Bayesian Estimation of the ETAS Model for Earthquake Occurrences

Gordon J. Ross[a,b,*]

[a]*Institute for Risk and Disaster Reduction, University College London (UCL), UK*
[b]*Department of Statistical Science, University College London (UCL), UK*

Abstract

The Epidemic Type Aftershock Sequence (ETAS) model is one of the best-performing methods for seismic forecasting. However most studies of ETAS use point estimates for the model parameters, which ignores the inherent uncertainty that arises from estimating these from historical earthquake catalogs, resulting in misleadingly optimistic forecasts. In contrast, Bayesian statistics allows parameter uncertainty to be explicitly represented, and fed into the forecast distribution. Despite its growing popularity in seismology, the application of Bayesian statistics to the ETAS model has been limited by the complex nature of the resulting posterior distribution which makes it infeasible to apply on catalogs containing more than a few hundred earthquakes. To combat this, we develop a new framework for estimating the ETAS model in a fully Bayesian manner, which can be efficiently scaled up to large catalogs containing thousands of earthquakes. We also provide easy-to-use software which implements our method.

1. Introduction

The Epidemic Type Aftershock Sequence (ETAS) model is widely used to quantify the degree of seismic activity in a geographical region, and to forecast the occurrence of future mainshocks and aftershocks [10, 9, 12, 11]. The temporal ETAS model is a point process where the probability of an earthquake occurring at time t depends on the previous seismicity H_t, and is defined by the conditional intensity function:

$$\lambda(t|H_t) = \mu + \sum_{t_i < t} \kappa(m_i|K, \alpha) h(t_i|c, p), \quad \kappa(m_i|K, \alpha) = K e^{\alpha(m_i - M_0)}, \quad h(t_i|c, p) = \frac{(p-1)c^{p-1}}{(t - t_i + c)^p}. \quad (1)$$

where the summation is over all previous earthquakes that occurred in the region, with the ith such earthquake occurring at time t_i and having magnitude m_i. The quantity M_0 denotes the magnitude of completeness of the catalog, so that $m_i \geq M_0$ for all i. The temporal ETAS model has 5 parameters: μ controls the background rate of seismicity, K and α determine the productivity (average number of aftershocks) of an earthquake with magnitude m_i, and c and p are the parameters of the Modified Omori Law (which has here been normalized to integrate to 1) and represent the speed at which the aftershock rate decays over time. For notational convenience we will write the unknown parameters as a vector $\theta = (\mu, K, \alpha, c, p)$, and write $Y_i = (t_i, m_i)$ to denote the ith earthquake with $Y = (Y_1, \ldots, Y_n)$ denoting a particular catalog containing n earthquakes.

Most applications of the ETAS model to earthquake catalogs have been carried out within the frequentist statistical framework, where the unknown parameters θ are estimated using the maximum likelihood technique [10, 17]. This results in a single estimated value denoted by $\hat{\theta}$, which is then plugged into Equation 1 and treated as the true value for the purpose of forecasting. However when the ETAS model is estimated using a real earthquake catalog, the estimated value $\hat{\theta}$ will not be exactly equal to the true parameter θ,

and this misspecification can result in unreliable forecasts. Although it is possible to compute the standard error of $\hat{\theta}$ under asymptotic assumptions [18], it is difficult to translate this into information about forecast uncertainty.

Bayesian statistics represents an alternative statistical framework for reasoning about uncertainty, which is becoming increasingly popular in seismology [11, 15, 8, 3],. In the Bayesian paradigm, we do not work with only a single estimate of θ but instead consider the whole posterior distribution $p(\theta|Y)$ which represents our uncertainty about θ based on both the observed earthquake catalog and any prior knowledge we have based on previous studies. This uncertainty can then be incorporated into forecasts in a straightforward manner [5]. However despite its advantages, the Bayesian framework is difficult to apply since the posterior distribution in the ETAS model is highly complex. As such, even studies which attempt Bayesian earthquake forecasting have had to resort to using frequentist-style point estimates for θ, which mitigates the benefits of the Bayesian framework [11, 2]. The only attempt to at providing a fully Bayesian treatment of the ETAS model is an unpublished thesis [16] which proposed using computational simulation based on the framework from [13] for parameter estimation. However as we will show, their approach is likely to provide inaccurate parameter estimates and is not scalable to catalogs containing more than a few hundred earthquakes, which limits its applicability.

The purpose of this article is to introduce a new computational strategy for Bayesian estimation of the ETAS model, which allows for efficient simulation from the posterior distribution. We first show why the direct simulation used in [16] is problematic, and instead derive a new parameterization of the ETAS model which allows for greatly improved estimation. This is based on a latent variable formulation introduced in [17] for the purposes of frequentist estimation, and also studied by [13] and [14]. Comparing this approach to the direct simulation method shows that it provides much more accurate parameter estimates and can also be deployed on seismic catalogs containing thousands of earthquakes. This allows for fully Bayesian forecasting of future seismic activity, based on realistically sized historical catalogs. Finally, we have implemented our estimation procedure in the **bayesianETAS** R package to allow practicing seismologists to automatically fit the Bayesian ETAS model without needing to understand the full details of the below mathematical calculations. We feel that the provision of easy-to-use software is important, since the non-trivial computer programming knowledge required to fit Bayesian models is often one of the factors that limits their widespread adoption.

2. Direct Bayesian Estimation of the ETAS Model

Suppose that the observations $Y = (Y_1, \ldots, Y_n)$ have been generated by a probability model $p(Y_1, \ldots, Y_n|\theta)$ with θ an unknown parameter vector. In Bayesian statistics, we begin with a prior distribution $\pi(\theta)$ which encodes all that is known about θ based on previous studies. In cases where we do not want previous knowledge to affect our analysis, $\pi(\theta)$ can be chosen to be non informative. After analyzing the data, the posterior distribution $p(\theta|Y_1, \ldots, Y_n)$ encodes all information about θ based on both the prior and the data, and is given by:

$$p(\theta|Y) = \frac{p(Y_1, \ldots, Y_n|\theta)\pi(\theta)}{\int p(Y_1, \ldots, Y_n|\theta)\pi(\theta)d\theta}. \tag{2}$$

Knowledge of the posterior distribution allows point estimation of θ to be derived as in the maximum likelihood framework, but also allows all uncertainly about θ to be represented. This uncertainly can then be incorporated into forecasts by simply averaging the forecast distribution over the posterior. We do not intend to give a full treatment of Bayesian inherence here, but an interested reader can consult a standard reference such as [5].

Unfortunately in most real-world situations, the probability model will be too complicated to allow the integral in Equation 2 to be computed analytically. As such, Bayesian inference typically involves using computer simulation to draw M **samples** $\theta^{(1)}, \ldots, \theta^{(M)} \sim p(\theta|Y)$ from the posterior distribution. These samples can then be used to compute all relevant quantities of interest. Although many sampling schemes exist, the most widely used is the independent random walk Markov Chain Monte Carlo (MCMC) method

based on the Metropolis-Hastings algorithm. The simplest version of the MCMC method works as follows. First $\theta^{(1)}$ is initialized to a random value. Then, for each $2 \le k \le M$ a particular parameter $\theta_i^{(k)}$ is selected from the $\theta^{(k)}$ vector, and a new value $\theta_i^{(k+1)}$ is proposed as $\theta_i^{(k+1)} = \theta_i^{(k)} + \epsilon$ where $\epsilon \sim N(0, \sigma_i^2)$. This new value $\theta^{(k+1)}$ is accepted with probability $p(\theta^{(k+1)}|Y)/p(\theta^{(k)}|Y)$. If it is not accepted, then $\theta^{(k+1)}$ is replaced with the previous $\theta^{(k)}$ value. This procedure is then repeated until the desired number M of $\theta^{(k)}$ values have been produced. It can be shown that these can be considered as a sample from the posterior distribution $p(\theta|Y)$. For a fuller description of MCMC, a reference such as [5]. or [6] can be consulted.

For the case of point process models, it can be shown [1] that the likelihood function of the n observations $Y_i = (t_i, m_i)$ observed on the time interval $[0, T]$ is given by:

$$p(Y|\theta) = \prod_{i=1}^{n} \lambda(t_i|H_t, \theta) e^{-\int_0^\infty \lambda(z|\theta, H_z) dz},\tag{3}$$

and plugging in the ETAS intensity from Equation 1 and taking logs gives the loglikelihood of the ETAS model [17]:

$$\log p(Y|\theta) = \sum_{i=1}^{n} \log \left[\mu + \sum_{j=1}^{i-1} \frac{(p-1)c^{p-1}Ke^{\alpha(m_j - M_0)}}{(t_i - t_j + c)^p} \right] - \mu T - \sum_{i=1}^{n} Ke^{\alpha(m_i - M_0)} \left(1 - \frac{c^{p-1}}{(T - t_i + c)^{p-1}} \right).\tag{4}$$

Maximizing this function over θ gives the maximum likelihood estimate $\hat{\theta}$ used in frequentist inference. In the Bayesian framework, we are instead concerned with the posterior $p(\theta|Y) \propto \pi(\theta)p(Y|\theta)$ where $\pi(\theta)$ is the prior which will be discussed later. The normalizing constant of this posterior cannot be computed analytically, and so simulation techniques must be used instead to draw samples from this distribution.. In [16], a simple independent random walk MCMC algorithm was introduced for this purpose. Given current values of the parameters $\theta^{(k)} = (\mu^{(k)}, K^{(k)}, \alpha^{(k)}, c^{(k)}, p^{(k)})$, a new value $\mu^{(k+1)} = \mu^{(k)} + \epsilon$ is proposed where $\epsilon \sim N(0, \sigma_\mu^2)$ and the prior is used to prevent negative values being accepted. This proposal is then accepted or rejected as described above. The other parameters are then updated in a similar way, and the procedure is repeated until the desired number of samples has been drawn. We call this the direct approach to estimation, since it uses a standard MCMC algorithm based on the raw posterior.

Although this procedure is simple and theoretically valid, there are grounds to doubt whether it will actually work well in practice. First, evaluating the likelihood function in Equation 4 is an $O(n^2)$ operation due to the double summation and this evaluation must take place whenever a new parameter value is proposed in the MCMC algorithm. As such, it is computationally very demanding, and cannot feasibly be run on a catalog containing more than a few hundred earthquakes. Second, in a seminal paper [17] studied the performance of frequentist maximum likelihood estimation for the ETAS model based on directly maximizing the likelihood function in Equation 4, and found that the resulting parameter estimates often differed substantially from their true values. This is because the likelihood function is multi-modal and the components of θ are highly correlated. Since MCMC methods can also suffer from serious convergence issues when the parameters are correlated, it is reasonable to believe that this direct MCMC procedure will suffer from the same problem, and in Section 4 we show that this is indeed the case. To avoid both these problems, we now introduce an alternative estimation scheme which allows for more reliable parameter estimation.

3. Latent Variable Formulation

We now develop an alternative sampling posterior scheme based on introducing latent variables. These have the effect of breaking the dependence between the parameters in the likelihood function. We will show that conditional on the latent variables, the parameter sets $\{\mu\}$, $\{K, \alpha\}$ and $\{p, c\}$ are all independent of each other, which greatly improves the convergence of MCMC sampling.

It has previously been shown [13, 17, 14] that the ETAS model can be reinterpreted as a branching process in the following sense. Suppose that the ith earthquake occurs at time t_i, so that $i - 1$ earthquakes

have occurred previously. Equation 1 can be interpreted as showing that the ETAS intensity function at time t_i is a sum of i different Poisson processes. The first is a homogenous Poisson process with intensity μ, while the other $i-1$ each correspond to one of the previous earthquakes. Specifically, for each $1 \leq j \leq i-1$, the earthquake at time t_j triggers an inhomogeneous Poisson process with intensity $\lambda(t) = Ke^{\alpha(m_j-M_0)}(p-1)c^{p-1}(t_i-t_j+c)^{-p}$. Based on standard results about the superposition of Poisson processes [1] we can interpret event t_i as having been generated by a single one of these i processes. We hence introduce the latent branching variables $B = \{B_1, \ldots, B_n\}$ where $B_i \in \{0, 1, \ldots, i-1\}$ indexes the process which generated t_i:

$$B_i \sim \begin{cases} 0 & \text{if } t_i \text{ was produced by the background process (i.e. it is a mainshock)} \\ j & \text{if } t_i \text{ was triggered by the previous earthquake at time } t_j \text{ (i.e. it is an aftershock.)} \end{cases}$$

Conditional on knowing B, we can partition the earthquakes into $n+1$ sets $S_0, \ldots, S_n$ where:

$$S_j = \{t_i; B_i = j\}, \quad 0 \leq j < n$$

so that S_0 is the set of mainshock events which were not triggered by previous earthquakes, and S_j is the set of direct aftershocks triggered by the earthquake at time t_j. It is clear that these sets are mutually exclusive and that their union contains all the earthquakes in the catalog. Additionally, we can see that the earthquakes in set S_0 are generated by a homogenous Poisson process with intensity μ, while the events in each set S_j for $j > 0$ are generated by a single inhomogenous Poisson process with intensity $\lambda(t) = Ke^{\alpha(m_j-M_0)}(p-1)c^{p-1}(t-t_j+c)^{-p}$. Each of these Poisson processes has a likelihood function of the form given in Equation 3 .The ETAS likelihood function from Equation 4 can hence be rewritten (conditional on knowing the branching variables) as:

$$p(Y|\theta, B) = e^{-\mu T} \mu^{|S_0|} \prod_{j=1}^{n} \left(e^{-\kappa(m_j|K,\alpha)H(T-t_j|c,p)} \kappa(m_j|K,\alpha)^{|S_j|} \prod_{t_i \in S_j} h(t_i-t_j|c,p) \right) \tag{5}$$

where $\kappa(\cdot)$ and $h(\cdot)$ are defined in Equation 1, $|S_j|$ denotes the number of earthquakes in the set $|S_j|$, and $H(z) = \int_0^z h(t)dt$. The key point of this reparameterization is that it makes μ independent of the other model parameters in the posterior, while also drastically weakening the dependence between (c, p) and (K, α). Indeed, this dependence is now restricted entirely to their interaction in the $e^{-\kappa(m_j|K,\alpha)H(T-t_j|c,p)}$ term. This greatly improves the performance of the MCMC sampler. In the next section we will discuss how this procedure is carried out.

3.1. Parameter Estimation

Our new MCMC scheme consists of sequentially sampling the parameters in the blocks $\{\mu\}, \{K, \alpha\}, \{c, p\}$ that are now only weakly dependent, given the latent variables B. Since the true values of B are unknown, they must also be estimated within the MCMC scheme. We begin by choosing arbitrary initial values $\left(\mu^{(1)}, K^{(1)}, \alpha^{(1)}, c^{(1)}, p^{(1)}, B^{(1)}\right)$ for the parameters. We then repeatedly sample new values $\left(\mu^{(k+1)}, K^{(k+1)}, \alpha^{(k+1)}, c^{(k+1)}, p^{(k+1)}, B^{(k+1)}\right)$ from the posterior by repeatedly iterating the following four steps:

1. Sample a new value of $B^{(k+1)}$ from $p(B|Y, \theta^{(k)})$. As shown by [19] in a very different context (stochastic declustering), each individual branching variable $B_i^{(k+1)}$ can be sampled exactly from its conditional posterior. Note that each B_i can take values only in the discrete set $\{0, 1, \ldots, j-1\}$, i.e. each earthquake can only be triggered by either a previous earthquake, or the background process. Assuming a uniform prior on each B_i, the probability of it being caused by any of the i processes is simply the proportion of the overall intensity that can be attributed to that process, i.e.:

$$p(B_i^{(k+1)} = j|Y, \theta) = \begin{cases} \dfrac{\mu}{\mu + \sum_{j=1}^{i-1} \kappa(m_j)h(t_j)} & \text{if } j = 0 \\[3mm] \dfrac{\kappa(m_j)h(t_j)}{\mu + \sum_{j=1}^{i-1} \kappa(m_j)h(t_j)} & \text{if } j \in 1, 2, \ldots, i-1 \end{cases}$$

Each B_i can hence be drawn independently from the discrete distribution on $\{0, \ldots, i - 1\}$, with weights given by the above. Note that unlike the MCMC sampler used in [13] for Hawkes processes, this samples $B^{(k+1)}$ exactly from its conditional posterior, which drastically improves computational efficiency.

2. Sample a new value of $\mu^{(k+1)}$ from $p(\mu|Y, \theta, B)$. Using Equation 5 we can see this only depends on the events in the background process S_0:

$$p(\mu|Y, \theta, B) \propto \pi(\mu)e^{-\mu T}\mu^{|S_0|}$$

This is equivalent to estimating the intensity function μ of a homogenous Poisson process on $[0, T]$, with event times S_0. In this case, the Gamma distribution is the conjugate prior: $\pi(\mu) = \text{Gamma}(\alpha_\mu, \beta_\mu)$. The posterior distribution is then $p(\mu|Y, \theta, B) = \text{Gamma}(\alpha_\mu + |S_0|, \beta_\mu + T)$ which can be sampled from directly.

3. Sample new values of $K^{(k+1)}, \alpha^{(k+1)}$ from $p(K, \alpha|Y, \theta, B)$. Using Equation 5, we can see this is given by:

$$p(K, \alpha|Y, \theta, B) \propto \pi(K, \alpha) \prod_{j=1}^{n} e^{-\kappa(m_j|K, \alpha)H(T-t_j|c, p)} \kappa(m_j|K, \alpha)^{|S_j|}$$

which is similar to a standard Poisson regression model with observations $|S_j|$ each with associated predictor m_j and unknown parameters (K, α). Although there is no conjugate prior in this case, it is straightforward to use random walk MCMC to draw a sample from this posterior as described in Section 2

4. Sample new values of $c^{(k+1)}, p^{(k+1)}$ from $p(c, p|Y, \theta, B)$. Using Equation 5, we can see this is given by:

$$p(c, p|Y, \theta, B) \propto \pi(c, p) \prod_{j=1}^{n} e^{-\kappa(m_j|K, \alpha)H(T-t_j|c, p)} \prod_{t_i \in S_j} h(t_i - t_j|c, p)$$

where we can see from Equation 1 that $h(\cdot)$ is the Modified Omori Law rescaled to integrate to 1 (i.e. it is a probability distribution). This is hence equivalent to estimating the parameters of the univariate distribution $h(\cdot|c, p)$ given the $n - |S_0|$ independent and identically distributed observations $t_i - B_{t_i}$ for each $t_i \notin S_0$. Again, this can be done using the basic random walk MCMC sampler (more advanced sampling techniques for univariate distributions could also be used to speed up computation).

4. Performance Analysis

To demonstrate the efficiency of our estimation scheme, we use it to estimate the ETAS model on the relocated Southern Californian earthquake catalog of [7] which can be obtained from the SCEC data center (http://scedc.caltech.edu/research-tools/alt-2011-dd-hauksson-yang-shearer.html). This catalog contains earthquakes from 1981 to 2011 in a rectangular region from 30° to 37.5^o latitude and from -113° to -122° longitude. In order to compare the estimation efficiency on catalogs of difference sizes, we formed a number of subcatalogs containing only earthquakes above magnitude M_0. Specifically, we created catalogs of length $n \in (100, 200, 500, 1000, 2000, 5000)$ by choosing $M_0 \in (5.02, 4.35, 4.04, 3.75, 3.36, 3.06)$. The larger catalog sizes here are more realistic than the 142 earthquake catalog considered by [16] when testing their direct MCMC scheme.

We use non informative priors for the ETAS parameters: μ is given a conjugate Gamma(0.1, 0.1), and improper uniform priors are placed on $\log K$ and $\log \alpha$. For c and p, [8] show in the context of the Omori law that improper priors do not lead to proper posteriors and recommends a Uniform prior on a restricted interval. We follow their suggestion and assign c a Uniform prior on $[0, 8]$, and p a Uniform prior on $[1, 8]$. For the proposal standard deviations used in the direct MCMC scheme, we chose values $\sigma_\mu = 0.05, \sigma_K =$

	DIrect MCMC					Latent Variables				
n	μ	K	α	c	p	μ	K	α	c	p
100	0.36	0.70	0.32	0.25	0.32	0.03	0.16	0.03	0.02	0.06
200	1.86	3.59	0.87	1.58	10.26	0.18	1.3	0.09	0.19	0.69
500	7.49	10.23	6.92	13.49	20.62	0.97	2.24	0.34	2.43	6.10
1000	18.74	29.49	24.96	91.20	145.96	2.27	3.85	1.21	2.07	6.36
2000	337.01	96.56	96.02	290.80	342.06	2.38	2.72	2.93	2.32	2.95
5000	2384.03	1143.01	1086.84	4594.58	1702.90	11.38	15.07	17.72	16.94	17.57

Table 1: Number of minutes required to draw samples equivalent to an effective sample size of 200 when running on a catalog of length n. For example when the catalog contains 5,000 earthquakes, the direct scheme requires 2384 minutes (39 hours) to produce 200 roughly independent samples of the μ parameter, compared to only 11.38 minutes when using the latent variable approach

$0.15, \sigma_\alpha = 0.15, \sigma_c = 0.25, \sigma_p = 0.30$ based on a short pilot run of the simulation. These values result in an acceptance rate of between 20% and 30%, which is thought to be optimal [4].

Since the MCMC algorithm uses a random walk scheme, the resulting samples can be highly correlated which means they cannot be considered as independent draws from the posterior. This means that the samples produced will be equivalent to a far smaller number of independent samples. As such, our main performance metric is the effective sample size (ESS) which measures how many independent samples the MCMC draws are equivalent to. Typically, a few hundred independent samples from the posterior are required for accurate inference.

Table 1 shows the number of minutes that both the latent variable and direct schemes require to produce 200 effective samples for each catalog length, using the highly optimized C++ code from our **bayesianETAS** package, running on a Macbook Pro with an i7 2.4Ghz processor. It can be seen that as the catalog size increases, the direct method takes longer and longer to produce samples since carrying out a $O(n^2)$ likelihood evaluation per proposal, combined with the high degree of correlation in the parameters, seriously limits scalability. For the catalog containing 5,000 earthquakes, it takes around 4594 minutes (76 hours) to draw samples of all parameters equivalent to an effective sample size of only 200. In contrast, the latent variable scheme requires under 18 minutes to do the same. This is roughly around a 2500% improvement, and the size of the improvement grows with the length of the catalog. For catalogs containing more than 1,000 earthquakes, the direct scheme is hence not computationally feasible while the latent variable approach can provide a high number of posterior samples in a reasonable length of time even for large catalogs.

To illustrate further, we consider a particular simulation run on the 5,000 earthquake catalog. For both the direct MCMC scheme and our latent variable formulation, 5,500 samples of each parameter were drawn from the posterior for each simulated catalog, with the first 500 treated as a burn-in period and discarded. The direct method took 202 minutes to complete, and resulted in an ESS of $(17, 35, 37, 9, 24)$ for the parameters (μ, K, α, c, p) respectively. while the latent variable method took 54 minutes to produce an ESS of $(958, 723, 615, 643, 621)$. As such, we can see that the performance improvement is in both the overall running time, and the number of effective samples. In fact, the low number of effective samples for the direct MCMC scheme is unlikely to allow for quantities such as forecast uncertainty to be computed with accuracy. To highlight this, Figure 1 plots a kernel density estimate of the resulting posterior distribution as computed by both methods. It can be seen that the latent variable method produces a smooth posterior distribution which is expected given the much larger number of effective samples, while the direct approach suffers from high variability and multimodality, and also underestimating the variance of c and p. We also note in passing how much more information the full posterior distributions give compared to the simple maximum likelihood estimate (shown in blue) used in most non-Bayesian studies, which can then be translated into more nuanced forecasting of future seismicity..
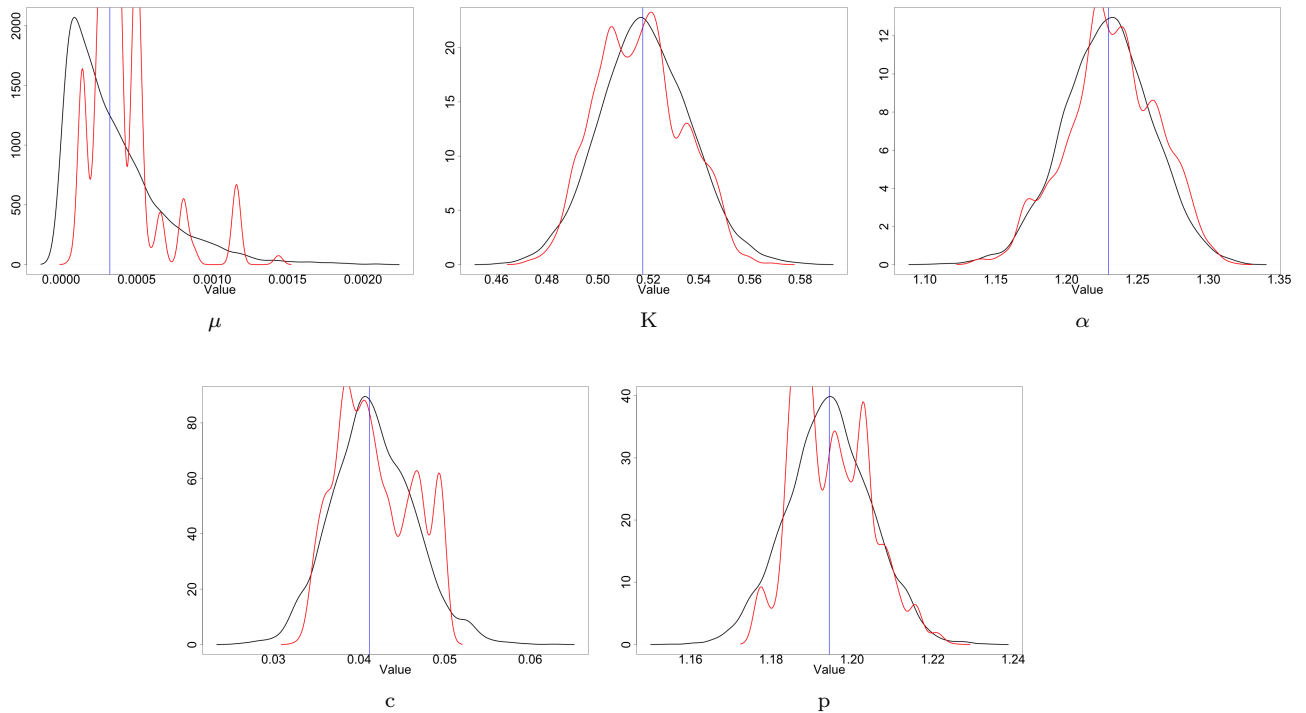
Figure 1: Posterior distributions for each of the 5 ETAS parameters on the simulated catalog using the direct MCMC method (red) and the latent variable approach (black). The maximum likelihood estimate of each parameter is shown in blue.

5. Concluding Remarks

The classic frequentist methods commonly used in seismic forecasting typically assume that model parameters are known exactly, which can result in forecasts which are overly confident. To mitigate this, Bayesian approaches are becoming more common in seismology. However despite being one of the most popular forecasting models, the ETAS framework has rarely received a fully Bayesian treatment, and even studies which attempt Bayesian forecasting end up resorting to frequentist-style plug-in estimates of the model parameters [11, 2].. This is due to the highly complex nature of the posterior distribution. In this work, we have introduced a new posterior sampling scheme which can be scaled up to catalogs containing thousands of earthquakes, and demonstrated its efficiency. Our approach can easily be deployed on realistic seismic catalogs containing thousands of earthquakes.

6. Supplementary Information

Computer code implementing the Bayesian estimation framework introduced in this paper has been written in the language R, and is now available from CRAN along with detailed instructions: https://cran.r-project.org/web/packages/bayesianETAS/index.html

7. References

[1] D. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume I.* Springer, New York, 2003.
[2] H. Ebrahimian, F. Jalayer, D. Asprone, A. M. Lombardi, W. Marzocchi, A. Prota, and G.Manfredi. Adaptive daily forecasting of seismic aftershock hazard. *Bulletin of the Seismological Society of America*, 104(1):145–161, 2014.
[3] L. Faenza, C. Meletti, and L. Sandri. Bayesian inference on earthquake size distribution: A case study in Italy. *Bulletin of the Seismological Society of America*, 100(1):349–363, 2010.

[4] A. Gelman, W. R. Gilks, and G. O. Robetrs. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, 7(1):110–120, 1997.

[5] M. Glickman and D. van Dyk. Basic Bayesian methods. *Methods in Molecular Biology*, 404:319–338, 2003.

[6] G. Hamra, R. MacLehose, and D. Richardson. Markov Chain Monte Carlo: An introduction for epidemiologists. *International Journal of Epidemiology*, 42(2):627–634, 2013.

[7] E. Hauksson, W. Yang, and P. Shearer. Waveform relocated earthquake catalog for southern california (1981 to 2011),. *Bulletin of the Seismological Society of America*, 102:2239–2244, 2012.

[8] M. Holschneider, C. Narteau, P. Shebalin, Z. Peng, and D. Schorlemmer. Bayesian analysis of the modified Omori law. *Journal of Geophysical Research*, 117(86):B06317,, 2012.

[9] W. Marzocchi and A. M. Lombardi. Real-time forecasting following a damaging earthquake. *Geophysical Research Letters*, 36(21), 2009.

[10] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

[11] T. Omi, Y. Ogata, Y. Hirata, and K. Aihara. Intermediate-term forecasting of aftershocks from an early aftershock sequence: Bayesian and ensemble forecasting approaches. *Journal of Geophysical Research*, 120(4):2561–2578, 2016.

[12] T. Omi, Y. Ogata, Y. Hirata, and K. Alihara. Estimating the ETAS model from an early aftershock sequence. *Geophysical Research Letters*, 41:850–857, 2014.

[13] J. G. Rasmussen. Bayesian inference for Hawkes processes. *Methodology and Computing in Applied Probability*, 15(15), 2013.

[14] G. J. Ross. The nonparametric Bayesian Hawkes process with application to modeling human activity patterns. Technical report, Technical Report, 2016.

[15] R. Shcherbakov. Bayesian confidence intervals for the magnitude of the largest aftershock. *Geophysical Research Letters*, 41(18):6380–6388, 2014.

[16] N. Vargas. Bayesian point process modelling of earthquake occurrences. Technical report, Ruprecht-Karls University Heidelberg, 2012.

[17] A. Veen and F. P. Schoenberg. Estimation of space-time branching process models in seismology using an EM-type algorithm. *Journal of the American Statistical Association*, 103:614–626, 2008.

[18] Q. Wang, F. Schoenberg, and D. Jackson. Standard errors of parameter estimates in the ETAS model. *Bulletin of the Seismological Society of America*, 106(5):1989–2001, 2010.

[19] J. Zhuang, Y. Ogata, and D. Vere-Jones. Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380, 2002.