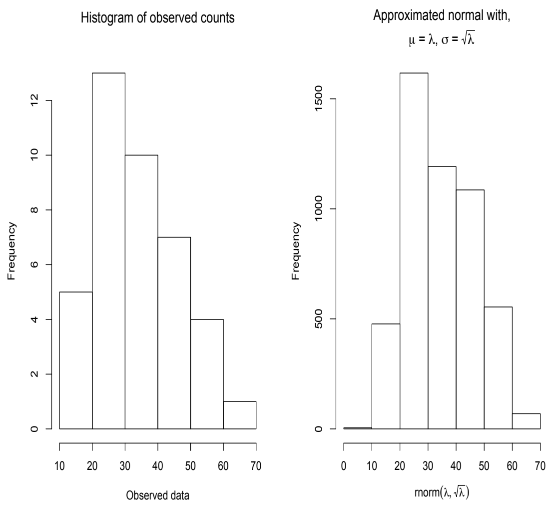


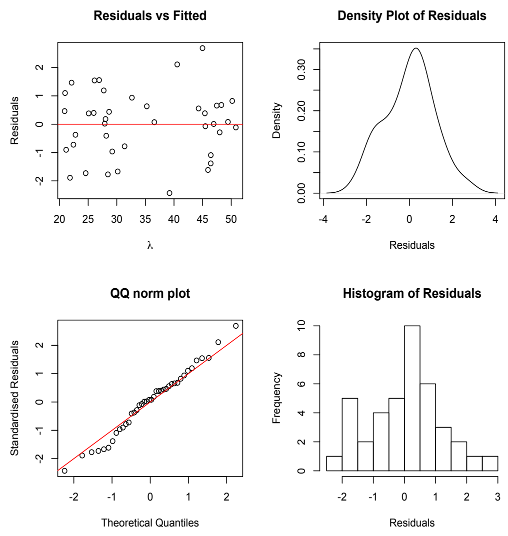


The diagrams above show the Q-Q plots at x[10] and x[40] which are both lightly tailed suggesting that the assumption of normality was not valid. Perhaps for future investigation we would need to model the ***x*** vector under different assumptions.

Considering the model as a whole the Poisson distribution can be approximated to a normal distribution with $N(\lambda, \sqrt{\lambda})$, provided that λ is suitably high. (Greater than 20 is considered reasonable).

$\lambda(t) = \exp(x(t))$, $\min(\lambda) \approx 20$, so a normal approximation of the model is considered.





The normal approximation is taken with $\mu = \lambda$, $\sigma = \sqrt{\lambda}$. The first diagram shows a histogram of the observed data on the left and a histogram of the approximated normal distribution on the right over 5000 iterations. The approximation appears to be a reasonable fit to the data. The residuals of the data are examined to see if they are indeed i.i.d. $N(0,1)$.

In the second diagram the Residual vs. Fitted plot shows the residuals are scattered uniformly along the horizontal axis centered around the mean = 0 showing no signs of heteroscedasticity. The histogram of the residuals follows a relatively normal trend although centered slightly off-centre from the zero-mean and the QQ norm plot is encouraging considering the small sample size although appears to be slightly light on the tails.

Analysis of the residual plots suggests that the initial model assumptions were reasonable but having more observations would be desirable in order to be more confident in our reasoning.

Section 6 - Conclusions

The Poisson model appears to be a good fit overall for our model considering the limited number of observations. Taking θ = 9.91 we could look at predicting future counts for the data. It would have been nice to have more observations and this could be something to bear in mind for future analysis if the experiment is repeated.