

The School of Mathematics



THE UNIVERSITY
of EDINBURGH

**Research on the Reliability of
Earthquake Forecast Using the ETAS
Model**

by

Haoyang Jiao

Dissertation Presented for the Degree of
MSc in Statistics with Data Science

July 2024

Supervised by
Dr Finn Lindgren, Dr Francesco Serafini and Dr Xindi Song

Executive Summary

This research paper investigates the accuracy and reliability of the Epidemic-Type Aftershock Sequence (ETAS) model, a specific instance of the Hawkes process model, in forecasting future earthquake events. The study utilizes the R-package ETAS.inlabru developed at the University of Edinburgh and focuses on synthetic data generated from known parameters to control for confounding factors.

Key findings:

- Model Performance: The ETAS model demonstrates good performance in forecasting future earthquake events, with an average deviation of 3.65
- Data Quality Impact: Catalogs containing more diverse high-magnitude events or higher-magnitude events improve parameter estimation reliability, particularly for the α parameter.
- Initial Parameter Values: Correct setting of initial values for certain parameters (μ and p , or combinations including α , c , and K) leads to more accurate posterior distributions of all model parameters.
- Catalog Characteristics: Larger catalogs with more events, especially large events, result in more reliable parameter estimation across all model parameters.
- Forecasting Accuracy: The model's forecast precision gradually decreases as the forecasting period extends.
- Model Reliability: The near-identical empirical distribution of "time between events" in synthetic catalogs from estimated and known parameters indicates high reliability of the ETAS model fitted using the INLA method.

This research contributes to understanding the ETAS model's effectiveness in earthquake forecasting and identifies key factors influencing its reliability. These findings have significant implications for improving earthquake prediction capabilities, which are crucial for disaster prevention, economic security, and social stability.

Acknowledgments

I am grateful to the supervisors of this project, Dr Finn Lindgren, Dr Daniel Paulin, Dr Xindi Song and Dr Man Ho Suen, for their support and valuable advice. I would also like to thank Dr Francesco Serafini for his articles and tutorials, which give me a lot of help.

University of Edinburgh – Own Work Declaration

This sheet must be filled in, signed and dated - your work will not be marked unless this is done.

Name: Haoyang Jiao

Matriculation Number: s2333015

Title of work: Research on the Reliability of Earthquake Forecast Using ETAS Model

I confirm that all this work is my own except where indicated, and that I have:

- Clearly referenced/listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Not sought or used the help of any external professional academic agencies for the work
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Complied with any other plagiarism criteria specified in the Course handbook

I understand that any false claim for this work will be penalised in accordance with the University regulations (<https://teaching.maths.ed.ac.uk/main/msc-students/msc-programmes/statistics/data-science/assessment/academic-misconduct>).

Signature

A handwritten signature in black ink, appearing to read "Haoyang Jiao". The signature is fluid and cursive, with "Haoyang" on the first line and "Jiao" on the second line.

Date 2024-07-09

Contents

1	Introduction	1
2	Methods	2
2.1	Models	2
2.2	Techniques	3
3	Results	4
3.1	Relationship between estimates of α and number of different high magnitude events . .	4
3.2	Influence of correctly specifying intial parameters value on posterior distributions of parameters	5
3.3	Relationship between sequence characteristics and posterior distributions of parameters	10
3.4	Forecast accuracy	12
3.5	Empirical distribution of “time between events”	16
4	Conclusion	18
5	Appendices	19

List of Tables

1	Statistics about forecast accuracy	15
2	The absolute percentage deviation and standard deviation of posterior distributions of parameters when one parameters' intial values were correctly set	19
3	The absolute percentage deviation and standard deviation of posterior distributions of parameters when two parameters' intial values were correctly set	20
4	The absolute percentage deviation and standard deviation of posterior distributions of parameters when three parameters' intial values were correctly set	21
5	The absolute percentage deviation and standard deviation of posterior distributions of parameters when four parameters' intial values were correctly set	22

List of Figures

1	Posterior distribution of alpha versus number of different high magnitude events and the level of magnitudes	4
2	Posterior distributions of parameters when setting one of them fixed to the values used when generating the data	5
3	Posterior distributions of parameters when setting two of them fixed to the values used when generating the data	7
4	Posterior distributions of parameters when setting mu and p fixed to the values used when generating the data	8
5	Posterior distributions of parameters when setting three of them fixed to the values used when generating the data	9
6	Posterior distributions of parameters when setting three of them (alpha, c, p or alpha, mu, p or c, mu, p) fixed to the values used when generating the data	10
7	Posterior distributions of parameters when setting four of them fixed to the values used when generating the data	11
8	Relationship between posterior distributions of parameters and number of events . . .	11
9	Relationship between posterior distributions of parameters and number of large events	12
10	Synthetic catalogue	13
11	Forecast and true cumulative count of events	14
12	Forecast precision of ETAS model	15
13	Empirical CDF of "time between events" in synthetic catalogue from estimated vs known parameters	16
14	Empirical PDF of "time between events" in synthetic catalogue from estimated vs known parameters	17

1 Introduction

Accurately predicting future earthquake events through the integration of geophysical insight and statistical modeling is of great importance for improving the disaster prevention and mitigation capabilities of the nation and individuals, protecting economic security, maintaining social stability, and promoting related scientific research. When constructing such an earthquake forecasting model for a specific region, factors to be considered include the local geological structure, historical earthquake records, and the potential patterns about how “mainshocks” generate “aftershocks”.

The Poisson process model is a classical statistical model applied to time and spatial point data, which assumes the data points are independent, conditionally on an intensity process. However, what we are more concerned about is the ability to predict future “aftershocks” based on the historical “mainshocks”, so we need to extend the Poisson process model by using the Hawkes process model, which can predict the magnitude and number of future earthquakes based on the magnitude and number of past earthquakes.

The focus of this paper’s research is the accuracy and reliability of the Hawkes process model in forecasting future earthquake events. Specifically, this paper, based on the R-package `ETAS.inlabru` developed at the University of Edinburgh for the Epidemic-Type Aftershock Sequence (ETAS) model, which is a particular instance of the Hawkes process model, investigates the number of different high magnitude events and the level of these magnitudes required for effective estimation of model parameters, the impact of the initial values of model parameters on the reliability of the estimated parameter values, the relationship between the posterior distribution of model parameters and the characteristics of the earthquake sequence, the accuracy of the model’s prediction of the number of future earthquake events, and the difference in the empirical distribution of “time between events” between synthetic data from known parameters and synthetic data from estimated parameters. By addressing these questions, we can clearly determine the accuracy and reliability of the ETAS model in forecasting future earthquake events, as well as identify the types of earthquake sequence data that is needed to provide reliable estimates of the model parameters.

It is important to note that in the research process, we did not use real data, but rather assumed the model parameters were known beforehand, and then used these parameters to simulate a large number of synthetic catalogs, and assumed these parameters and the generated earthquake data were true. The advantage of this approach is that it allows us to control for additional confounding factors during the analysis. By varying the specified influencing factors, we can observe the changes in the posterior distribution of the model parameters, effectively eliminating the influence of other factors. This enables us to clearly determine the impact of the specified influencing factors on the posterior distribution of the model parameters.

2 Methods

2.1 Models

The Epidemic-Type Aftershock Sequence (ETAS) model belongs to the family of Hawkes point processes. The temporal Hawkes process is a point process model with conditional intensity given by

$$\lambda(t|H_t) = \mu + \sum_{t_h \in H_t} g(t - t_h)$$

where H_t is the history of the process up to time t . Generally speaking, H_t contains all the events occurred before t . The quantity $\mu > 0$ is usually called the background rate, and is interpreted as the rate at which events occur spontaneously. The function $g(t - t_h)$ is called triggering function and measures the influence of having an event in t_h on time t . If we look at $g(t - t_h)$ as a function of t is the intensity of the point process representing the offspring of the event in t_h . In essence, an Hawkes process model can be seen as the superposition of a background process with intensity μ and all the aftershock processes generated by the observations in H_t each one with intensity $g(t - t_h)$. This makes Hawkes process model particularly suitable to describe phenomena in which an event has the ability to trigger additional events, phenomena characterized by cascades of events such as earthquakes, infectious diseases, wildfires, financial crisis, and similar.

The ETAS model is a particular instance of Hawkes process model which has proven to be particularly suitable to model earthquake occurrence. Earthquakes are usually described and modelled as marked time points where the marking variable is the magnitude of the event. So the history of the process is composed by time-magnitude pairs $\{(t_h, m_h), h = 1, \dots, N_h\}$. Various slightly different ETAS formulations exist, usually characterized by slightly different triggering functions, the one that is implemented in the `ETAS.inlabru` R-package has conditional intensity given by

$$\lambda(t|H_t) = \mu + \sum_{t_h \in H_t} K \exp\{\alpha(m_h - M_0)\} \left(\frac{t - t_h}{c} + 1\right)^{-p}$$

where M_0 is the cutoff magnitude such that $m_h \geq M_0$ for any h . This value is decided a priori based on the quality of the catalogue used.

The parameters of the model are:

- $\mu \geq 0$, the background rate.
- $K \geq 0$, a general productivity parameter, it plays a role in determining the average number of aftershocks induced by any event in the catalogue.
- $\alpha \geq 0$, a magnitude scaling parameter, it determines how the number of aftershocks changes based on the magnitude of the event generating the aftershocks. It has to be non-negative to reflect the fact that stronger earthquakes generate more aftershocks.
- $c > 0$, a time offset parameter, smaller values are associated with catalogues with fewer missing events.
- $p > 1$, an aftershock decay parameter, it determines the rate at which the aftershock activity decreases over time. It has to be greater than 1 otherwise an event may generate an infinite number of aftershocks in an infinite interval of time which is thought to be unphysical.

As any other Bayesian analysis we need to decide the priors of the parameters. The approximation method that we use considers the parameters in two different scales: the original ETAS scale, and the internal scale. The internal scale is used by `ETAS.inlabru` package to perform the calculations. In the internal scale the parameters does not have any constraint and have a standard normal distribution as prior. We need to set up the priors for the parameters in the ETAS scale. This is done by considering

a copula transformation $\eta(X)$ such that if $X \sim N(0, 1)$ then, $\eta(X)$ has the desired distribution. For all models fitted in this paper, the prior distributions of the parameters were set as:

$$\mu \sim \Gamma(0.3, 0.6)$$

$$K, \alpha, c \sim Unif(0, 10)$$

$$p \sim Unif(1, 10)$$

The true values of the ETAS parameters were set as $\mu = 0.301, K = 0.136, \alpha = 2.439, c = 0.071, p = 1.178$.

2.2 Techniques

1 Generate catalogues

When simulating the data, we are actually generating synthetic catalogues from a temporal ETAS model with fixed parameters spanning a given interval of time. The information that needs to be considered includes the true values of the ETAS parameters $\boldsymbol{\theta}$, the parameter of the magnitude distribution β , the cutoff magnitude M_0 , the starting time and end time of the catalogues and the set of known events. Unless otherwise specified, we uniformly set β to 2.353, M_0 to 2.5, and the starting time and end time of the catalogues to 0 and 365, respectively.

2 Hawkes process log-likelihood approximation

Hawkes process log-likelihood approximation technique allows us to express the Hawkes process log-likelihood as a sum of linear functions of the parameters $\boldsymbol{\theta}$. The log-likelihood can be decomposed into three main components:

$$L(\boldsymbol{\theta}) = -\Lambda_0(\chi) - \sum_{h=1}^n \Lambda_h(\chi) + SL(H)$$

The expected number of background events $\Lambda_0(\chi)$, the expected number of induced events $\sum_{h=1}^n \Lambda_h(\chi)$, and the sum of the log-intensities $SL(H)$.

The technique used for Hawkes process log-likelihood approximation is based on approximating these three components separately. The approximation in `ETAS.inlabru` package is performed with respect to the mode of the posterior distribution $\boldsymbol{\theta}^*$, which is determined by an iterative algorithm. The algorithm starts from a linearisation point $\boldsymbol{\theta}_0^*$, finds the mode of the linearised (with respect to $\boldsymbol{\theta}_0^*$) posterior using the INLA method, namely $\bar{\boldsymbol{\theta}}_1^*$, the value of the linearisation point is updated to $\boldsymbol{\theta}_1^* = \alpha \boldsymbol{\theta}_0^* + (1 - \alpha) \bar{\boldsymbol{\theta}}_1^*$. This process is repeated until, for each parameter, the difference between two consecutive linearization points is less than 1% of the marginal posterior standard deviation. The value 1% is the default value used by the R-package `ETAS.inlabru`. In addition, if the number of iterations exceeds the set maximum number of iterations, it means the algorithm has not converged. Unless otherwise specified, the initial value of parameters $\boldsymbol{\theta}_0^*$ were set as $\mu_0 = 0.5, K_0 = 0.1, \alpha_0 = 1, c_0 = 0.1, p_0 = 1.1$, and the maximum number of iterations was set to be 100. During the iteration process, the parameters of the binning strategy used were uniformly set to $\delta = 1, \Delta = 0.1, n_{max} = 5$.

3 Results

The following are five research questions of interest to me. Unless otherwise specified, the setting used are the same as those mentioned earlier, such as the model parameters used for generating synthetic catalogs, the parameter of the magnitude distribution β , the cutoff magnitude M_0 , the time range, and the parameters of the binning strategy δ, Δ, n_{max} used for fitting the model.

3.1 Relationship between estimates of α and number of different high magnitude events

Theoretically speaking, increasing the number of different high magnitude events will lead to better estimates of α (which regulates how the number of event offspring scales with the magnitude), as opposed to having the same number of high magnitude events but all with the same magnitude.

To study how the posterior distribution for α changes with the number of different high magnitude events, and the level of these magnitudes, I conducted several simulation experiments. Specifically, I defined three categories of historical known events, corresponding to 1, 2, and 3 events, respectively. The 1 event occurred at the midpoint of the catalogue's starting time and end time, i.e., $\frac{0+365}{2} = 182.5$; the 2 known events occurred at the $\frac{1}{3}$ and $\frac{2}{3}$ quantiles of the catalogue's starting time and end time, i.e., 121.6667 and 243.3333; the 3 known events occurred at the $\frac{1}{4}, \frac{2}{4}$, and $\frac{3}{4}$ quantiles of the catalogue's starting time and end time, i.e., 91.25, 182.50, and 273.75. The magnitudes were set to 4, 4.25, 4.50, ..., 5.75, 6, for each case, I randomly sampled some random numbers from the range $[magnitude - 0.25, magnitude + 0.25]$, where the number of random numbers was equal to the number of historical known events. This resulted in 27 historical known events and, consequently, 27 synthetic catalogues, for which 27 ETAS models were fitted. The experimental results showed that the algorithm of each ETAS model converged.

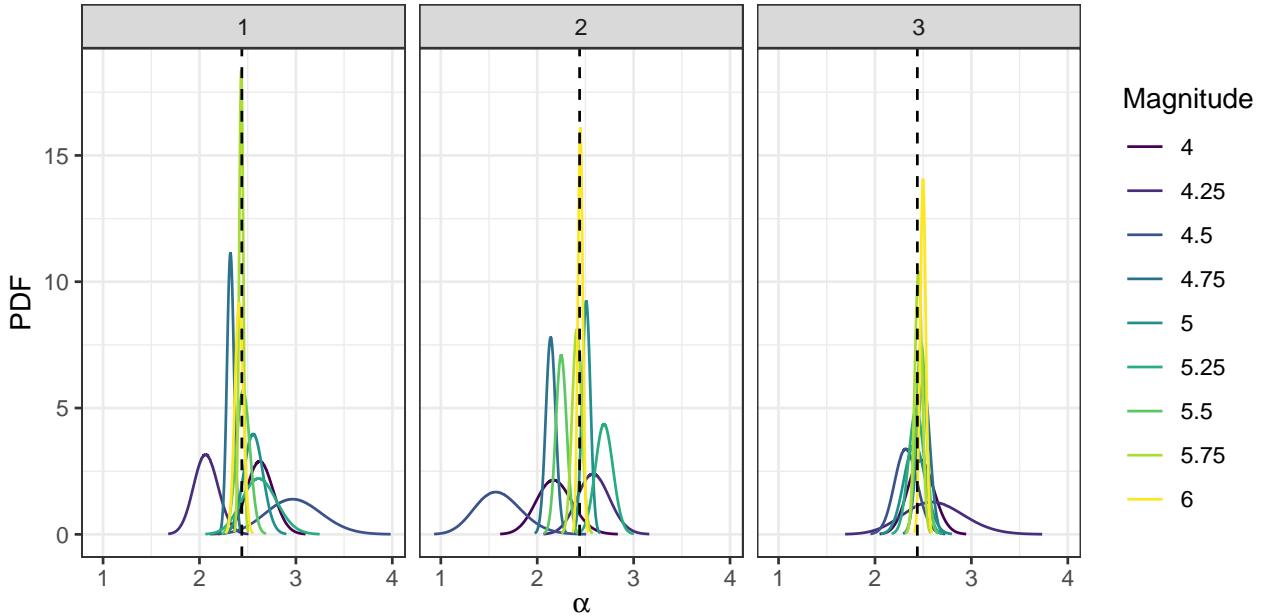


Figure 1: Posterior distribution of alpha versus number of different high magnitude events and the level of magnitudes

The posterior distributions of α for each scenario are shown in Figure 1. The 3 panels in Figure 1 correspond to different numbers of historical known events, and different colors correspond to different magnitudes, with colors tending towards yellow indicating higher magnitudes and colors tending towards deep purple indicating lower magnitudes. The vertical black dashed lines represent the true value of α . On one hand, regardless of the number of historical known events, within the same panel,

as the level of magnitudes increases, the standard deviation of the posterior distribution of α exhibits a decreasing trend, and their means seem to be closer to the true parameter value. On the other hand, given a certain magnitude, especially a high magnitude, across different panels, as the number of historical known events increases, the standard deviation of the posterior distribution of α seems to decrease, and their means seem to be closer to the true value of α .

3.2 Influence of correctly specifying intial parameters value on posterior distributions of parameters

To understand which parameters are the ones that we need “to get right” in order to estimate the others correctly, I will compare the posterior distributions when we fit the model on all parameters or when considering some of them fixed to the values used when generating the data, i.e., the true parameter values.

Specifically, I first generated a synthetic catalogue without known events using the true parameters, and then, on this catalogue, I set all the initial parameter values to the true parameter values and ran the INLA algorithm to obtain the posterior distributions of all parameters when they are all correctly specified. Next, I correctly specified 0, 1, 2, 3, and 4 of the initial parameter values, where the incorrect values were set to three times the true values, and ran the INLA algorithm for each case to obtain the posterior distributions under different correctly specifying scenarios. In total, I fitted 32 ETAS models.

Due to the large number of models involved, to study the impact of correctly specifying the initial parameter values on the posterior distributions more clearly, I will present the posterior distributions in 4 separate figures, corresponding to the cases where 1, 2, 3, and 4 of the parameters are correctly specified, respectively.

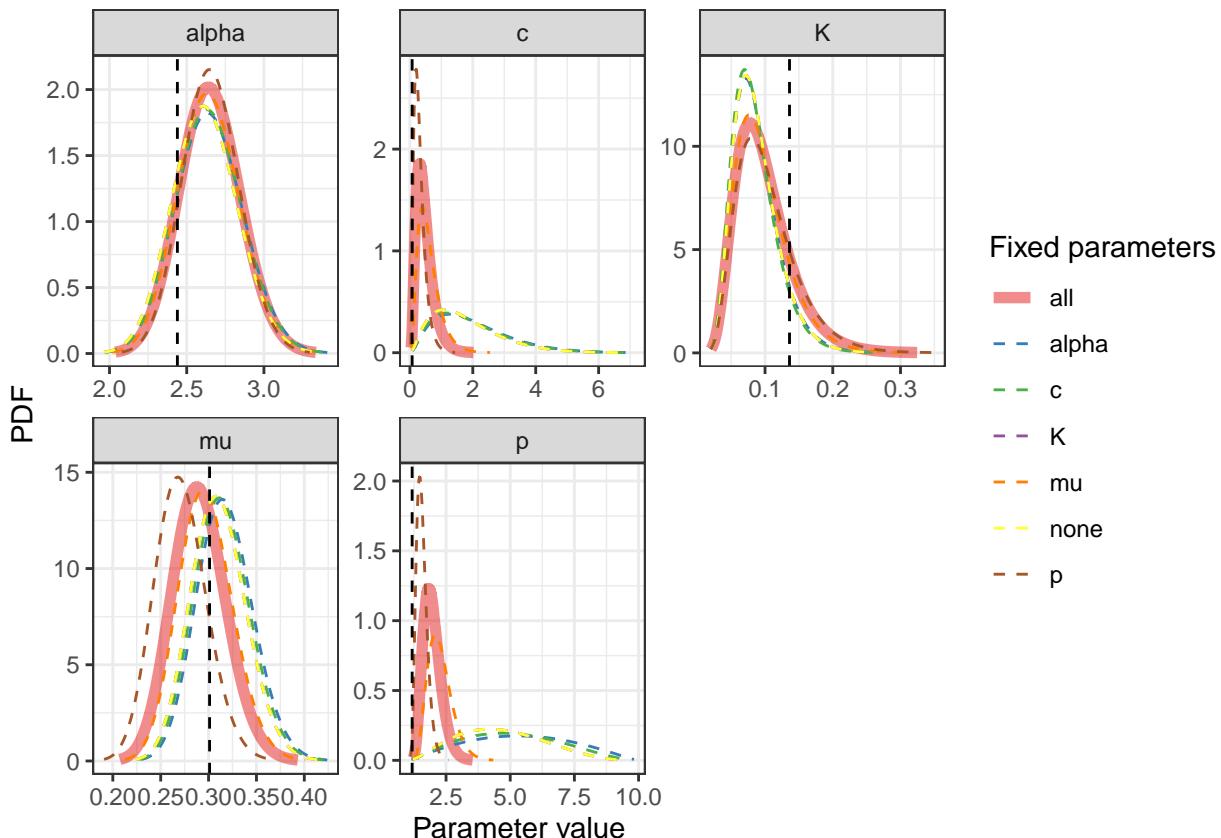


Figure 2: Posterior distributions of parameters when setting one of them fixed to the values used when generating the data

Figure 2 shows the posterior distributions when one parameter is correctly specified. The red thick line corresponds to the posterior distribution when all parameters are correctly specified, the yellow dashed

line corresponds to the posterior distribution when all parameters are incorrectly specified, and the other colored dashed lines correspond to the posterior distributions when a single parameter is correctly specified. It can be seen that when all parameters are incorrectly specified and when c , α , or K is correctly specified, the posterior distributions of the different parameters are very close to each other. However, the posterior distributions of c and p are quite different from the case when all parameters are correctly specified. In contrast, when p is correctly specified, the posterior distributions of α and K are relatively close to the case when all parameters are correctly specified, but the posterior distributions of c , μ , and p are not as good. Overall, when μ is correctly specified, the posterior distributions of all parameters are closer to the case when all parameters are correctly specified, although the posterior distributions of c and p are still not perfect. Table 2 provides specific details on the standard deviation of the posterior distribution when the initial values of one parameter are correctly set, as well as the absolute percentage deviation between the mode of the posterior distribution and the true parameter values.

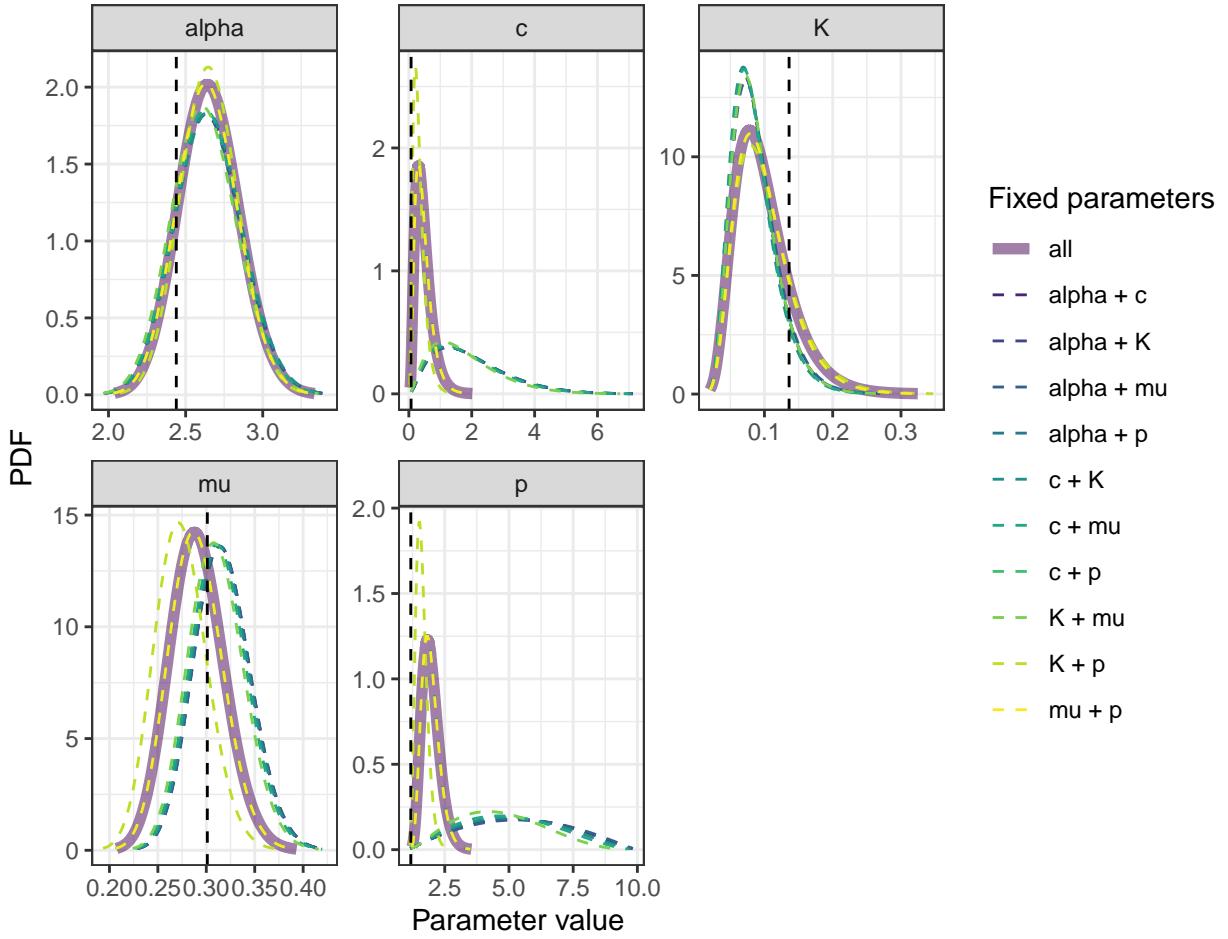


Figure 3: Posterior distributions of parameters when setting two of them fixed to the values used when generating the data

Figure 3 shows the posterior distributions when two parameters are correctly specified. The purple thick line corresponds to the posterior distribution when all parameters are correctly specified, and the other colored dashed lines correspond to the posterior distributions when different pairs of parameters are correctly specified. It can be seen that among these 10 cases, there is one case where the posterior distributions of all parameters almost perfectly match the posterior distributions when all parameters are correctly specified. After investigation, it was found that the two parameters correctly specified in this case are μ and p . The posterior distributions when μ and p are correctly specified are shown in Figure 4, and they are nearly identical to the posterior distributions when all parameters are correctly specified. Table 3 provides specific details on the standard deviation of the posterior distribution when the initial values of two parameters are correctly set, as well as the absolute percentage deviation between the mode of the posterior distribution and the true parameter values.

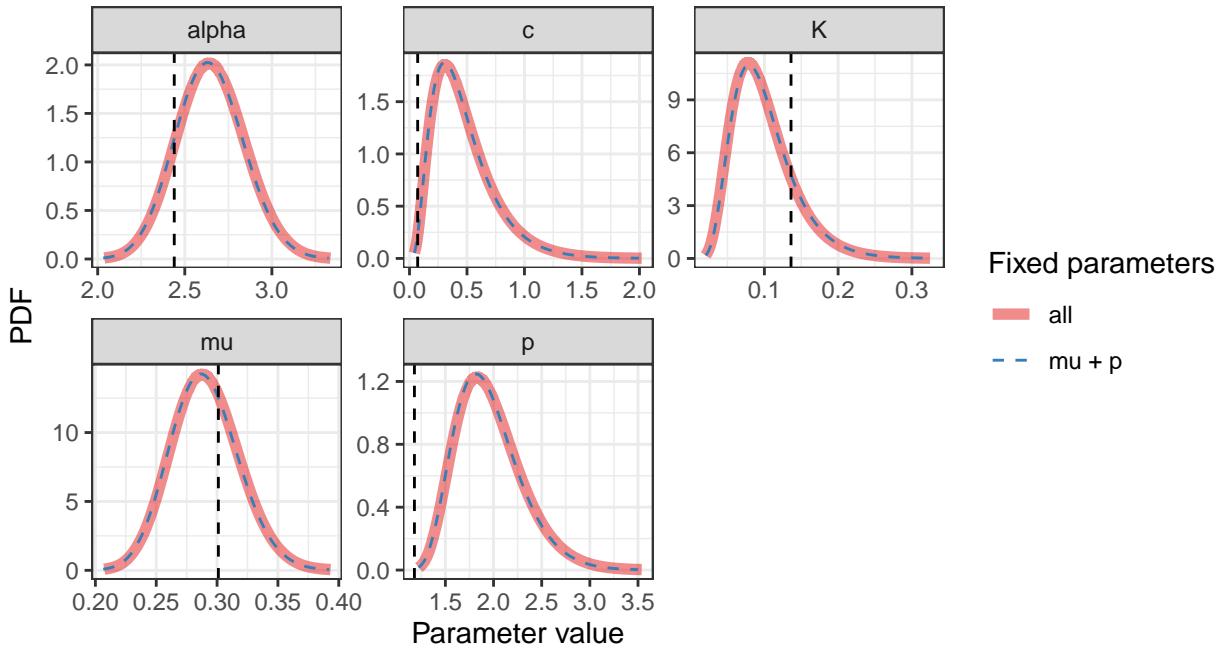


Figure 4: Posterior distributions of parameters when setting μ and p fixed to the values used when generating the data

Figure 5 shows the posterior distributions when three parameters are correctly specified. The purple thick line corresponds to the posterior distribution when all parameters are correctly specified, and the other colored dashed lines correspond to the posterior distributions when different triplets of parameters are correctly specified. Similar to the pattern observed in Figure 2, among these 10 cases where three parameters are correctly specified, there are a few cases where the posterior distributions of all parameters almost perfectly match the posterior distributions when all parameters are correctly specified. After investigation, it was found that the three parameters that satisfy this condition can be:

- α, c, p
- α, μ, p
- c, μ, p

The posterior distributions when these triplets of parameters are correctly specified are shown in Figure 6, and they are nearly identical to the posterior distributions when all parameters are correctly specified. Table 4 provides specific details on the standard deviation of the posterior distribution when the initial values of three parameters are correctly set, as well as the absolute percentage deviation between the mode of the posterior distribution and the true parameter values.

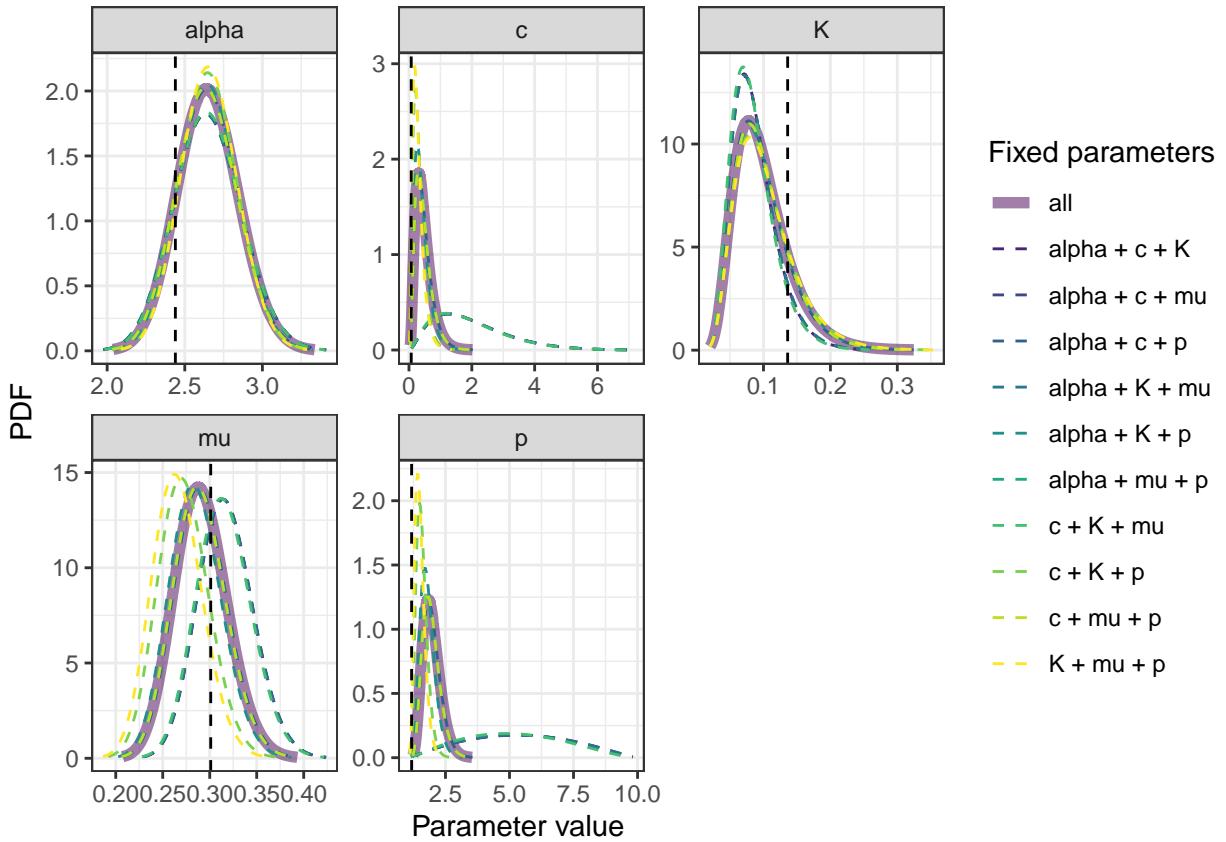


Figure 5: Posterior distributions of parameters when setting three of them fixed to the values used when generating the data

Figure 7 shows the posterior distributions when four parameters are correctly specified. The purple thick line corresponds to the posterior distribution when all parameters are correctly specified, and the other colored dashed lines correspond to the posterior distributions when different sets of four parameters are correctly specified. It can be seen that when the initial values of α , K , μ , and p are correctly specified, the posterior distributions of all parameters are almost identical to the posterior distributions when all parameters are correctly specified. Table 4 provides specific details on the standard deviation of the posterior distribution when the initial values of four parameters are correctly set, as well as the absolute percentage deviation between the mode of the posterior distribution and the true parameter values.

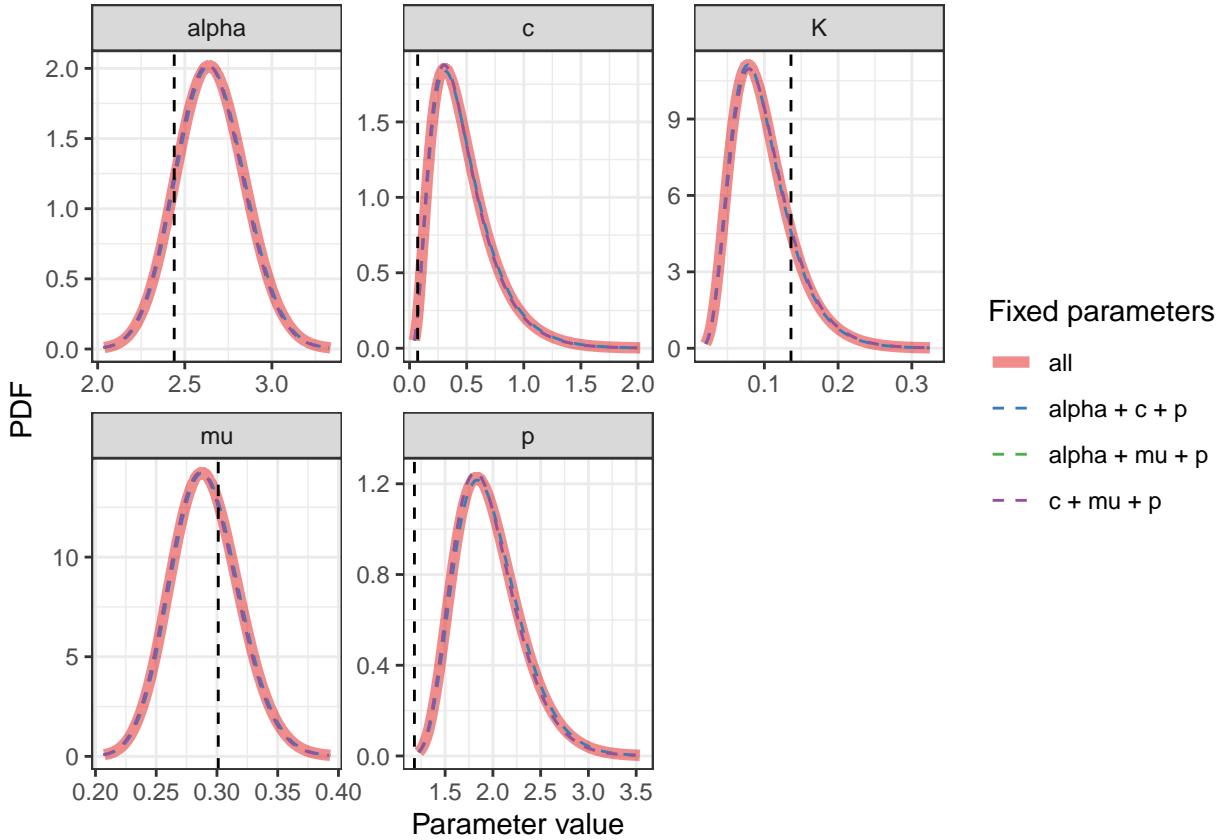


Figure 6: Posterior distributions of parameters when setting three of them (α , c , p or α , μ , p or c , μ , p) fixed to the values used when generating the data

3.3 Relationship between sequence characteristics and posterior distributions of parameters

Here, we use the 27 synthetic catalogues generated in the first research question again to investigate the relationship between the parameters' posterior distributions of the model fitted on these synthetic catalogues and the characteristics of the synthetic catalogues. Specifically, the characteristics considered include the number of events in the sequence and the number of large events, where we define large events as those with magnitudes greater than or equal to 5.

Figure 8 and Figure 9 reveal the relationships between the number of events/large events and the posterior distributions of the model parameters fitted on the corresponding synthetic catalogues, respectively. The black dashed lines represent the true parameter values, and each curve corresponds to one model. It is clear that as the number of events/large events increases, the mode of the posterior distribution of the parameters becomes closer to the true parameter values, and the standard deviation of the posterior distribution also decreases.

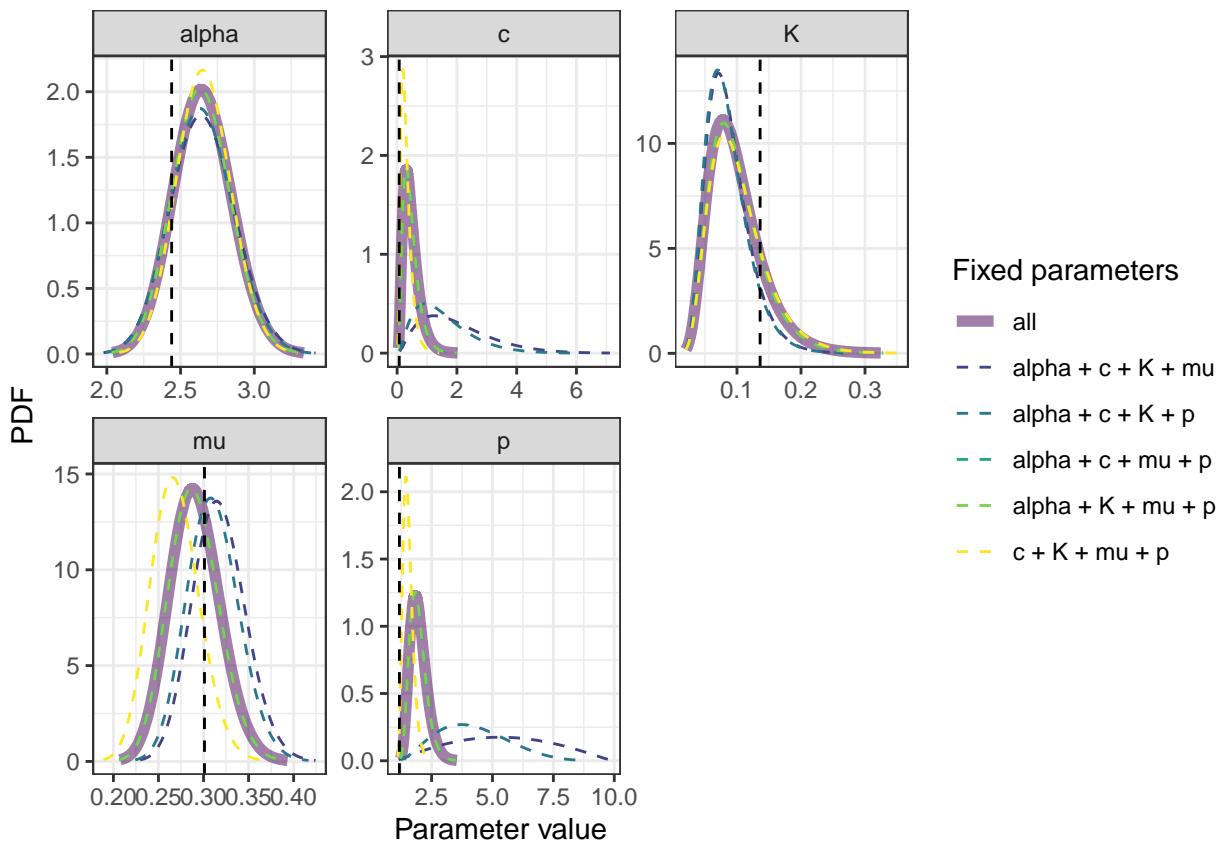


Figure 7: Posterior distributions of parameters when setting four of them fixed to the values used when generating the data

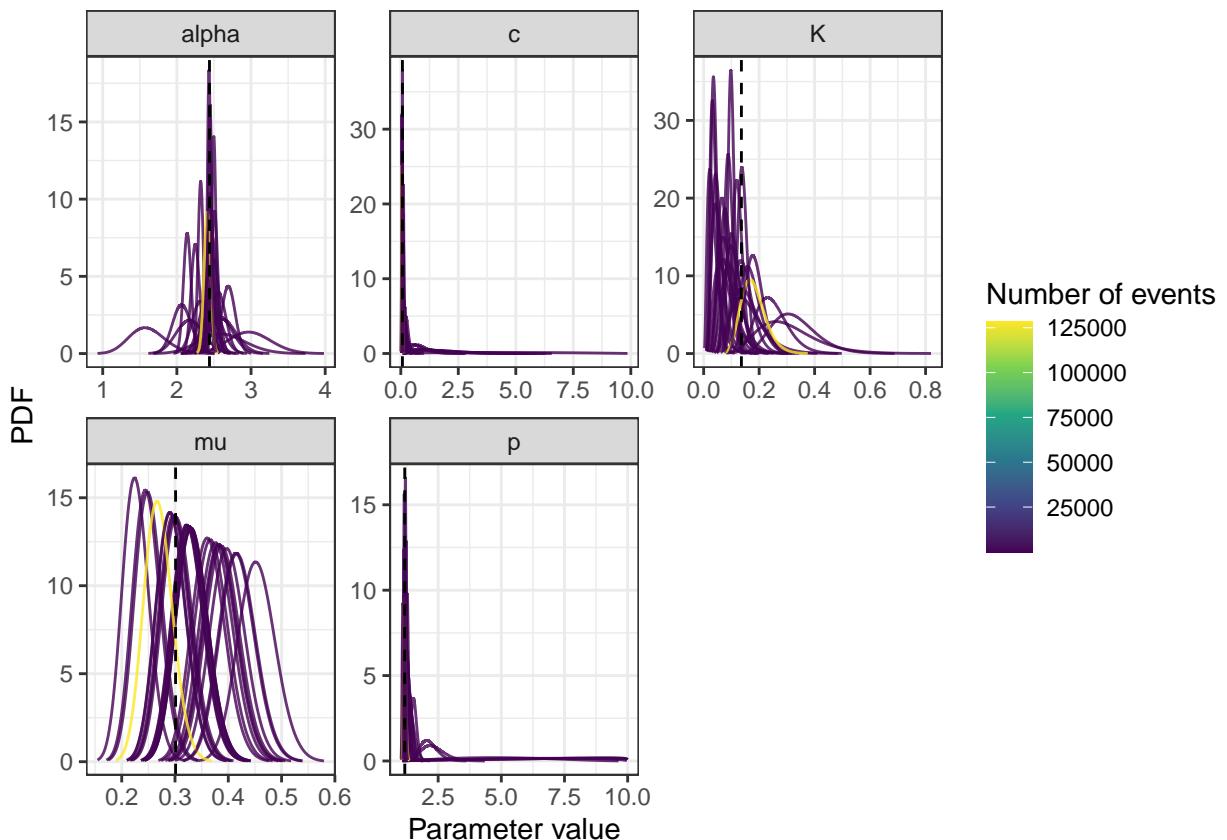


Figure 8: Relationship between posterior distributions of parameters and number of events

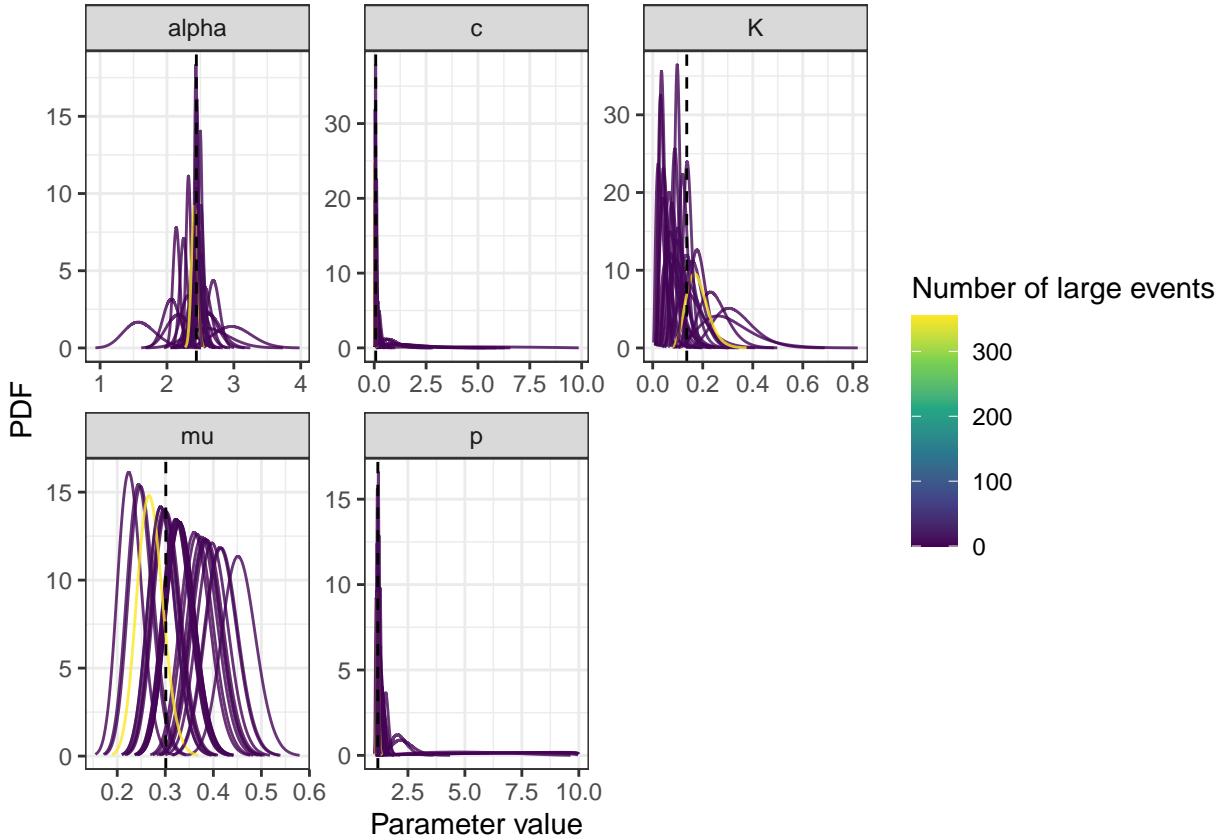


Figure 9: Relationship between posterior distributions of parameters and number of large events

3.4 Forecast accuracy

Here, we have regenerated a synthetic catalogue, and the difference from the previous synthetic catalogues is that the historical known event in generating this synthetic catalogue is an earthquake of magnitude 6.5 that occurred at time 91.25. The generated earthquake sequence is shown in Figure 10, where the red dashed line indicates the time of the occurrence of the known event, and a large number of aftershocks occurred after that.

Based on this synthetic catalogue, we fit the ETAS model using the parameter settings mentioned earlier, and then generate 100 samples from the joint posterior of the ETAS parameters. In addition, it should be noted that the parameter of the magnitude distribution β used to generate the synthetic catalogue needs to be updated. The magnitude distribution is an exponential distribution:

$$m - M_0 \sim \text{Exp}(\beta)$$

The parameter β is usually estimated independently from the ETAS parameters. use the maximum likelihood estimator for β which is given by

$$\hat{\beta} = \frac{1}{m - M_0}$$

where m is the mean of the observed magnitudes values.

When forecasting future earthquake events, we set the starting time of the forecasting period to 1 minute after the occurrence of the largest magnitude earthquake, i.e., 1 minute after 91.25, and the end time of the forecasting period is the starting time plus the forecast length. Here, we set the forecast length to 1 to 50, so we will obtain 50 forecasting periods in the end.

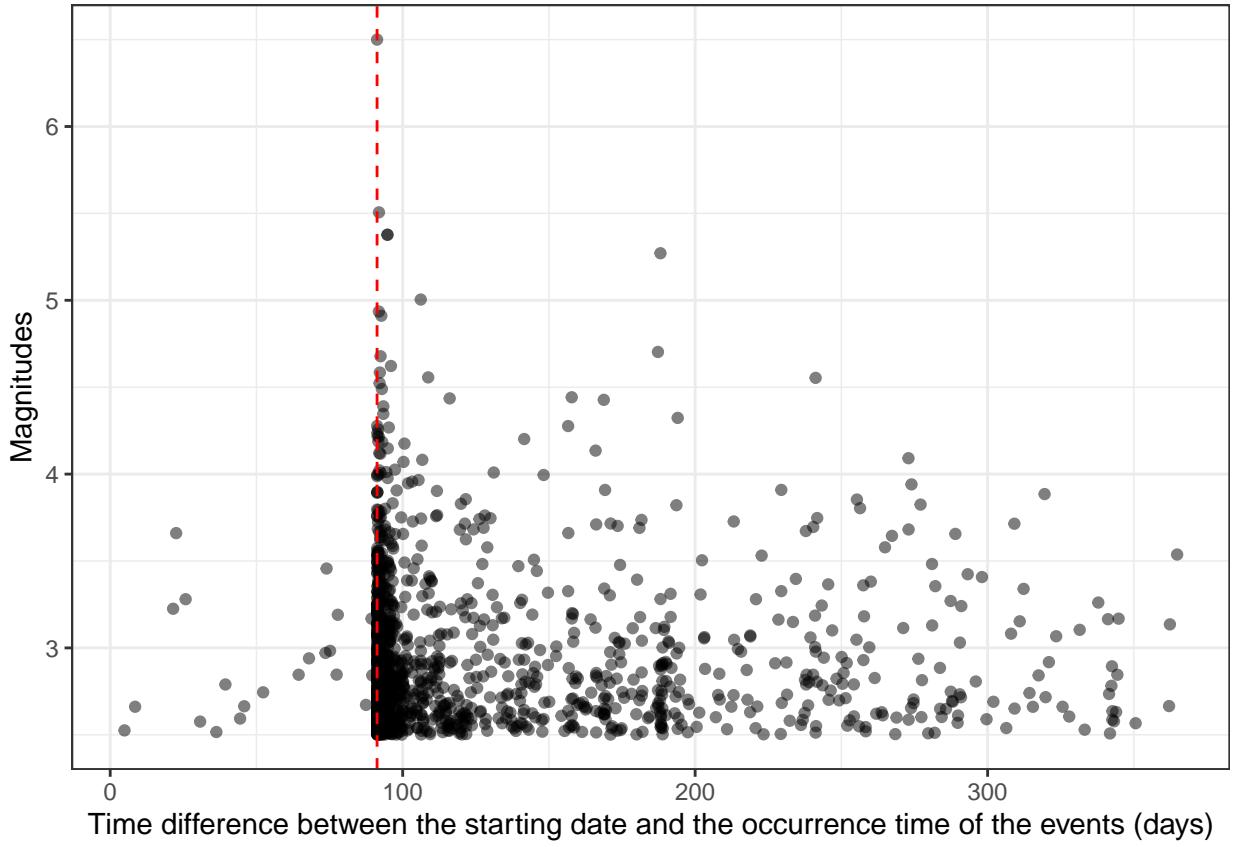


Figure 10: Synthetic catalogue

For each forecasting period, we use the known events used to generate the synthetic catalogue, the updated parameter of the magnitude distribution β , and 100 samples from the joint posterior of the ETAS parameters. For each posterior sample, we generate a synthetic catalogue composing the forecast. Therefore, for a single forecasting period, we can generate 100 earthquake sequences, and the number of events in each sequence is the number of forecast events per catalogue in that specific forecasting period. Based on these 100 numbers of forecast events, we calculate the median, precision, 2.5%, and 97.5% quantiles. In addition, we need to use the initially generated synthetic catalogue to calculate the number of observed events in the forecasting period.

Figure 11 shows the number of events in the 50 forecasting periods starting 1 minute after the occurrence of the largest magnitude earthquake. The black dots represent the number of observed events in the forecasting periods, the red line represents the median of the number of forecast events in the forecasting periods, and the orange area represents the 95% confidence interval of the number of forecast events in the forecasting periods. Note that the y-axis is log-transformed, which is done to make the visualization clearer and more intuitive. In fact, we are forecasting the cumulative count of events after 1 minute of the occurrence of the largest magnitude earthquake. It can be seen that the growth of the number of aftershocks is the fastest within the first 5 days after the largest magnitude earthquake, and the growth of the aftershocks gradually slows down after 5 days. The median of the number of forecast events is close to the number of observed events, and the 95% confidence interval also includes the number of observed events in the forecasting periods. In other words, the forecast accuracy of the ETAS model is good.

To evaluate the forecast accuracy of the ETAS model numerically, the following statistical metrics can be considered:

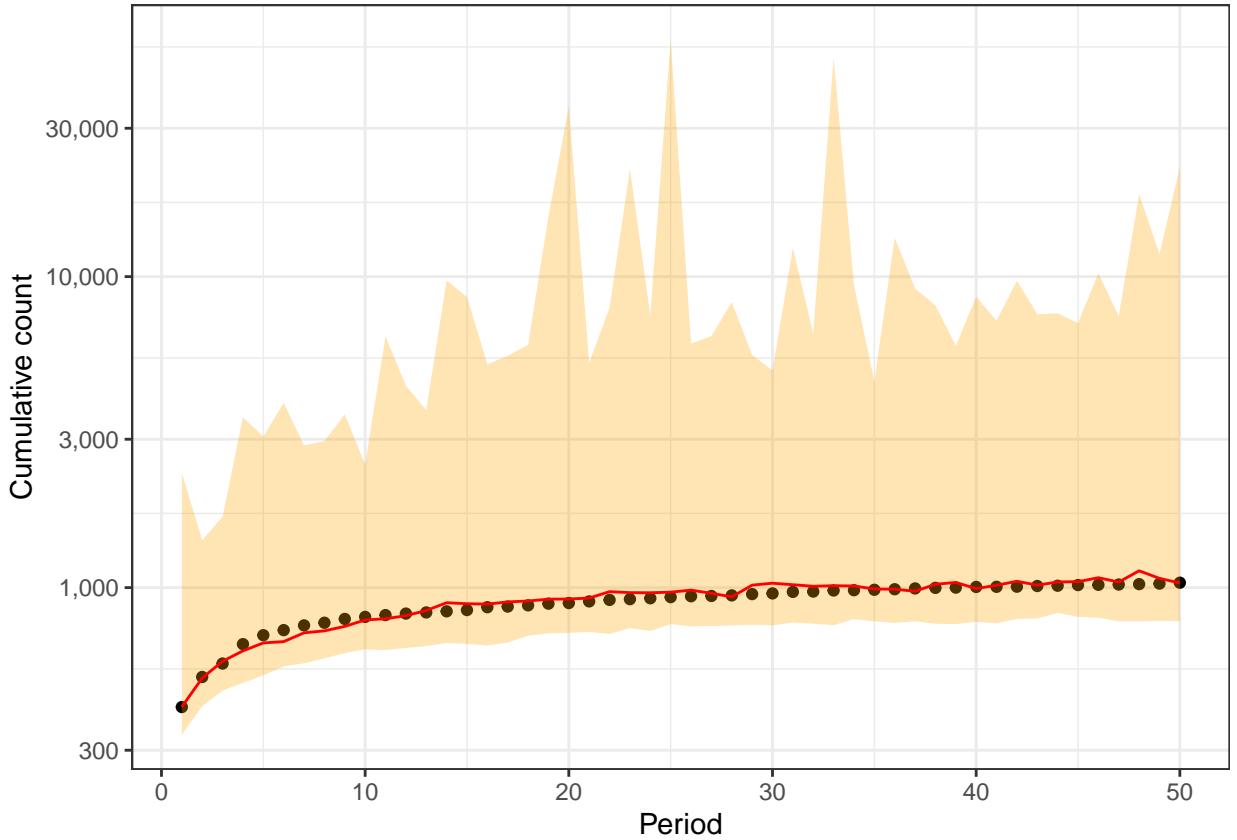


Figure 11: Forecast and true cumulative count of events

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2}$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t|$$

$$MAPE = 100 \times \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{y_t}$$

$$R^2 = \left[\frac{\sum_{t=1}^n (y_t - \bar{y}_t)(\hat{y}_t - \bar{\hat{y}}_t)}{\sqrt{\sum_{t=1}^n (\hat{y}_t - \bar{\hat{y}}_t)^2} \sqrt{\sum_{t=1}^n (y_t - \bar{y}_t)^2}} \right]^2$$

Where y_t is the number of observed events, \hat{y}_t is the number of forecast events, n is the total number of forecasting periods. The more the $RMSE$, MAE , $MAPE$, and *Theil's U* are closer to 0, the higher forecast accuracy, and the higher the R^2 , the stronger the correlation between the number of forecast events and the number of observed events.

Table 1 presents the values of statistical metrics related to the forecast accuracy of the ETAS model in these 50 forecasting periods. On average, the mean error in predicting the cumulative count of events is between 30 and 40, which can be considered relatively small given the scale of the y-axis in Figure 10. On average, the deviation between the forecast and observation is 3.65%, and the correlation between the two is 0.96.

To evaluate the forecast precision of the ETAS model, for each forecasting period, I calculated the

Table 1: Statistics about forecast accuracy

Statistics	Value
RMSE	37.7182
MAE	31.6000
MAPE	3.5234
R2	0.9642

inverse of the standard deviation of the forecast from 100 synthetic catalogs, which represents the precision. Then, I plotted the relationship between the forecast precision and the forecasting period, as shown in Figure 12. It can be seen that as the forecasting period is extended, the forecast precision decreases, which is because the forecast is for the cumulative count of events, and therefore the forecast error accumulates as the forecasting period is lengthened.

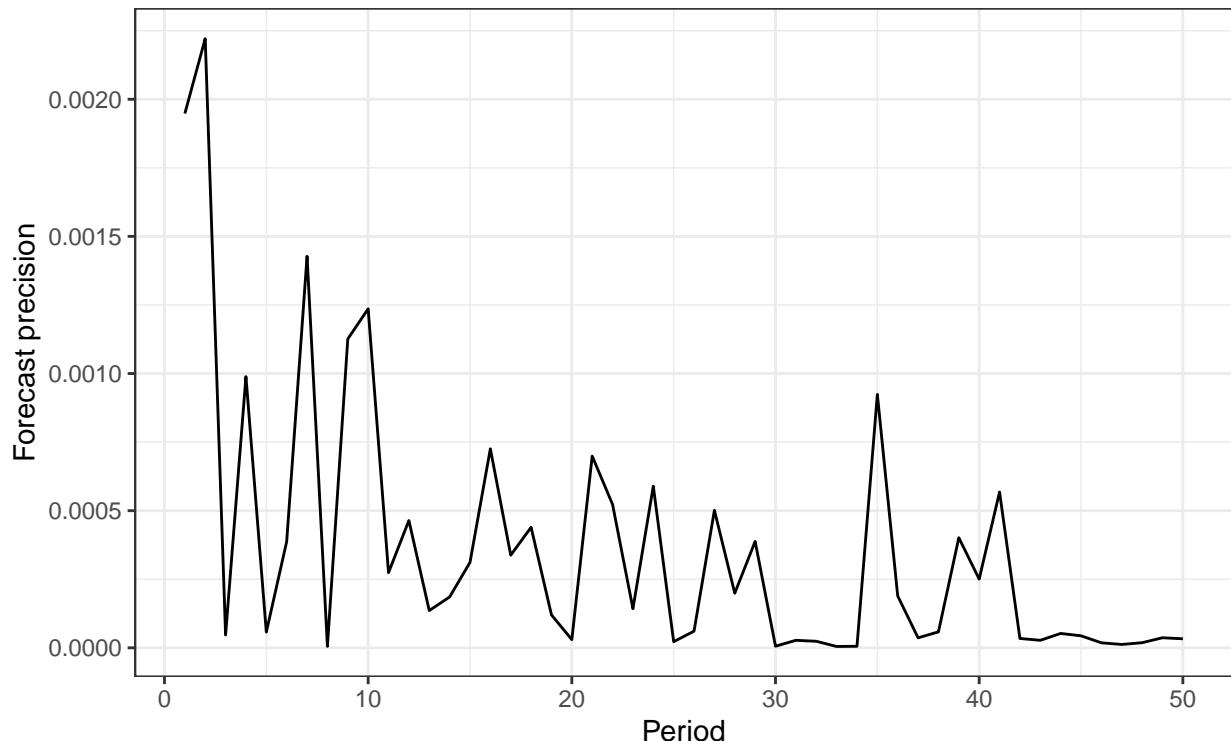


Figure 12: Forecast precision of ETAS model

3.5 Empirical distribution of “time between events”

To compare the empirical distribution of “time between events” in the synthetic catalog from estimated vs. known parameters, I directly used the synthetic catalog generated without historical known events and using the true parameter values in the second research question, as well as the corresponding posterior distribution of the ETAS model. First, I found the mode of the posterior distribution of all the parameters of the fitted ETAS model, which served as the estimated parameters, and updated the parameter of the magnitude distribution β based on the synthetic catalog. Then, I regenerated the synthetic catalog using the estimated parameters and the estimated β . Finally, I calculated the “time between events” for the synthetic catalogs from both the estimated and known parameters.

Figures 13 and 14 show the empirical CDF and empirical PDF of the “time between events” in the synthetic catalog from estimated vs. known parameters, respectively. It can be seen that the empirical distributions of the “time between events” are very similar in the two cases, meaning that even the synthetic catalog generated using the estimated parameters has a “time between events” distribution that is close to the one calculated using the “true” data. Therefore, from this perspective, the ETAS model fitted to the data is highly reliable.

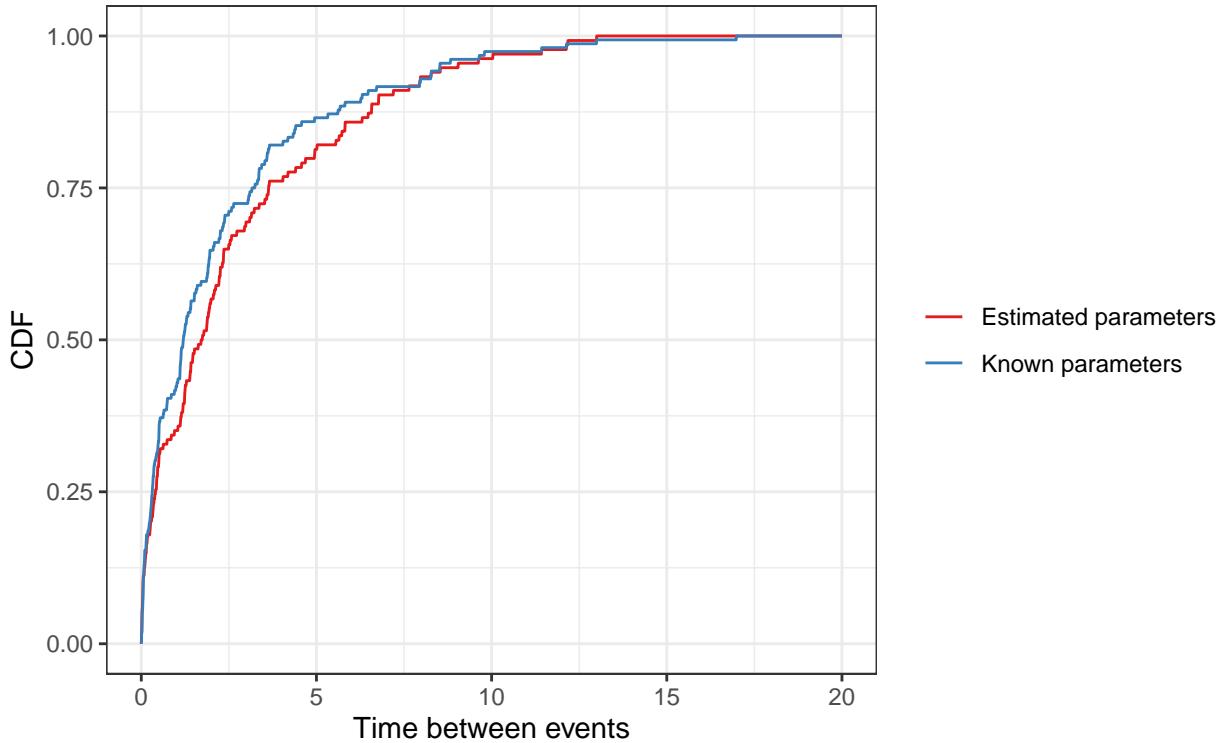


Figure 13: Empirical CDF of “time between events” in synthetic catalogue from estimated vs known parameters

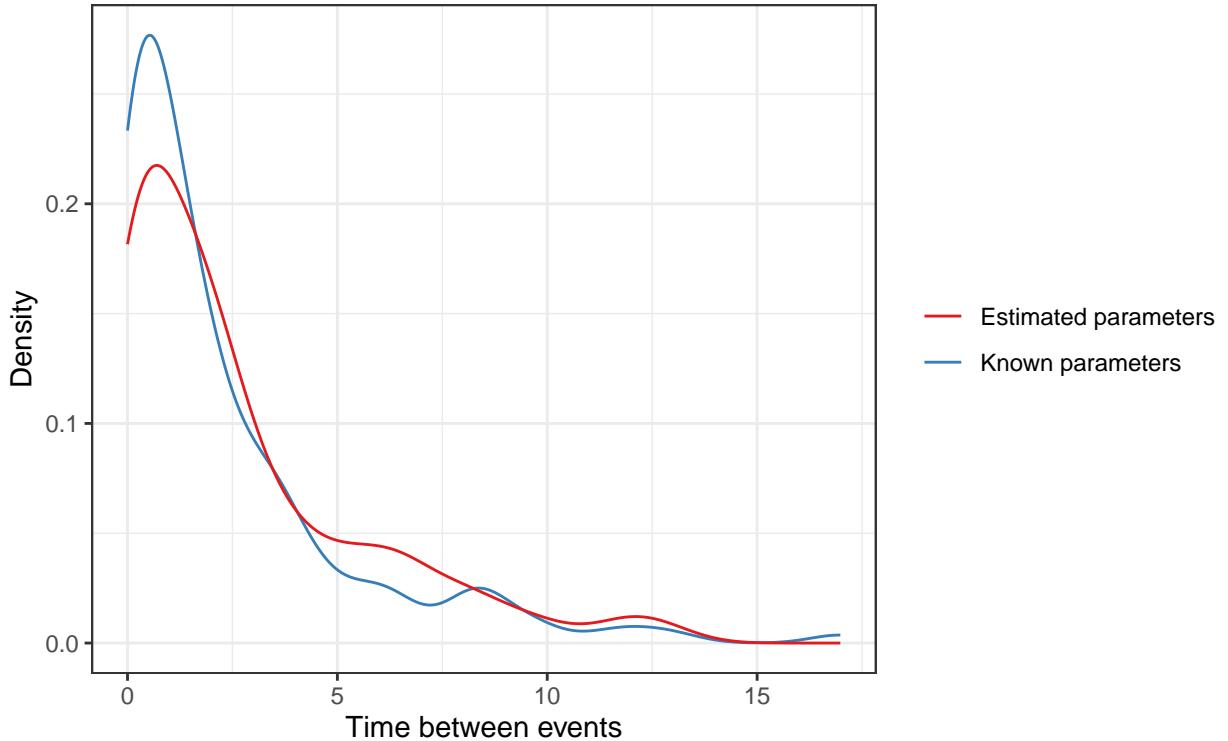


Figure 14: Empirical PDF of “time between events” in synthetic catalogue from estimated vs known parameters

To formally compare the existence of significant differences between the empirical distribution of “time between events” in two different scenarios, a two-sample Kolmogorov-Smirnov (KS) test is employed for testing. The two-sample KS test compares the empirical distribution functions from two groups.

Suppose the “time between events” data from two groups:

$$X_1, \dots, X_n \sim F_{known}$$

$$Y_1, \dots, Y_m \sim F_{estimate}$$

The two-sample KS test performs a test of the following hypothesis:

$$H_0 : F_{known} = F_{estimate} \text{ vs. } H_1 : F_{known} \neq F_{estimate}$$

The two-sample KS test statistic is defined as the maximum distance between the two empirical distribution functions:

$$KS_{n,m} = \sup_t |\hat{F}_{n,known}(t) - \hat{F}_{m,estimate}(t)|$$

where $\hat{F}_{n,known}(t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t)$ and $\hat{F}_{m,estimate}(t) = \frac{1}{m} \sum_{j=1}^m I(Y_j \leq t)$ denote the empirical distribution functions from the X and Y samples.

After conducting 10,000 iterations of the permutation test, the p-value of the KS test for the empirical distributions of “time between events” in synthetic data from estimated versus known parameters is greater than 0.05. This suggests that we do not have sufficient evidence to reject the null hypothesis,

indicating that there is no statistically significant difference in “time between events” between synthetic data from estimated versus known parameters.

4 Conclusion

The performance of the model fitted using the `ETAS.inlabru` package in forecasting future earthquake events is quite good. More different magnitude events and higher magnitude events often make the parameter estimation results of the model more reliable, and the correct setting of some initial parameter values can make the posterior distribution of the model parameters more reasonable. Specifically:

- Catalogs containing more different high-magnitude events or higher-magnitude events can make the mode of the posterior distribution of α closer to the true α value and can also make the standard deviation of the posterior distribution of α smaller.
- When constructing the ETAS model, if the initial values of the two parameters μ and p are set correctly, when the algorithm converges, the posterior distributions of all the parameters of the model will perfectly match the posterior distributions of the parameters obtained when all the initial values are set with the correct parameter values. The same situation will also occur when the initial values of three parameters ($\alpha, c, p; \alpha, \mu, p; c, \mu, p$) or four parameters (α, K, μ, p) are set correctly.
- The larger the number of events and the number of large events in the catalog, the closer the mode of the posterior distribution of the model parameters (not just α) to the true parameter values, and the smaller the standard deviation, that is, the more reliable the parameter estimation of the model.
- When forecasting the future cumulative count of events, the average deviation between the forecasts and true values is 3.65%, and the correlation between the two is 0.96. In addition, the forecast precision gradually decreases as the forecasting period is extended.
- The empirical distribution of the “time between events” in synthetic catalogs from estimated and known parameters is almost identical, which means that the ETAS model fitted using the INLA method is highly reliable.

References

- [1] M. Naylor, F. Serafini, F. Lindgren, and I. G. Main. Bayesian modeling of the temporal evolution of seismicity using the `etas.inlabru` package. *Frontiers in Applied Mathematics and Statistics*, 9, 2023.
- [2] F. Serafini, F. Lindgren, and M. Naylor. Approximation of bayesian hawkes process with `inlabru`. *Environmetrics*, 34(5):e2798, 2023.

5 Appendices

A Code

In order to make the report more clear, the codes for generating the results can be found in <https://github.com/JIAOHAOYANG/S2333015-dissertation>.

B Some tables with different number of parameters' intial values were correctly set

Table 2: The absolute percentage deviation and standard deviation of posterior distributions of parameters when one parameters' intial values were correctly set

Correct Set Inital Value	Parameter	Abs Diff	SD
K	K	47.358	0.022
K	alpha	6.827	0.151
K	c	1416.929	0.730
K	mu	2.129	0.021
K	p	254.414	1.280
alpha	K	47.456	0.022
alpha	alpha	7.962	0.155
alpha	c	1638.413	0.811
alpha	mu	4.370	0.021
alpha	p	349.019	1.619
c	K	48.756	0.021
c	alpha	7.809	0.153
c	c	1549.001	0.791
c	mu	3.323	0.021
c	p	300.502	1.463
mu	K	43.591	0.026
mu	alpha	7.856	0.143
mu	c	452.164	0.216
mu	mu	2.308	0.020
mu	p	76.974	0.322
none	K	47.358	0.022
none	alpha	6.827	0.151
none	c	1417.037	0.730
none	mu	2.130	0.021
none	p	254.435	1.280
p	K	41.015	0.029
p	alpha	8.506	0.131
p	c	175.964	0.112
p	mu	11.009	0.019
p	p	24.739	0.144

Table 3: The absolute percentage deviation and standard deviation of posterior distributions of parameters when two parameters' intial values were correctly set

Correct Set Initial Value	Parameter	Abs Diff	SD
K + mu	K	47.358	0.022
K + mu	alpha	6.827	0.151
K + mu	c	1417.208	0.730
K + mu	mu	2.130	0.021
K + mu	p	254.468	1.280
K + p	K	41.069	0.028
K + p	alpha	8.449	0.132
K + p	c	191.375	0.117
K + p	mu	9.972	0.019
K + p	p	28.050	0.152
alpha + K	K	47.816	0.022
alpha + K	alpha	8.016	0.155
alpha + K	c	1630.540	0.811
alpha + K	mu	4.267	0.021
alpha + K	p	342.375	1.602
alpha + c	K	47.682	0.022
alpha + c	alpha	7.929	0.155
alpha + c	c	1630.847	0.811
alpha + c	mu	4.200	0.021
alpha + c	p	341.445	1.599
alpha + mu	K	46.922	0.022
alpha + mu	alpha	7.767	0.155
alpha + mu	c	1645.579	0.811
alpha + mu	mu	4.343	0.021
alpha + mu	p	352.653	1.627
alpha + p	K	41.785	0.027
alpha + p	alpha	7.870	0.139
alpha + p	c	325.694	0.164
alpha + p	mu	4.793	0.020
alpha + p	p	54.154	0.232
c + K	K	49.301	0.021
c + K	alpha	8.279	0.154
c + K	c	1578.669	0.801
c + K	mu	3.846	0.021
c + K	p	313.905	1.519
c + mu	K	48.618	0.022
c + mu	alpha	7.675	0.152
c + mu	c	1531.801	0.784
c + mu	mu	3.148	0.021
c + mu	p	293.986	1.437
c + p	K	41.789	0.027
c + p	alpha	7.869	0.139
c + p	c	325.991	0.165
c + p	mu	4.786	0.020
c + p	p	54.209	0.232

mu + p	K	41.795	0.027
mu + p	alpha	7.868	0.139
mu + p	c	326.479	0.165
mu + p	mu	4.773	0.020
mu + p	p	54.299	0.232

Table 4: The absolute percentage deviation and standard deviation of posterior distributions of parameters when three parameters' intial values were correctly set

Correct Set Inital Value	Parameter	Abs Diff	SD
K + mu + p	K	40.931	0.029
K + mu + p	alpha	8.561	0.129
K + mu + p	c	154.588	0.106
K + mu + p	mu	12.736	0.019
K + mu + p	p	20.022	0.133
alpha + K + mu	K	47.524	0.022
alpha + K + mu	alpha	7.881	0.155
alpha + K + mu	c	1632.995	0.811
alpha + K + mu	mu	4.208	0.021
alpha + K + mu	p	342.862	1.603
alpha + K + p	K	41.844	0.027
alpha + K + p	alpha	8.288	0.137
alpha + K + p	c	269.485	0.144
alpha + K + p	mu	6.293	0.020
alpha + K + p	p	43.751	0.197
alpha + c + K	K	47.875	0.022
alpha + c + K	alpha	8.050	0.155
alpha + c + K	c	1630.229	0.810
alpha + c + K	mu	4.288	0.021
alpha + c + K	p	342.742	1.603
alpha + c + mu	K	47.502	0.022
alpha + c + mu	alpha	7.981	0.155
alpha + c + mu	c	1637.987	0.811
alpha + c + mu	mu	4.376	0.021
alpha + c + mu	p	348.911	1.618
alpha + c + p	K	42.662	0.027
alpha + c + p	alpha	8.191	0.140
alpha + c + p	c	333.992	0.168
alpha + c + p	mu	4.379	0.020
alpha + c + p	p	55.879	0.238
alpha + mu + p	K	41.785	0.027
alpha + mu + p	alpha	7.869	0.139
alpha + mu + p	c	325.704	0.164
alpha + mu + p	mu	4.793	0.020
alpha + mu + p	p	54.155	0.232
c + K + mu	K	49.046	0.021
c + K + mu	alpha	8.102	0.154
c + K + mu	c	1573.319	0.800

Table 5: The absolute percentage deviation and standard deviation of posterior distributions of parameters when four parameters' intial values were correctly set

Correct Set Initial Value	Parameter	Abs Diff	SD
alpha + K + mu + p	K	41.778	0.027
alpha + K + mu + p	alpha	7.870	0.139
alpha + K + mu + p	c	325.245	0.164
alpha + K + mu + p	mu	4.805	0.020
alpha + K + mu + p	p	54.070	0.232
alpha + c + K + mu	K	47.777	0.022
alpha + c + K + mu	alpha	8.128	0.155
alpha + c + K + mu	c	1636.253	0.810
alpha + c + K + mu	mu	4.462	0.021
alpha + c + K + mu	p	349.532	1.620
alpha + c + K + p	K	48.956	0.022
alpha + c + K + p	alpha	7.844	0.151
alpha + c + K + p	c	1214.124	0.619
alpha + c + K + p	mu	2.224	0.021
alpha + c + K + p	p	216.121	1.069
alpha + c + mu + p	K	41.787	0.027
alpha + c + mu + p	alpha	7.869	0.139
alpha + c + mu + p	c	325.903	0.164
alpha + c + mu + p	mu	4.788	0.020
alpha + c + mu + p	p	54.192	0.232
c + K + mu + p	K	41.021	0.029
c + K + mu + p	alpha	8.558	0.130
c + K + mu + p	c	166.162	0.109
c + K + mu + p	mu	11.743	0.019
c + K + mu + p	p	22.595	0.139
c + K + mu	mu	3.694	0.021
c + K + mu	p	310.970	1.507
c + K + p	K	41.038	0.029
c + K + p	alpha	8.489	0.132
c + K + p	c	183.068	0.114
c + K + p	mu	10.522	0.019
c + K + p	p	26.252	0.148
c + mu + p	K	41.785	0.027
c + mu + p	alpha	7.869	0.139
c + mu + p	c	325.717	0.164
c + mu + p	mu	4.793	0.020
c + mu + p	p	54.158	0.232