Bayesian updating on time intervals at different magnitude thresholds in a marked point process and its application to synthetic seismic activity

Hiroki Tanaka[1] and Ken Umeno[1]

[1]Kyoto University

January 11, 2023

Abstract

We present a Bayesian updating method on the inter-event times at different magnitude thresholds in a marked point process, toward the probabilistic forecasting of an upcoming large event using temporal information on smaller events. Bayes' theorem in a marked point process that yields the one-to-one relationship between intervals at lower and upper magnitude thresholds is presented. This theorem is extended to Bayesian updating for an uncorrelated marked point process that yields the relationship between multiple consecutive lower intervals and one upper interval. The inverse probability density function and its approximation function are derived. For the former, the condition for having a peak is shown. The latter is easier to apply to the time series of the ETAS model, and it consists of the kernel part, which includes the product of the conditional probabilities, and the correction term. The maximum point of the kernel part is shown to be not significantly affected by the correction term when applying the Bayesian updating to the ETAS model time series numerically. The occurrence time of the upcoming large event is estimated using this maximum point, and its accuracy is evaluated considering the relative error with the actual occurrence time. Moreover, forecasting is evaluated to be effective by the continuity of the updates with the accuracy within an acceptable range prior to the upcoming large event. Under these conditions, the statistical analysis indicates that forecasting is relatively effective immediately or long after the last major event in which stationarity is dominant in the time series.

Bayesian updating on time intervals at different magnitude thresholds in a marked point process and its application to synthetic seismic activity

Hiroki Tanaka [1] and Ken Umeno [1]

[1]Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto-shi, JAPAN

Key Points:

- Bayesian updating for time intervals at different magnitude thresholds in marked point process is discussed for large event forecasting
- Inverse probability density and its approximation functions are derived analytically for a time series with no correlation between events
- Forecasting of large events by Bayesian method is examined for the ETAS model time series and its effectiveness is discussed statistically

Corresponding author: Hiroki Tanaka, `tanaka.hiroki.43s@st.kyoto-u.ac.jp`

Abstract

We present a Bayesian updating method on the inter-event times at different magnitude thresholds in a marked point process, toward the probabilistic forecasting of an upcoming large event using temporal information on smaller events. Bayes' theorem in a marked point process that yields the one-to-one relationship between intervals at lower and upper magnitude thresholds is presented. This theorem is extended to Bayesian updating for an uncorrelated marked point process that yields the relationship between multiple consecutive lower intervals and one upper interval. The inverse probability density function and its approximation function are derived. For the former, the condition for having a peak is shown. The latter is easier to apply to the time series of the ETAS model, and it consists of the kernel part, which includes the product of the conditional probabilities, and the correction term. The maximum point of the kernel part is shown to be not significantly affected by the correction term when applying the Bayesian updating to the ETAS model time series numerically. The occurrence time of the upcoming large event is estimated using this maximum point, and its accuracy is evaluated considering the relative error with the actual occurrence time. Moreover, forecasting is evaluated to be effective by the continuity of the updates with the accuracy within an acceptable range prior to the upcoming large event. Under these conditions, the statistical analysis indicates that forecasting is relatively effective immediately or long after the last major event in which stationarity is dominant in the time series.

Plain Language Summary

In order to forecast future large earthquakes, it is important to use as much information as possible on the seismic activity at hand. The number of small earthquakes is much larger than that of large earthquakes, and we propose a method to use this information to forecast probabilistically the timing of future large earthquakes. Theoretical analysis is performed on a simple time series. The theoretical results are applied to a seismic activity model, and it is shown that this method is relatively effective in forecasting the timing of future large events when stationary activity is dominant; in this model, either immediately after a large event or after sufficient time has passed since the last large event. Therefore, this method can be applied to reduce secondary disasters after a major earthquake and to evaluate the risk of earthquake occurrence over a long period of time.

1 Introduction

The probabilistic forecasting of the timing of future major earthquakes is important for seismic risk assessment. Therefore, it is necessary to effectively use the information on the temporal properties of seismic activity represented by a marked point process with the magnitude as the mark as indicated in Figure 1. A basic approach involves using the hazard rate based on the inter-event time distribution of earthquakes (Scholz, 2002). The inter-event time distribution is defined as a probability density function of the length of the interval between adjacent points in the point process determined by setting a magnitude threshold for the marked point process. For the magnitude threshold M (m), we denote the inter-event times using a variable τ_M (τ_m), and the inter-event time distribution it follows by $p_M(\tau_M)$ ($p_m(\tau_m)$).

The inter-event time distribution of earthquakes has been studied for not only risk assessment but also to understand the statistical nature of seismicity. For example, it has been studied with the aim to unify it with other established laws (Bak et al., 2002; Aizawa et al., 2013; Aizawa & Tsugawa, 2014) such as the GR law (Gutenberg & Richter, 1944) and the Omori–Utsu law (Omori, 1894; Utsu, 1961); further, its scaling universality (Corral, 2004) has been discussed using the ETAS model (Saichev & Sornette, 2006; Touati et al., 2009; Bottiglieri et al., 2010; Lippiello et al., 2012).

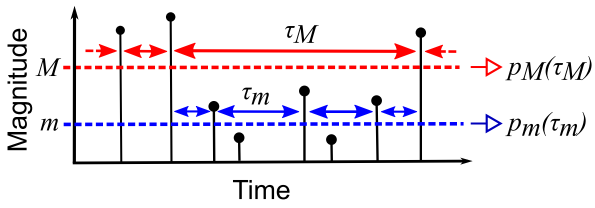

Figure 1. Schematic of a marked point process with the magnitude as a mark.

The ETAS model is a stochastic model that combines the GR and Omori–Utsu laws (Ogata, 1988, 1998); this model can generate a time series like that of the seismic activity. The ETAS model generates an inhomogeneous Poisson process with a history-dependent occurrence rate. Let t_j and M_j ($j \in \mathbb{N}$) represent the occurrence time and magnitude of the j-th event before time t; then, the occurrence rate ($\lambda(t)$) at time t is given by

$$\lambda(t) = \lambda_0 + \sum_{j:t_j<t} \frac{K10^{\alpha(M_j - M_0)}}{(t - t_j + c)^{\theta+1}}. \tag{1}$$

The magnitude is generated randomly and independently obeying the GR law, $P(M) \propto 10^{-bM}$. Here, M_0 represents the minimum magnitude and $(\lambda_0, K, \alpha, c, \theta, b)$ represent the parameters that characterize the activity. In particular, λ_0 represents the constant rate for background seismicity. The combination of the remaining parameters yields the branching ratio $n_{\mathrm{br}} = \dfrac{K}{\theta c^\theta}\dfrac{b}{b-\alpha}$ (when $\theta > 0$) (Helmstetter & Sornette, 2002) that determines the stationarity of the time series as well as the average number of aftershocks generated by a mainshock (Helmstetter & Sornette, 2003).

The ETAS model provides a standard seismicity for detecting anomalous activity (Ogata, 1988). This model has been extended to a spatio-temporal version (Ogata, 1998), and its application to the evaluation of seismic risk has been studied actively. The conditional intensity function provides the risk of an event at a given time, space, and magnitude based on the history of seismic activity, which includes small earthquakes.

In the aforementioned probabilistic evaluation using the inter-event time distribution, temporal information on events smaller than the magnitude threshold set on the marked point process is not utilized. Therefore, in this paper, we propose another approach to probabilistically forecast major earthquakes based on the inter-event time distribution while considering the temporal information on smaller events. This is achieved by utilizing a conditional probability that yields the statistical relationship between the inter-event times at different two magnitude thresholds (Tanaka & Aizawa, 2017).

For two magnitude thresholds m and $M(= m + \Delta m, \Delta m > 0)$ set in the time series, the conditional probability ($p_{mM}(\tau_m|\tau_M)$, hereafter referred to as the conditional probability density function) is defined as the probability density function of the length of a lower interval (τ_m) under the condition that it is included in the upper interval of length τ_M (Figure 1). Inter-event time distributions at magnitude thresholds m and M are connected by an integral equation with this conditional probability density function in its kernel (Tanaka & Aizawa, 2017). This integral equation is given as

$$N_m p_m(\tau_m) = N_M \int_{\tau_m}^{\infty} \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M) p_M(\tau_M) d\tau_M, \tag{2}$$

where N_m and N_M represent the total number of intervals at magnitude thresholds m and M, respectively. Further, $\langle\langle\tau_m\rangle\rangle_{\tau_M}$ represents the average of the conditional probability density function, $\langle\langle\tau_m\rangle\rangle_{\tau_M} := \int_0^\infty \tau_m p_{mM}(\tau_m|\tau_M) d\tau_m$.

Thus, the conditional probability density function yields the statistical relation between inter-event times at different magnitude thresholds. This suggests that the information on the lower intervals can be utilized for estimating the length of the upper interval through the conditional probability density function. Given this context, we consider the Bayes' theorem and the Bayesian updating on the intervals at different magnitude thresholds in this paper; further, we report the results of the numerical analysis of the properties of the inverse probability density function (Tanaka & Umeno, 2021).

In §2, we derive the Bayes' theorem for intervals at different magnitude thresholds in the marked point process. In §3, the inverse probability density function is derived for the uncorrelated time series that corresponds to the background seismicity of the ETAS model. In §4, the Bayesian updating method is considered for the uncorrelated time series, and the inverse probability density function and its approximation function are derived. These functions are calculated numerically and compared in §5. In §6, Bayesian updating is applied to the time series of the ETAS model. The approximation function is examined numerically, and the property of the maximum point of its kernel part is analyzed statistically considering the effectiveness for forecasting. Finally, §7 presents additional discussions and conclusions.

2 Bayes' theorem for inter-event times at different magnitude thresholds

We consider the Bayes' theorem between the inter-event times at different magnitude thresholds (m and M) in a marked point process, and we derive the general relationship between the conditional probability density function $p_{mM}(\tau_m|\tau_M)$ and the inverse probability density function $p_{Mm}(\tau_M|\tau_m)$ (Tanaka & Umeno, 2021). Here $p_{Mm}(\tau_M|\tau_m)$ represents the probability density function of the upper interval under the condition that it includes a lower interval of length τ_m.

Let the total number of the pairs of the upper interval of length within $[\tau_M, \tau_M + d\tau_M)$ and the lower interval of length within $[\tau_m, \tau_m + d\tau_m)$ by $N_{mM}(\tau_M, \tau_m)$ (Figure 2). Hereafter, we express this $N_{mM}(\tau_M, \tau_m)$ as the number of the pairs of the intervals such that the length of the upper interval *is* τ_M and the length of the lower interval *is* τ_m, for simplicity, and other numbers of the intervals are expressed in the same way. $N_{mM}(\tau_M, \tau_m)$ can be represented in two ways:

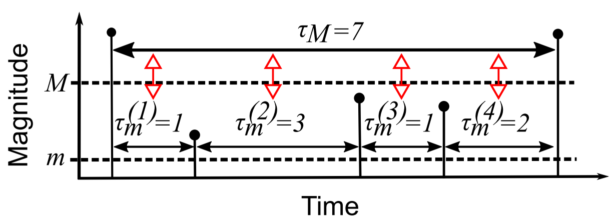

Figure 2. Schematic of the approach to count the number of pairs of upper and lower intervals whose lengths are τ_M and τ_m, respectively. Four pairs are shown in the figure, and $N_{mM}(7,1) = 2$, $N_{mM}(7,2) = 1$, and $N_{mM}(7,3) = 1$.

1) Derive $N_{mM}(\tau_M, \tau_m)$ by counting the cumulative total number of the upper intervals of length τ_M that include the lower interval of length τ_m (Figure 3(a)). Among the N_m lower intervals in the time series, there are $N_m p_m(\tau_m) d\tau_m$ intervals of length τ_m. There exists only one upper interval that includes each of such lower intervals. The probability that the length of that upper interval is τ_M is given by $p_{Mm}(\tau_M|\tau_m) d\tau_M$. There-

fore

$$N_{mM}(\tau_M,\tau_m) = N_m p_m(\tau_m) p_{Mm}(\tau_M|\tau_m) d\tau_m d\tau_M. \tag{3}$$

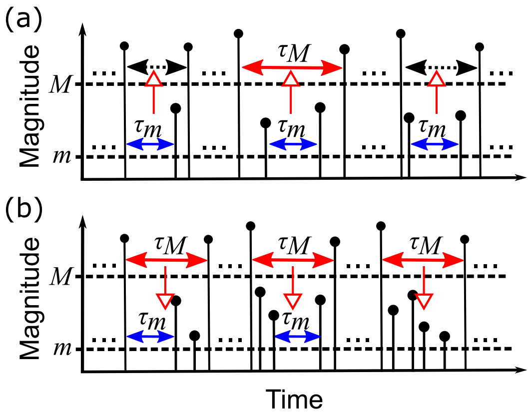

Figure 3. Schematic of the two approaches for calculating $N_{mM}(\tau_M,\tau_m)$. (a) The first approach involves counting the cumulative total number of the upper intervals of length τ_M that include the lower interval of length τ_m. (b) The second approach involves counting the number of the lower intervals of length τ_m included in the upper interval of length τ_M.

2) Derive $N_{mM}(\tau_M,\tau_m)$ by counting the total number of the lower intervals of length τ_m included in the upper interval of length τ_M (Figure 3(b)). The number of the upper intervals of length τ_M in the time series is $N_M p_M(\tau_M) d\tau_M$. Therefore, the number of the lower intervals included in these upper intervals is

$$N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} d\tau_M.$$

Among them, the proportion of the lower intervals whose length is τ_m is $p_{mM}(\tau_m|\tau_M)d\tau_m$. Therefore

$$N_{mM}(\tau_M,\tau_m) = N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M) d\tau_m d\tau_M. \tag{4}$$

From Equations (3) and (4)

$$N_m p_m(\tau_m) p_{Mm}(\tau_M|\tau_m) = N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M). \tag{5}$$

By using the average intervals at each magnitude threshold

$$\langle\tau_m\rangle := \int_0^\infty \tau_m p_m(\tau_m) d\tau_m,$$
$$\langle\tau_M\rangle := \int_0^\infty \tau_M p_M(\tau_M) d\tau_M,$$

and $N_M/N_m = \langle\tau_m\rangle/\langle\tau_M\rangle$, Equation (5) is rewritten as

$$p_{Mm}(\tau_M|\tau_m) = \left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}}\right) \frac{p_{mM}(\tau_m|\tau_M) p_M(\tau_M)}{p_m(\tau_m)}. \tag{6}$$

Equation (6) can be considered as the Bayes' theorem for a marked point process. The parenthesized part is from the difference in the number of intervals for each magnitude threshold ($\langle\tau_m\rangle/\langle\tau_M\rangle$) and the inclusion relationship between the upper and lower intervals $(\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M})$, i.e., a lower interval is always included in only one upper interval, whereas an upper interval includes $\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M}$ lower intervals on average. This part disappears by using generalized probability density functions

$$
\begin{aligned}
z_m(\tau_m) &:= \frac{\tau_m}{\langle\tau_m\rangle} p_m(\tau_m), \\
z_M(\tau_M) &:= \frac{\tau_M}{\langle\tau_M\rangle} p_M(\tau_M), \\
z_{mM}(\tau_m|\tau_M) &:= \frac{\tau_m}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M).
\end{aligned}
$$

These functions satisfy the normalization condition of the probability density function. Equations (2) and (6) are simplified as

$$
z_m(\tau_m) = \int_0^\infty z_{mM}(\tau_m|\tau_M) z_M(\tau_M) d\tau_M, \tag{7}
$$

$$
p_{Mm}(\tau_M|\tau_m) = \frac{z_{mM}(\tau_m|\tau_M) z_M(\tau_M)}{z_m(\tau_m)}. \tag{8}
$$

These equations indicate that $p_{Mm}(\tau_M|\tau_m)$ satisfies the normalization condition.

3 Bayes' theorem for uncorrelated time series

In this section, we derive $p_{Mm}(\tau_M|\tau_m)$ for an uncorrelated time series generated by the ETAS model with $\lambda(t) \equiv \lambda_0$ (Tanaka & Umeno, 2021). In this case, the magnitudes and inter-event times obey the following probability density functions independently.

$$
P(m) \propto 10^{-bm}, \tag{9}
$$

$$
p_m(\tau_m) = \frac{1}{\langle\tau_m\rangle} e^{-\frac{\tau_m}{\langle\tau_m\rangle}}. \tag{10}
$$

First, we derive $p_{mM}(\tau_m|\tau_M)$, which can be expressed generally as

$$
p_{mM}(\tau_m|\tau_M) = \frac{\sum_{i=1}^\infty i\ \rho_{mM}(\tau_m|i,\tau_M)\Psi_{mM}(i|\tau_M)}{\sum_{i=1}^\infty i\ \Psi_{mM}(i|\tau_M)}, \tag{11}
$$

where $i(\in \mathbb{Z})$ represents the number of lower intervals included in the upper interval of length τ_M; $\Psi_{mM}(i|\tau_M)$ represents the probability mass function of such i under the condition that the length of the upper interval is τ_M; and $\rho_{mM}(\tau_m|i,\tau_M)$ represents the probability density function of the length of a lower interval given that the length of the upper interval is τ_M and the number of the lower intervals in it is i. We can calculate the conditional probability density function and other related amounts when we know these functions.

In the case of the uncorrelated time series, these functions can be obtained as follows. For the selected stationary Poisson process, the average number of events included in the upper interval of length τ_M is $(\tau_M/\langle\tau_m\rangle - \tau_M/\langle\tau_M\rangle)$, because no event greater than M occurs in the interval considered, and therefore, $\tau_M/\langle\tau_M\rangle$-events larger than M occurring in the interval of length τ_M on average must be excluded from the average number $\tau_M/\langle\tau_m\rangle$ of events occurring in the interval of length τ_M. Then, the average occurrence rate in the upper interval of length τ_M is $(1/\langle\tau_m\rangle - 1/\langle\tau_M\rangle)$. The number of events with $m <$ magnitude $\leq M$ in τ_M is one less than that of the lower intervals, and therefore, the probability of including i lower intervals is equal to the probability of including $(i-1)$ events with an average occurrence rate $(1/\langle\tau_m\rangle - 1/\langle\tau_M\rangle)$. Therefore

$$
\Psi_{mM}(i|\tau_M) = \frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right)^{i-1}}{(i-1)!} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}, \tag{12}
$$

where

$$A_{\Delta m} := \frac{\langle \tau_M \rangle}{\langle \tau_m \rangle} - 1 \\ = 10^{b\Delta m} - 1.$$

The second transformation in the above equation does not strictly hold for a time series with a finite number of events because the number of the events is different from that of the intervals by 1. However, we consider that the statistical properties are for infinite samples, and in a time series containing an infinite number of events, the two are equivalent and the equality holds.

The other function $\rho_{mM}(\tau_m|i, \tau_M)$ is obtained as follows. For $i = 1$

$$\rho_{mM}(\tau_m|1, \tau_M) = \delta(\tau_M - \tau_m), \tag{13}$$

where $\delta(x)$ represents the Dirac's delta function. For $i \geq 2$ (Webb, 1974)

$$\rho_{mM}(\tau_m|i, \tau_M) = \frac{(i-1)}{\tau_M}\left(1 - \frac{\tau_m}{\tau_M}\right)^{i-2} \theta(\tau_M - \tau_m), \tag{14}$$

where $\theta(x)$ represents the unit step function that returns 1 for $x > 0$ and 0 for $x \leq 0$.

From Equations (12)–(14), $p_{mM}(\tau_m|\tau_M)$ is derived as (Appendix A)

$$p_{mM}(\tau_m|\tau_M) = \frac{e^{-A_{\Delta m}\frac{\tau_M}{\langle \tau_M \rangle}}\delta(\tau_M - \tau_m) + \frac{A_{\Delta m}}{\langle \tau_M \rangle}e^{-A_{\Delta m}\frac{\tau_m}{\langle \tau_M \rangle}}\left\{A_{\Delta m}\frac{\tau_M - \tau_m}{\langle \tau_M \rangle} + 2\right\}\theta(\tau_M - \tau_m)}{\left(A_{\Delta m}\frac{\tau_M}{\langle \tau_M \rangle} + 1\right)}. \tag{15}$$

This conditional probability composed of Equations (12)–(14) certainly has exponential distributions as the solution of Equation (2) (Appendix A).

Second, we derive $p_{Mm}(\tau_M|\tau_m)$. From Equations (10) and (15), $p_{Mm}(\tau_M|\tau_m)$ is obtained as (Appendix A)

$$p_{Mm}(\tau_M|\tau_m) = \frac{e^{-\frac{\tau_M - \tau_m}{\langle \tau_m \rangle}}\delta(\tau_M - \tau_m) + \frac{A_{\Delta m}}{\langle \tau_M \rangle}e^{-\frac{\tau_M - \tau_m}{\langle \tau_M \rangle}}\left\{A_{\Delta m}\frac{\tau_M - \tau_m}{\langle \tau_M \rangle} + 2\right\}\theta(\tau_M - \tau_m)}{\left(A_{\Delta m} + 1\right)^2}. \tag{16}$$

We emphasize that $p_{Mm}(\tau_M|\tau_m)$ has a peak at

$$\tau_M^{\max} = \tau_m + \langle \tau_M \rangle \left(1 - \frac{2}{A_{\Delta m}}\right), \tag{17}$$

when the next condition is satisfied (Appendix A).

$$\Delta m > \frac{\log_{10} 3}{b}. \tag{18}$$

4 Bayesian updating for uncorrelated time series

The Bayes' theorem shows a one-to-one relationship between an upper and a lower interval. In this section, we extend it to the relationship between an upper interval and multiple consecutive lower intervals by considering Bayesian updating for the uncorrelated time series (Tanaka & Umeno, 2021). We derive the inverse probability density function $p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$, as well as its approximation function, for the upper interval under the condition that it includes the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$.

4.1 Inverse probability density function

As in §2, we derive the inverse probability density function by expressing the total number of combinations of the upper interval of length τ_M and the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ included in it denoted by $N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ in two ways.

First, we derive $N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ by counting the cumulative total number of the upper intervals of length τ_M that include the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ (Figure 4(a)). We begin with the case $n = 2$. The intervals in the uncorrelated time series emerge independently, and therefore, the total number of the two consecutive lower intervals of lengths $\tau_m^{(1)}$ and $\tau_m^{(2)}$ is

$$N_m p_m(\tau_m^{(1)}) p_m(\tau_m^{(2)}) d\tau_m^2.$$

Among them, some pairs do not belong to the same upper interval (the case of (3) in Figure 4(a)). In that case, the magnitude of the event sandwiched between the two lower intervals is larger than M. In the uncorrelated time series, the proportion that the consecutive lower intervals belong to the same upper interval equals to the probability that the magnitude of the event sandwiched between the two lower intervals is smaller than M. It is given by the GR law as

$$\begin{aligned} 1 - \frac{P(M)}{P(m)} &= 1 - 10^{-b\Delta m} \\ &= 1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}. \end{aligned}$$

Therefore

$$N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)}) = N_m \left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right) p_m(\tau_m^{(1)}) p_m(\tau_m^{(2)}) p_{Mm}(\tau_M | \tau_m^{(1)}, \tau_m^{(2)}) d\tau_m^2 d\tau_M. \quad (19)$$

Equation (19) is generalized for $n(\geq 2)$ consecutive lower intervals.

$$\begin{aligned} &N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\ &= N_m \left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1} \left(\prod_{i=1}^{n} p_m(\tau_m^{(i)})\right) p_{Mm}(\tau_M | \tau_m^{(1)}, \cdots, \tau_m^{(n)}) d\tau_m^n d\tau_M. \end{aligned} \quad (20)$$

Second, we derive $N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ by counting the total number of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ included in the upper interval of length τ_M (Figure 4(b)). To this end, we start with the case $n = 2$ again (Figure 4(b)). When the upper interval of length τ_M includes $i(\geq 2)$ lower intervals, the first interval of the two consecutive lower intervals is selected from $(i-1)$ intervals except for the rightmost one. The probability that this first interval has length $\tau_m^{(1)}$ is $\rho_{mM}(\tau_m^{(1)} | i, \tau_M) d\tau_m$. The second lower interval is fixed at adjacent to the first one. This second interval is one of the $(i-1)$ intervals that divide the remaining length $\tau_M - \tau_m^{(1)}$, and therefore, the probability that the second interval has length $\tau_m^{(2)}$ is $\rho_{mM}(\tau_m^{(2)} | i-1, \tau_M - \tau_m^{(1)}) d\tau_m$. Thus, considering all $i(\geq 2)$

$$\begin{aligned} &N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)}) \\ &= N_M p_M(\tau_M) d\tau_M \sum_{i=2}^{\infty} (i-1) \Psi_{mM}(i | \tau_M) \rho_{mM}(\tau_m^{(1)} | i, \tau_M) \rho_{mM}(\tau_m^{(2)} | i-1, \tau_M - \tau_m^{(1)}) d\tau_m^2. \end{aligned} \quad (21)$$

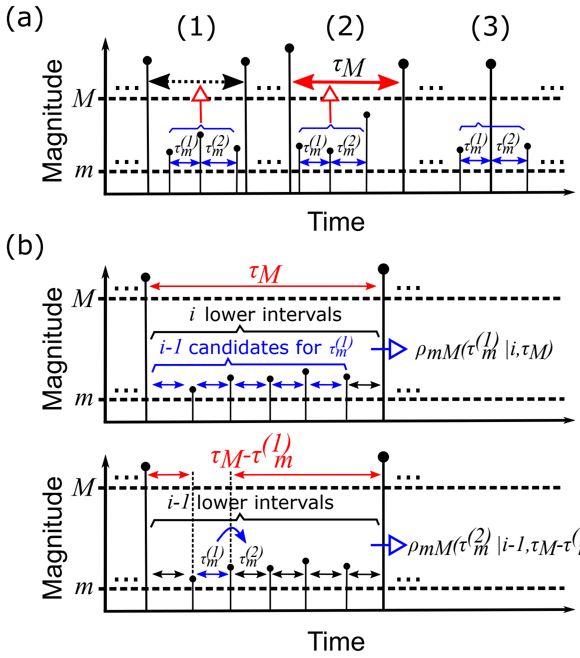

Figure 4. Schematic of the two approaches to calculate $N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)})$. (a) The first approach involves counting the cumulative total number of the upper intervals of length τ_M that include the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \tau_m^{(2)}\}$. (b) The second approach involves counting the number of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \tau_m^{(2)}\}$ included in the upper interval of length τ_M.

Equation (21) is generalized for the case $n(\geq 2)$ lower intervals as

$$N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) = N_M p_M(\tau_M) d\tau_M$$
$$\times \sum_{i=n}^{\infty} (i-n+1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1}\tau_m^{(k)}\right) d\tau_m^n. \tag{22}$$

From Equations (20) and (22), $p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$ is derived as

$$p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{1}{\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}} \frac{p_M(\tau_M)}{\prod_{i=1}^{n} p_m(\tau_m^{(i)})}$$
$$\times \sum_{i=n}^{\infty} (i-n+1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1}\tau_m^{(k)}\right). \tag{23}$$

Furthermore, the explicit form of the inverse probability density function is derived by substituting Equations (10) and (12)–(14) into Equation (23) as (Appendix B)

$$p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^2 \left[e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_m\rangle}} \delta\left(\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}\right) \right.$$
$$\left. + \frac{A_{\Delta m}}{\langle\tau_M\rangle} e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}} \left\{ \frac{A_{\Delta m}}{\langle\tau_M\rangle}\left(\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}\right) + 2 \right\} \theta\left(\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}\right) \right]. \tag{24}$$

Equation (24) includes the case $n = 1$ (Equation (16)). In addition, Equation (24) is identical to Equation (16) when τ_m is replaced with $\mathcal{T} := \sum_{i=1}^{n} \tau_m^{(i)}$; this implies that the occurrence pattern of small events does not affect that of upper intervals. This seems natural for the uncorrelated time series.

The same property as Equations (17) and (18) holds for $p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$; it has a peak at

$$\tau_M^{\max} = \mathcal{T} + \langle \tau_M \rangle \frac{\frac{\langle \tau_M \rangle}{\langle \tau_m \rangle} - 3}{\frac{\langle \tau_M \rangle}{\langle \tau_m \rangle} - 1} \; (> \mathcal{T}),$$

under the condition

$$\Delta m > \frac{\log_{10} 3}{b}. \tag{25}$$

In the above mentioned Bayesian updating, the position of the consecutive lower intervals in an upper interval is not restricted. However, update can be started only from the lower interval immediately after the event with the magnitude above M. In such a method, the inverse probability density function is different from Equation (24) (Appendix C). At a glance, this updating method seems suitable under the situation wherein the information on the lower intervals observed one after another is imported sequentially; however, seismic catalogs are known to be incomplete immediately after a large earthquake. In that case, the lower intervals should be considered not from the leftmost one but from somewhere else. Therefore, in the present paper, we limit ourselves to examine the property of the inverse probability density function of the unrestricted updating method that is more appropriate for application to earthquake catalogs.

4.2 Approximation function of inverse probability density function

Equation (23) indicates that new information on the lower intervals cannot be added by the product of the conditional probabilities as is usual in Bayesian updating. In this sub-section, we derive its approximation function with a convenient form applicable to the time series with correlations between events.

To this end, we use the approximate derivation of $N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ described below instead of the second approach for deriving Equation (22). In the following, the upper and the lower consecutive intervals are assumed to satisfy

$$\tau_M \geq \sum_{i=1}^{n} \tau_m^{(i)}. \tag{26}$$

First, consider the case $n = 2$. There are $N_M p_M(\tau_M) d\tau_M$ upper intervals of length τ_M in the time series. These upper intervals are as shown in Figure 5(a), and we use them to generate a new time series by connecting them in the order of appearance as in Figure 5(b). Let the number of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \tau_m^{(2)}\}$ in this new time series be denoted by $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$. The total number of the lower intervals in this new time series is given as

$$N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} d\tau_M.$$

Therefore, based on the assumption that $\tau_m^{(1)}$ and $\tau_m^{(2)}$ emerge independently, $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$ is approximately calculated as

$$N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M) \approx N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m^{(1)}|\tau_M) p_{mM}(\tau_m^{(2)}|\tau_M) d\tau_m^2 d\tau_M. \tag{27}$$

$N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$ is not equivalent to $N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)})$ because $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$ includes cases where the two consecutive lower intervals do not belong to the same up-

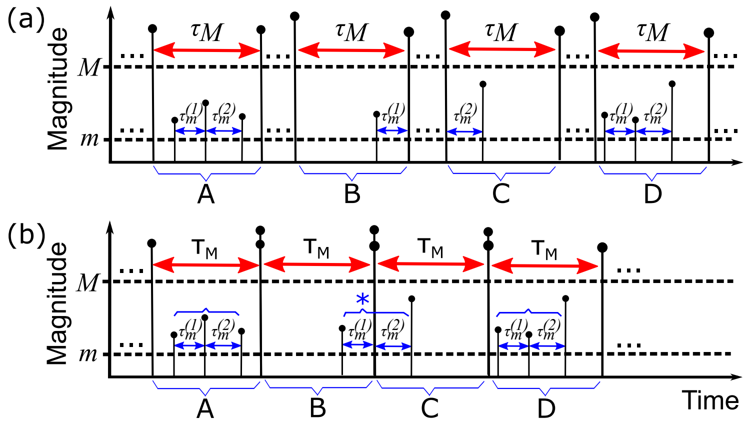

Figure 5. Schematic of another approach to count the total number of consecutive lower intervals of lengths $\tau_m^{(1)}$ and $\tau_m^{(2)}$ included in the upper interval of length τ_M. (a) First, pick up all upper intervals of length τ_M from the time series. (b) Second, generate new time series by connecting these upper intervals in the order of appearance. Third, $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$ is calculated by counting the total number of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \tau_m^{(2)}\}$ in this new time series. In this counting process, an approximate calculation using the product of the conditional probability is conducted. Finally, $N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)})$ is obtained by excluding such pairs where the two consecutive lower intervals are not included in the same upper interval (the cases indicated with $*$) from $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$.

per interval (the case indicated by $*$ in Figure 5(b)). Therefore, it is necessary to count such cases in the time series, and subtract them from $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$.

These cases to exclude occur when an upper interval of length τ_M whose rightmost lower interval has length $\tau_m^{(1)}$ is adjacent to the left of another upper interval whose leftmost lower interval has length $\tau_m^{(2)}$. The probability density that the length of the rightmost or leftmost lower interval of the upper interval of length τ_M is τ_m is, because the position of the rightmost or leftmost interval is confirmed among the i-lower intervals, calculated as

$$\begin{aligned} P^{\mathrm{R}}(\tau_m|\tau_M) &= P^{\mathrm{L}}(\tau_m|\tau_M) \\ &= \sum_{i=1}^{\infty} \Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m|i, \tau_M). \end{aligned} \tag{28}$$

Here, the probability density for the rightmost lower interval is denoted by $P^{\mathrm{R}}(\tau_m|\tau_M)$, and the leftmost by $P^{\mathrm{L}}(\tau_m|\tau_M)$. Equation (28) can be explicitly written using Equations (12)–(14) as (Appendix D)

$$\begin{aligned} P^{\mathrm{R}}(\tau_m|\tau_M) &= P^{\mathrm{L}}(\tau_m|\tau_M) \\ &= e^{-A_{\Delta m}\frac{\tau_M}{\langle \tau_M \rangle}}\delta(\tau_M - \tau_m) + \frac{A_{\Delta m}}{\langle \tau_M \rangle} e^{-A_{\Delta m}\frac{\tau_m}{\langle \tau_M \rangle}}\theta(\tau_M - \tau_m). \end{aligned} \tag{29}$$

By using $P^{\mathrm{L}}(\tau_m|\tau_M)$ and $P^{\mathrm{R}}(\tau_m|\tau_M)$, the number of cases to exclude can be expressed for a sufficiently large N_M (because $N_M p_M(\tau_M)d\tau_M$ in Equation (30) is precisely $N_M p_M(\tau_M)d\tau_M - 1$) as

$$N_M p_M(\tau_M) P^{\mathrm{R}}(\tau_m^{(1)}|\tau_M) P^{\mathrm{L}}(\tau_m^{(2)}|\tau_M) d\tau_m^2 d\tau_M. \tag{30}$$

Therefore, $N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)})$ is approximately derived as

$$N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)}) \approx N_M p_M(\tau_M) \\ \times \left(\frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m^{(1)}|\tau_M) p_{mM}(\tau_m^{(2)}|\tau_M) - P^{\mathrm{R}}(\tau_m^{(1)}|\tau_M) P^{\mathrm{L}}(\tau_m^{(2)}|\tau_M) \right) d\tau_m^2 d\tau_M. \quad (31)$$

Next, we consider the case $n(\geq 3)$. Equation (27) is generalized as

$$N'_{mM}(\tau_m^{(1)}, \cdots, \tau_m^{(n)}|\tau_M) \approx N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \left(\prod_{i=1}^{n} p_{mM}(\tau_m^{(i)}|\tau_M) \right) d\tau_m^n d\tau_M. \quad (32)$$

From this $N'_{mM}(\tau_m^{(1)}, \cdots, \tau_m^{(n)}|\tau_M)$, the cases wherein the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ are not included in the same upper interval need to be excluded. Considering the condition of Equation (26), a sequence of consecutive lower intervals is divided by only one boundary event with a magnitude above M (Figure 6). Let the probability that the rightmost or leftmost lower intervals of the upper interval of length τ_M is $\{\tau_m^{(1)}, \cdots, \tau_m^{(l)}\}$ ($l \geq 2$) be $P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M)$. Then, as the position of the rightmost or leftmost lower intervals is confirmed among the $i(\geq l)$ lower intervals, $P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M)$ is

$$P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M) \\ = \sum_{i=l}^{\infty} \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{l} \rho_{mM} \left(\tau_m^{(j)} | i - j + 1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)} \right). \quad (33)$$

By substituting Equations (12)–(14) into Equation (33) (Appendix E)

$$P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M) = \prod_{i=1}^{l} P_i(\tau_m^{(i)}|\tau_M), \text{ where } P_i(\tau_m^{(i)}|\tau_M) := \left(\frac{A_{\Delta m}}{\langle \tau_M \rangle} \right) e^{-A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle}}. \quad (34)$$

There are $(n-1)$ possible choices for the boundary position of the consecutive lower intervals (Figure 6(a)), each with an equal probability $\prod_{i=1}^{n} P_i$. The number of consecutive upper intervals in the new time series is almost $N_M p_M(\tau_M) d\tau_M$, and therefore, the number of cases to be excluded is

$$N_M p_M(\tau_M)(n-1) \left(\prod_{i=1}^{n} P_i(\tau_m^{(i)}|\tau_M) \right) d\tau_m^n d\tau_M.$$

Then

$$N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\ \approx N_M p_M(\tau_M) \left\{ \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \prod_{i=1}^{n} p_{mM}(\tau_m^{(i)}|\tau_M) - (n-1) \prod_{i=1}^{n} P_i(\tau_m^{(i)}|\tau_M) \right\} d\tau_m^n d\tau_M. \quad (35)$$

Therefore, from Equations (20) and (35), the approximation function ($p_{Mm}^{\mathrm{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$) of the inverse probability density function is derived as

$$p_{Mm}^{\mathrm{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle} \frac{1}{\left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1}} \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \left(\prod_{i=1}^{n} \frac{p_{mM}(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})} \right) p_M(\tau_M) \\ - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle} \frac{(n-1)}{\left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1}} \left(\prod_{i=1}^{n} \frac{P_i(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})} \right) p_M(\tau_M). \quad (36)$$

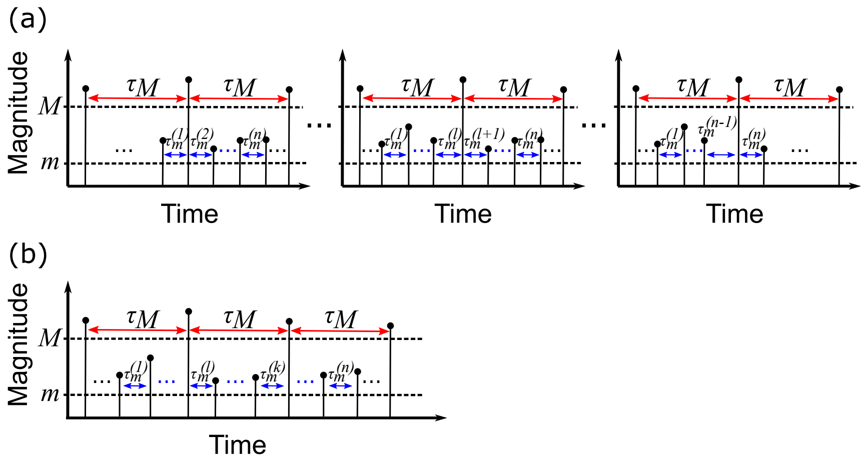

Figure 6. Schematic of the patterns of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ excluded from $N'_{mM}(\tau_m^{(1)}, \cdots, \tau_m^{(n)}|\tau_M)$. (a) There are $(n - 1)$ ways to divide the sequence of lower intervals by the event with a magnitude greater than M at the boundary of the upper intervals of length τ_M. (b) The sequence can not be divided by more than one boundary according to condition (26).

Equation (36) is composed of two parts: the first term of the r.h.s. involves the product of the conditional probability density functions, and we refer to this part as the kernel part of the approximation function ($p_{Mm}^{\mathrm{kernel}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$) hereafter.

$$p_{Mm}^{\mathrm{kernel}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{1}{\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}} \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} \left(\prod_{i=1}^{n} \frac{p_{mM}(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})}\right) p_M(\tau_M). \tag{37}$$

The second term of the r.h.s. is referred to as the correction term, and we denote the part other than $(n-1)$ by $p_{Mm}^{\mathrm{correct}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$ as

$$\text{correction term} = (n-1)p_{Mm}^{\mathrm{correct}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}), \tag{38}$$

$$\text{where } p_{Mm}^{\mathrm{correct}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{1}{\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}} \left(\prod_{i=1}^{n} \frac{P_i(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})}\right) p_M(\tau_M).$$

Equation (36) can be explicitly written as (Appendix F)

$$\begin{aligned} p_{Mm}^{\mathrm{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) &= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} \left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) \left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle} + 1\right) e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}} \\ &\times \prod_{i=1}^{n} \left\{1 - \left(\frac{\tau_m^{(i)} - \frac{\langle\tau_M\rangle}{A_{\Delta m}}}{\tau_M + \frac{\langle\tau_M\rangle}{A_{\Delta m}}}\right)\right\} - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} \left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) (n-1) e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}}. \end{aligned} \tag{39}$$

The kernel part is explicitly expressed as

$$\begin{aligned} p_{Mm}^{\mathrm{kernel}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) &= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} \left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) \left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle} + 1\right) \\ &\times e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}} \prod_{i=1}^{n} \left\{1 - \left(\frac{\tau_m^{(i)} - \frac{\langle\tau_M\rangle}{A_{\Delta m}}}{\tau_M + \frac{\langle\tau_M\rangle}{A_{\Delta m}}}\right)\right\}. \end{aligned} \tag{40}$$

Note that functions (36)–(40) do not satisfy the normalization condition. Furthermore, in some cases, $p_{Mm}^{\mathrm{approx}}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)})$ in Equations (36) and (39) may take negative values when the correction term is larger than the kernel part. The relationship between the inverse probability density function and its approximation function is discussed in Appendix G.

5 Examination of Bayesian updating method in uncorrelated time series

In this section, we compute the inverse probability density function given by Equation (24) and the (part of) approximation function (Equations (36)–(40)) for the numerically generated uncorrelated time series, and we compare their properties (Tanaka & Umeno, 2021). We examine the numerical method of the Bayesian updating by changing some conditions to see its utility.

5.1 Time series generation and Bayesian updating methods

The uncorrelated time series can be numerically generated by setting $\lambda(t) \equiv \lambda_0$ in Equation (1). In fact, it is numerically generated as the renewal process in which magnitudes and time intervals are generated randomly obeying Equations (9) and (10), respectively. We set the parameter values to be $b = 1$ and $\lambda_0 = 0.0007$. The magnitude thresholds are set to $(M, m) = (5.0, 3.0)$. The b-value condition of Equation (25) is satisfied for these settings. The occurrence time of each event is recorded to 20 decimal places. For such time series, Bayesian updating is applied as explained below.

Bayesian updating is executed for each lower interval in the order of appearance starting from the one immediately after the event with a magnitude above M by substituting their lengths $\{\tau_m^{(1)},\tau_m^{(2)},\cdots,\tau_m^{(n)}\}$ into Equations (24), (39), and (40). The summation of the lower intervals at the n-th update $\sum_{i=1}^{n}\tau_m^{(i)}$ is equivalent to the elapsed time T from the previous event with a magnitude above M. Further, the updating is performed until the event immediately before the next large event with a magnitude above M (i.e., the rightmost lower interval in an upper interval is not used). Therefore, we consider only cases where at least one event is (or two lower intervals are) included in an upper interval.

In addition, we use the following numerical method based on Equations (36) and (37). First, we generate $\mathcal{N}$ time series each contains 10^5 events as sample data. From these sample data, numerically obtain the statistics required for the calculating Equations (36) and (37), i.e., $p_m(\tau_m)$, $p_M(\tau_M)$, $p_{mM}(\tau_m|\tau_M)$ and $P_i(\tau_m|\tau_M)$, and the average number of lower intervals included in the upper interval of length τ_M, $\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M}$. Although the last one is a quantity related to the conditional probability, we calculate it separately. Moreover, we calculate only $P_1(\tau_m|\tau_M)$ and use it instead of $P_i(\tau_m|\tau_M)$ for $i \geq 2$.

These statistics are obtained as a vector or a matrix on discretized intervals as

$$\begin{aligned}\tau_{m,j} &:= 10^{(j+0.5)\Delta\tau_m},\\ \tau_{M,k} &:= 10^{(k+0.5)\Delta\tau_M},\end{aligned} \tag{41}$$

where $j,k \in \mathbb{Z}$, such that

$$\begin{aligned}\boldsymbol{p}_m &= [p_{m,j}]_{j=j_{\min},\cdots,j_{\max}},\\ \boldsymbol{p}_M &= [p_{M,k}]_{k=k_{\min},\cdots,k_{\max}},\\ \boldsymbol{p}_{mM} &= \left[[p_{mM,jk}]_{j=j_{\min}^{(k)},\cdots,j_{\max}^{(k)}}\right]_{k=k_{\min},\cdots,k_{\max}},\\ \boldsymbol{P}_1 &= \left[[P_{1,jk}]_{j=j_{\min}^{(k)},\cdots,j_{\max}^{(k)}}\right]_{k=k_{\min},\cdots,k_{\max}}.\end{aligned} \tag{42}$$

In Equation (42), $j_{\min}, j_{\max}, k_{\min}$, and $k_{\max}$ represent the smallest and largest bin numbers of each distribution. For the statistics obtained as a matrix, the range of j depends on k, and this is indicated as $j_{\min}^{(k)}$ and $j_{\max}^{(k)}$. The ranges of j and k are different for distribution; however, the same symbol is used in Equation (42). In this paper, we fix $\Delta\tau_m = 0.1$, and in this section, we examine the cases $\mathcal{N} = 10^3, 10^5$ and $\Delta\tau_M = 0.1, 0.025$. In the case $\mathcal{N} = 10^5$, they are fully used only for $p_{mM}(\tau_m|\tau_M)$ and $P_1(\tau_m|\tau_M)$, and only 10^3 of them are used for $p_m(\tau_m)$, $p_M(\tau_M)$, and $\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M}$.

To use these amounts in numerical Bayesian updating, we perform the following interpolations between the data points and extrapolations outside the data range. We describe these procedures using the example of the case $\mathcal{N} = 10^3$ and $\Delta\tau_M = 0.1$.

First, for the inter-event time distributions ($\boldsymbol{p}_m$ and $\boldsymbol{p}_M$), we interpolate between the data points of each distribution (between $\tau_{m,j}$ and $\tau_{M,k}$, respectively) using cubic spline functions. Outside the data range (i.e., $\tau_m < \tau_{m,j_{\min}}$, $\tau_m > \tau_{m,j_{\max}}$ and $\tau_M < \tau_{M,k_{\min}}$, $\tau_M > \tau_{M,k_{\max}}$), we extrapolate the fitting curve for the edge 10 points (Figure S1). The distributions are defined for all continuous τ_m values and for all $\tau_{M,k}$ using this process.

Second, for the bivariate distributions ($\boldsymbol{p}_{mM}$ and $\boldsymbol{P}_1$), we perform the same interpolations and extrapolations for $\tau_{m,j}$ (Figures S2 and S3). Meanwhile, for $\tau_{M,k}$, the domain is extended using the average of the functions at $\{\tau_{M,k_{\min}}, \cdots, \tau_{M,k_{\min}+l_e-1}\}$ as the substitute for $\tau_{M,k}$ with $k < k_{\min}$, whereas using the functions at $\{\tau_{M,k_{\max}-l_e+1}, \cdots, \tau_{M,k_{\max}}\}$ as the substitute for $\tau_{M,k}$ with $k > k_{\max}$. We set $l_e = 5$ for $\Delta\tau_M = 0.1$ and $l_e = 20$ for $\Delta\tau_M = 0.025$.

Finally, for $\tau_{M,k}/\langle\langle\tau_m\rangle\rangle_{\tau_{M,k}}$, the interpolation and extrapolation procedures are conducted in the same way as $\boldsymbol{p}_M$, although the extrapolation functions are different (Figure S4).

Thus, the discrete variable $\tau_{m,j}$ becomes continuous as τ_m and the distribution functions are defined for all τ_m larger than 0. This makes it possible to return a value for any input of the length of a lower interval when performing Bayesian updating. Further, the distribution functions are defined for any k in Equation (41). We set the range of k to be $-120 \leq k \leq 70$ for $\Delta\tau_M = 0.1$, and $-480 \leq k \leq 280$ for $\Delta\tau_M = 0.025$. Although this yields the maximum range of the Bayesian updating, the updating at the n-th step is performed within the range $\max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\} < \tau_M$. The properties of the inverse probability density function and the (part of) approximation function are examined within this range.

The kernel parts of the approximation functions are computed by calculating Equation (37) in a step-by-step manner as

$$
\begin{aligned}
\ln p_{Mm}^{\text{kernel}}(\tau_{M,k}|\tau_m^{(1)}) &= \ln\left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\frac{\tau_{M,k}}{\langle\langle\tau_m\rangle\rangle_{\tau_{M,k}}}\right) + \ln p_{mM}(\tau_m^{(1)}|\tau_{M,k}) - \ln p_m(\tau_m^{(1)}) + \ln p_{M,k},\\
\ln p_{Mm}^{\text{kernel}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)}) &= -\ln\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln p_{mM}(\tau_m^{(2)}|\tau_{M,k}) - \ln p_m(\tau_m^{(2)}) + \ln p_{Mm}^{\text{kernel}}(\tau_{M,k}|\tau_m^{(1)}),\\
\ln p_{Mm}^{\text{kernel}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)},\tau_m^{(3)}) &= -\ln\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln p_{mM}(\tau_m^{(3)}|\tau_{M,k}) - \ln p_m(\tau_m^{(3)}) + \ln p_{Mm}^{\text{kernel}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)}),\\
&\vdots
\end{aligned}
\tag{43}
$$

The correction terms of the approximation functions are calculated by first update as

$$
\begin{aligned}
\ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)}) &= \ln\left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(1)}|\tau_{M,k}) - \ln p_m(\tau_m^{(1)}) + \ln p_{M,k},\\
\ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)}) &= -\ln\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(2)}|\tau_{M,k}) - \ln p_m(\tau_m^{(2)}) + \ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)}),\\
\ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)},\tau_m^{(3)}) &= -\ln\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(3)}|\tau_{M,k}) - \ln p_m(\tau_m^{(3)}) + \ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)}),\\
&\ \ \vdots
\end{aligned}
\tag{44}
$$

and then, we add $\ln(n-1)$ for each $\ln p_{Mm}^{\text{correct}}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)})$.

The approximation functions are obtained by adding together the kernel part and the correction term calculated by these separate updates. The approximation functions are calculated only for such k's that $p_{\text{sup}} > \ln p_{Mm}^{\text{kernel}}, \ln p_{Mm}^{\text{correct}} > p_{\text{inf}}$. Here, $p_{\text{sup}}(= 600)$ and $p_{\text{inf}}(= -600)$ yield the upper and lower limits of p_{Mm}^{kernel} and p_{Mm}^{correct} to ensure that these are within the range of the computer capacity. In addition, such k's for which the correction term is so large that Equation (36) becomes negative are excluded.

Figure S5 shows an example of Bayesian updating for the uncorrelated time series. The inverse probability density function given by Equation (24) has a characteristic peak that is not observed in $p_M(\tau_M)$. The correction term makes the kernel part obtained from Equation (40) closer to the inverse probability density function. Moreover, the numerical calculations based on Equations (36) and (37) with $\mathcal{N} = 10^3$ and $\Delta\tau_M = 0.1$ appear to be consistent with these results.

In the next subsection, we compare these functions statistically to examine numerical Bayesian updating method.

5.2 Examination of numerical Bayesian updating method

In this subsection, we compare the probability density functions and the (part of) approximation functions statistically. The Bayesian updating method described in the previous subsection is applied to 100 test data time series, each containing 10^5 events prepared separately from the sample data.

5.2.1 Comparison by distance

We define the distance for two square-integrable functions $f(\cdot)$ and $g(\cdot)$ as

$$D(f||g) := \int_T^\infty |f(\tau_M) - g(\tau_M)|^2\, d\tau_M. \tag{45}$$

The range of the integral is set to (T,∞) to exclude the Dirac's delta function at $\tau_M = T$ in the inverse probability density function. For $f = p_{Mm}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)})$ and $g = p_M(\tau_M)$, the distance can be analytically derived (Appendix H), whereas when $f(\cdot)$ or $g(\cdot)$ is the (part of) approximation function, the distance is calculated numerically as

$$D(f||g) \simeq \sum_{\substack{k;\tau_{M,k}>T \\ p_{\text{sup}}>\ln f,\ln g>p_{\text{inf}}}} |f(\tau_{M,k}) - g(\tau_{M,k})|^2 (\ln 10)\tau_{M,k}\Delta\tau_M. \tag{46}$$

$D(f||g)$ is calculated for each update throughout the 100 test data time series. If no k's satisfy $p_{\text{sup}} > \ln f, \ln g > p_{\text{inf}}$, it is not included in the following calculation. The average distance $\langle D(f||g)\rangle$ is calculated by averaging these distances for each elapsed time $T \in [10^{0.1l}, 10^{0.1(l+1)})$ with $l \in \mathbb{Z}$ from the previous event larger than M.

Figure 7(a) shows the average distance for the cases $f = p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$, $p_{Mm}^{\mathrm{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$, $p_{Mm}^{\mathrm{kernel}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$, and $g = p_M(\tau_M)$. In addition to the analytical calculation in Equation (45) for $D(p_{Mm}||p_M)$, the results of the numerical integration of Equation (46) are presented; the calculations using Equations (39) and (40) are indicated by $D'(\cdot||\cdot)$. The results of the calculation using Equations (36) and (37) with the numerical method in §5.1 with $\mathcal{N} = 10^3$ and $\Delta\tau_M = 0.1$ are presented by $D''(\cdot||\cdot)$. The results for $\mathcal{N} = 10^5$ with $\Delta\tau_M = 0.1$ and $\Delta\tau_M = 0.025$ are shown in Figure S6.

First, one can see that $\langle D'(p_{Mm}||p_M)\rangle$ is almost consistent with $\langle D'(p_{Mm}^{\mathrm{approx}}||p_M)\rangle$, which indicates that $p_{Mm}^{\mathrm{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$ derived in the previous section certainly approximates the inverse probability density function, regardless of the elapsed time (or regardless of the number of updates, because the occurrence rate is constant). However, these separate from $D(p_{Mm}||p_M)$ at around $T \sim 10^5$ and at a large T. As such separations disappear when $\Delta\tau_M = 0.025$ (Figure S6(c,d)), this is attributed to the coarseness of the numerical integration.

Second, $\langle D'(p_{Mm}^{\mathrm{kernel}}||p_M)\rangle$ is nearly consistent with $\langle D''(p_{Mm}^{\mathrm{kernel}}||p_M)\rangle$. This suggests that the numerical updating method in Equation (43) certainly calculates the kernel part. However, $\langle D''(p_{Mm}^{\mathrm{approx}}||p_M)\rangle$ gradually separates from $\langle D'(p_{Mm}^{\mathrm{approx}}||p_M)\rangle$ at a large T. This separation is more clearly illustrated in Figure 7(b), which shows the average distances between $f = p_{Mm}^{\mathrm{approx}}, p_{Mm}^{\mathrm{kernel}}$ and $g = p_{Mm}$ calculated by Equation (46). This separation can be attributed to the calculation of the correction term in Equation (44), in particular to the fluctuation in the numerically obtained $\boldsymbol{P}_1$ (Appendix I).

5.2.2 Comparison by maximum peak time

In the previous subsection, the approximation function calculated by the numerical Bayesian updating method is suggested to be separate from the inverse probability density function. However, we show that such a separation does not have a considerable effect around the maximum peak. To this end, we further compare the maximum points (hereafter, maximum peak time) of the inverse probability density function in Equation (23) and its approximation function in Equation (36) with the numerical updating method, each denoted by $\hat{\tau}_M^{\max}$ and $\tau_M^{\max,\mathrm{approx}}$. Both functions are descretized as Equation (41); the corresponding k in Equation (41) is denoted by $\hat{k}^{\max}$ and $k^{\max,\mathrm{approx}}$, respectively.

$\hat{k}^{\max}$ and $k^{\max,\mathrm{approx}}$ are numerically searched for each update. These are determined as such k that the function takes the maximum value within the range for which the above mentioned numerical results are obtained, while excluding its edges. Thus, if $\hat{k}^{\max}$ or $k^{\max,\mathrm{approx}}$ is located at such edges, it is not considered the peak and is set to $k = 80$ when $\Delta\tau_M = 0.1$ and $k = 320$ when $\Delta\tau_M = 0.025$. Further, when the numerical results of the approximation function are not obtained for any k (when the correction term exceeds the kernel part for all k), $k^{\max,\mathrm{approx}}$ is set to be 80 or 320.

Figure 8 shows the joint probability mass function (p.m.f.) of $(\hat{k}^{\max}, k^{\max,\mathrm{approx}})$ for $\mathcal{N} = 10^3$ and $\Delta\tau_M = 0.1$. Those for $\mathcal{N} = 10^5$ are presented in Figure S7. Here, the population is all the pairs of $(\hat{k}^{\max}, k^{\max,\mathrm{approx}})$ obtained for each update throughout the test data. The maximum peak search is conducted in the two ranges; (a) $\tau_M > \max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$, and (b) $\tau_M > T$. In the former case, the p.m.f. is bimodal; the higher peak exists around $\hat{k}^{\max} = k^{\max,\mathrm{approx}}$, and the other lower peak around $\hat{k}^{\max} > k^{\max,\mathrm{approx}}$. The second peak disappears in the latter case, and the first peak is intrinsic, i.e., the positions of the maximum peak are close between the inverse probability density function and its approximation function. The situation is the same for other cases (Figure S7). These results indicate that it is the off-peak region of the approximation function that contributes to the separation of the average distances.

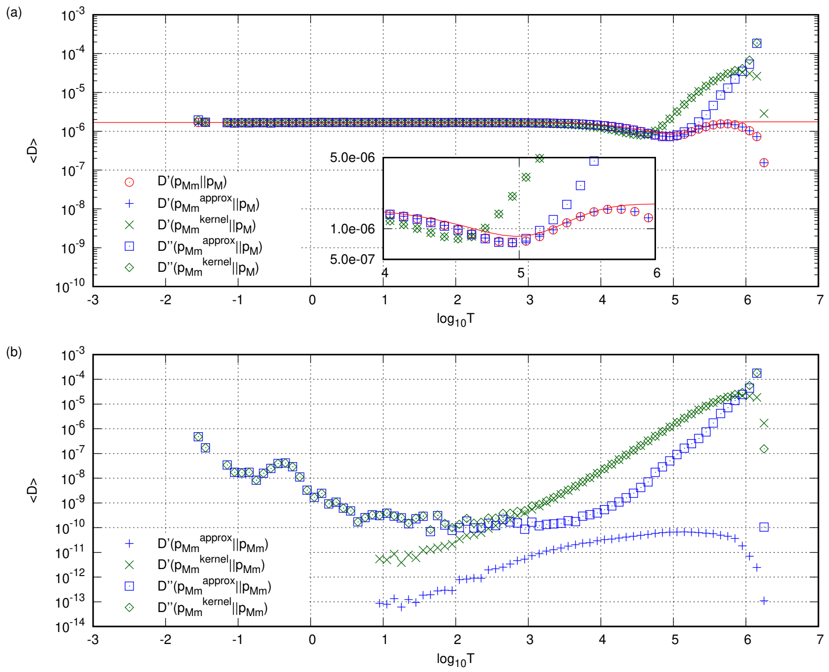

Figure 7. Average distances for each elapsed time (T) from the previous large event with a magnitude above M. (a) Distances between the inter-event time distribution and other function. $D(p_{Mm}||p_M)$ (Equation (H2) in Appendix H) is shown by the red curve, and the symbols are numerical results for Equation (46). (b) Distances between the inverse probability density function and other function numerically calculated by Equation (46).

The results obtained in this section indicate that the numerical method using 1100 time series (1000 for sample data and 100 for test data) is sufficient to calculate the kernel part as well as the maximum peak time of the approximation function that is important in the inference, and to examine their statistical property. Further, these results indicate that Bayesian updating can be applied with the numerical method even if the explicit functional forms of the inter-event time distribution and the conditional probability density function and so on are unclear, such as the time series of the ETAS model.

6 Bayesian updating for the time series of the ETAS model

In this section, Bayesian updating is applied to the time series of the ETAS model (Tanaka & Umeno, 2021). In this case, due to the correlations among events, it is difficult to derive the inverse probability density function and its approximation function analytically. Therefore, we compute the approximation function (Equation (36)) and its kernel part (Equation (37)) using the numerical Bayesian updating method. The maximum peak time of the kernel part is used as the estimate for the occurrence time of the next large event, and the effectiveness of forecasting based on that estimate is evaluated statistically.

6.1 Time series generation and Bayesian updating methods

We apply the numerical Bayesian updating method in §5.1 to the time series generated by Equation (1) with the parameter values $b = 1$, $\alpha = 0.8$, $\theta = 0.2$ ($p = 1.2$), $c = 0.01$, $M_0 = 3$, $\lambda_0 = 0.0007$, and $K = 0.0125$. The magnitude thresholds are

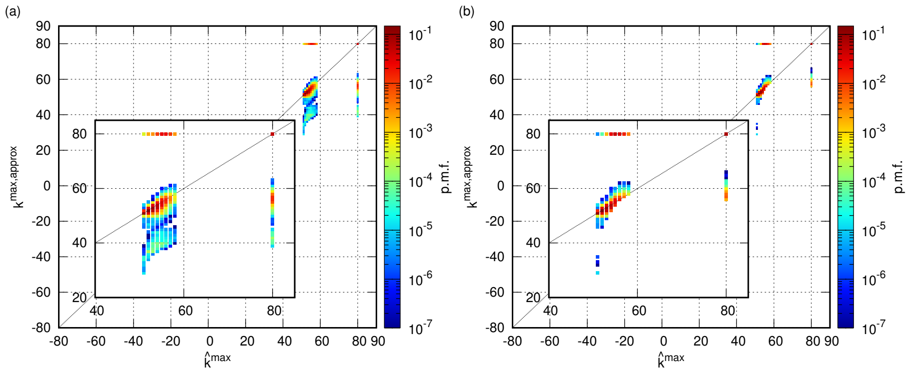

Figure 8. Joint probability mass function for $(\hat{k}^{\max}, k^{\max,\text{approx}})$. Numerical search of the maximum peak is conducted for (a) $\tau_M > \max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ and (b) $\tau_M > T$. The horizontal line at $k^{\max,\text{approx}} = 80$ and the vertical line at $\hat{k}^{\max} = 80$ correspond to the cases when the peak is not detected by the peak search.

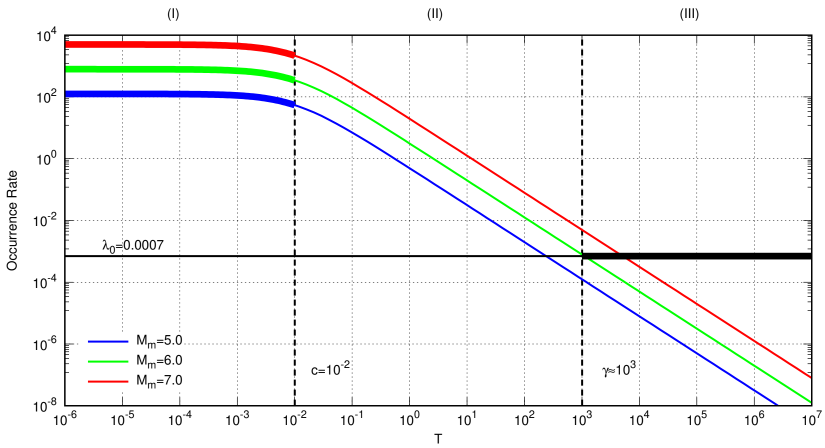

Figure 9. Omori–Utsu law for the parameter values in the text with a different mainshock magnitude M_m. The number of aftershocks per unit day against the elapsed time (T) from the mainshock obeys $\lambda(T) = K10^{\alpha(M_m-M_0)}/(T+c)^{\theta+1}$. The background rate ($\lambda_0 = 0.0007$) is also shown.

$(M, m) = (5.0, 3.0)$. Although the entire time series is stationary because the branching ratio ($n_{\text{br}} \approx 0.785$) is less than 1, it is locally non-stationary obeying the Omori–Utsu law after a large event, as shown in Figure 9. The activity can be categorized into three regimes with respect to the elapsed time (T) from the mainshock, as summarized in Table 1.

We prepare 1100 time series, with each containing 10^5 events. First, random numbers generated from five different seed values are used to generate 240 time series for each seed. Among them, those contain events with magnitude above 10 are excluded. This is because the aftershock sequence excited by such an unrealistic large event do not fit within a single time series, and then, the non-stationarity affects the statistics of the sample data. We use 1100 of the remaining time series. $\mathcal{N} = 1000$ are used as the sample data to obtain statistics with $\Delta\tau_M = 0.1$; the interpolation and extrapolation procedures are conducted with $l_e = 5$ in the same way as explained in §5.1 (Figures S8–S12). Bayesian updating (Equations (43) and (44)) is applied to the remaining 100 time se-

Table 1. *Three Regimes in the time series of the ETAS model*

Category	Regime	Property
(I)	$T \lesssim c(= 0.01)$	Stationary, high occurrence rate
(II)	$c \lesssim T \lesssim \gamma(\approx 10^3)$	Non-stationary, relaxation process
(III)	$\gamma \lesssim T$	Stationary, low occurrence rate ($\lesssim \lambda_0$)

ries. The maximum range of k is $-120 \le k \le 70$, and the n-th update from the occurrence time of the event above M is conduced in the range $\max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\} < \tau_M$. The numerical update is conducted when the lower interval is above 0 (for the occurrence times recorded to 20 decimal places); otherwise, the update is skipped.

The following normalizations are performed in the calculations of the Bayesian updating. The result of the calculation in Equation (43) can be very large. In order to compute the approximation functions together with Equation (44), it is necessary to use the function value of p_{Mm}^{kernel} as it is, though it can exceed p_{sup}. Therefore, to avoid such enlargement, we normalize the result of Equation (43) for each update by subtracting the following numerical integration from Equation (43).

$$\ln\left(\sum_{\substack{k;\tau_{M,k}>T \\ p_{\mathrm{sup}}>\ln p_{Mm}^{\mathrm{kernel}}>p_{\mathrm{inf}}}} p_{Mm}^{\mathrm{kernel}}(\tau_{M,k}|\tau_m^{(1)}, \cdots, \tau_m^{(n)})(\ln 10)\tau_{M,k}\Delta\tau_M\right). \tag{47}$$

Further, it is necessary to subtract Equation (47) from the correction term in Equation (44) at the same time (thereby the entire approximation function is multiplied by a constant). Thus, for each update of Equations (43) and (44), the numerical integration (47) is computed and subtracted from both.

6.2 Comparison of the approximation function and its kernel part

It is difficult to obtain $P_i(\tau_m|\tau_M)$ for the ETAS model, and therefore, we examine the contribution from the correction term to the approximation function as follows. Instead of $P_i(\tau_m|\tau_M)$, we calculate the probability density functions $P^{\mathrm{L}}(\tau_m|\tau_M)$ and $P^{\mathrm{R}}(\tau_m|\tau_M)$ (Figures S10 and S11). According to the Omori–Utsu law, we consider that these two are the end-members of $P_i(\tau_m|\tau_M)$. Then, the approximation functions are calculated numerically by replacing all $P_i(\tau_m|\tau_M)$'s in Equation (44) by either $P^{\mathrm{L}}(\tau_m|\tau_M)$ or $P^{\mathrm{R}}(\tau_m|\tau_M)$. We denote the maximum peak times of these approximation functions by $\tau_M^{\max,\mathrm{L}}$ and $\tau_M^{\max,\mathrm{R}}$, and their corresponding k's in Equation (41) by $k^{\max,\mathrm{L}}$ and $k^{\max,\mathrm{R}}$, respectively. Similarly, they are denoted by $\tau_M^{\max}$ and $k^{\max}$ for the kernel part, hereafter. The numerical search of $k^{\max,\mathrm{L}}$, $k^{\max,\mathrm{R}}$, and $k^{\max}$ is conducted in the same way as indicated in §5.2 in the range $\tau_M > \max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$.

Figure 10 shows the joint p.m.f. of $(k^{\max}, k^{\max,\mathrm{L}})$ and $(k^{\max}, k^{\max,\mathrm{R}})$, which is calculated in the same way as indicated in §5.2. The maximum peak time of the kernel part is not significantly affected by the correction term, and then, it can be used to infer that of the inverse probability density function. However, its confidence interval or average cannot be used because the correction term is not taken into account. In the following, we use the maximum peak time of the kernel part ($\tau_M^{\max}$) as the estimator of the occurrence time of the next large event with a magnitude above M, and we discuss the effectiveness of the forecasting based on the estimates.

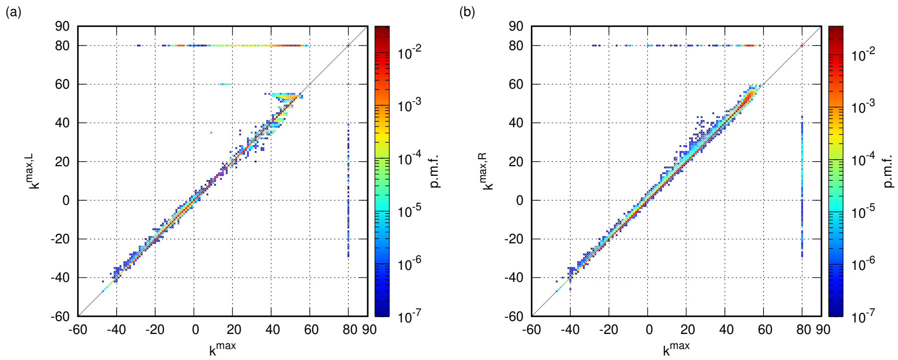

Figure 10. Joint probability mass functions for (a) $\left(k^{\max}, k^{\max,\mathrm{L}}\right)$ and (b) $\left(k^{\max}, k^{\max,\mathrm{R}}\right)$. The horizontal lines at $k^{\max,\mathrm{L}} = 80$ and $k^{\max,\mathrm{R}} = 80$ and the vertical line $k^{\max} = 80$ are the cases where the peak is not detected.

6.3 Estimation of the occurrence time of the next large event and effectiveness of forecasting

We denote the estimate at the n-th update by $\tau_M^{\max,\mathrm{n}} (= 10^{(k^{\max,\mathrm{n}}+0.5)\Delta\tau_M})$, and the actual elapsed time of the next large event from the previous one by τ_M^*. We evaluate the accuracy of the estimation at the n-th update using the relative error (δ_n), which is given as

$$\delta_n := \frac{\tau_M^* - \tau_M^{\max,\mathrm{n}}}{\tau_M^*}, \tag{48}$$

Equation (48) considers that the error $|\tau_M^* - \tau_M^{\max,\mathrm{n}}|$ gets larger as τ_M^* becomes longer. The relative error makes it possible to evaluate the accuracy in a manner that is comparable regardless of τ_M^*.

The accuracy at the n-th update is judged by whether δ_n is within the threshold (δ_{th})

$$-\delta_{\mathrm{th}} \le \delta_n \le \delta_{\mathrm{th}}. \tag{49}$$

When Equation (49) is satisfied, the estimation at the n-th update is judged to be plausible for the given threshold value δ_{th} in the present paper. This is equivalent for the actual occurrence time to be within the range

$$\frac{\tau_M^{\max,\mathrm{n}}}{(1+\delta_{\mathrm{th}})} \le \tau_M^* \le \frac{\tau_M^{\max,\mathrm{n}}}{(1-\delta_{\mathrm{th}})}. \tag{50}$$

Based on the above accuracy at each update, we further evaluate whether a series of estimations yields effective forecasting. Here, effective forecasting implies that $\tau_M^{\max}$ takes a nearly constant value around τ_M^* continuously from well before the occurrence time of the next large event. This can be quantitatively expressed as follows: Let $n_{\le \mathrm{th}}$ be the number of consecutive updates immediately before the next large event, in which Equation (49) is satisfied. Further, we denote the last update as the n_{fin}-th update. When the sequence of updates with a sufficiently long $n_{\le \mathrm{th}}$ exists in the range of $n \in (n_{\mathrm{fin}} - n_{\le \mathrm{th}}, n_{\mathrm{fin}}]$, we consider the forecasting to be effective. We judge the stability of $\tau_M^{\max,\mathrm{n}}$ by Equation (49), and therefore, δ_{th} should not be too large. In the present paper, we set $\delta_{\mathrm{th}} = 0.5$ and 0.25.

To observe the relationship between the effectiveness of forecasting and the stationarity of the time series, we examine the occurrence rate (R_n), variation of its log ($\Delta \log_{10} R_n$), and variation of log-estimate ($\Delta k_n^{\max}$) defined below.

$$R_n := 10/\left(t_{n+9} - t_n\right), \tag{51}$$
$$\Delta \log_{10} R_n := \log_{10} R_{n+10} - \log_{10} R_n, \tag{52}$$
$$\Delta k_n^{\max} := k^{\max,\mathrm{n}+10} - k^{\max,\mathrm{n}}, \tag{53}$$

where t_n represents the occurrence time of the n-th update.

Figures S11–S13 show examples of Bayesian updating for each regime in Table 1. Although these are only examples and not all updating proceeds in this way, these examples suggest that the stability of the estimate is related to the stationarity of the time series.

6.4 Statistical analysis of the effectiveness of forecasting

We show the results of the statistical analysis on the effectiveness of forecasting. Only the cases of $n_{\mathrm{fin}} \geq 30$ are used in the analysis to ensure that the temporal information of lower intervals is fully reflected in the estimate. Figure 11(a) shows the total number of upper intervals (N) obtained from the test data for each $\tau_M^* \in [10^{0.5l}, 10^{0.5(l+1)})$ with $l \in \mathbb{Z}$. Further, N_{30} represents the total number of upper intervals such that $n_{\mathrm{fin}} \geq 30$, which is shown with the ratio to N. The updates included in these N_{30} upper intervals are analyzed.

Figures 11(b–d) show the results of the statistical analysis with $\delta_{\mathrm{th}} = 0.5$. Figure 11(b) shows the probability (P_{fin}) of $n_{\leq \mathrm{th}} > 0$ (or $|\delta_{n_{\mathrm{fin}}}| \leq \delta_{\mathrm{th}}$) for each τ_M^*. The average of P_{fin} for the overall τ_M^* is about 0.52, and the P_{fin} for each τ_M^* is about the same, except for $\tau_M^* > \langle \tau_M \rangle$ in which P_{fin} takes a higher probability around 0.67. Of such $n_{\leq \mathrm{th}} > 0$ cases, the proportion ($P_{\geq 30}$) of those with relatively long $n_{\leq \mathrm{th}} \geq 30$ is also shown in Figure 11(b) (the probability distribution of $n_{\leq \mathrm{th}}$ is shown in Figure S16(a)). Thus the regions of high $P_{\geq 30}$ are overlapped with regimes (I) and (III), though the former is shifted toward larger τ_M^*. On the other hand, $P_{\geq 30}$ is lower in regime (II); it gradually decreases as τ_M^* gets larger. This is consistent with the average of $n_{\leq \mathrm{th}}$ ($\langle n_{\leq \mathrm{th}} \rangle$, this average is taken for $n_{\leq \mathrm{th}} > 0$), but also with the average of its proportion to n_{fin} ($\langle n_{\leq \mathrm{th}}/n_{\mathrm{fin}} \rangle$) as shown in Figure 11(c). This implies that, as the fraction of non-stationary times in $[0, \tau_M^*)$ increases in regime (II), the domination rate of $n_{\leq \mathrm{th}}$ in the total n_{fin}-updates decreases gradually. These properties are preserved for $\delta_{\mathrm{th}} = 0.25$ (Figure S17).

Figure 12 shows the joint probability density-mass functions of $\Delta \log_{10} R$ and $\Delta k^{\max}$ calculated numerically for each τ_M^*. The case $k^{\max} = 80$ is excluded from the population. If τ_M^* is in the regions of high $P_{\geq 30}$, the distribution is almost symmetrically concentrated near the origin. This implies that, when the time series is dominated by stationarity ($\Delta \log_{10} R \approx 0$), the estimated value is stable ($\Delta k^{\max} \approx 0$). On the other hand, if τ_M^* is in regime (II), the probability density function gradually has a region in the second quadrant as τ_M^* gets larger. This region indicates the existence of a non-stationary time series in which the estimate has an increasing trend ($\Delta k^{\max} > 0$).

These results present the following conclusions. First, the probability that the relative error is within the threshold at the last update ($|\delta_{\mathrm{n_{fin}}}| \leq \delta_{\mathrm{th}}$) is almost independent of the actual occurrence time (τ_M^*). This suggests that the length of the upper interval can be estimated by the inverse probability density function reflecting the temporal pattern of lower intervals, at the last update when the information of the lower intervals can be utilized fully. Second, the stationarity of the time series is related to the stability of the estimate; if the time series is non-stationary, it causes the estimate $\tau_M^{\max}$ to shift. Third, the domination rate of stationarity in the time series determines the effectiveness of forecasting. Immediately or long after the large event, the stationary time series is dominant. Therefore, based on the second point mentioned above, the estimate becomes stable, which leads to an effective forecasting with a relatively long $n_{\leq \mathrm{th}}$. However, these regions are not identical to regimes (I) and (Ⅲ). This is attributed to lag until the ratio of the non-stationary region in the time series becomes dominant. On the other hand, in regime (II), the non-stationarity becomes gradually dominant, which leads to the shifting of $\tau_M^{\max}$ and shortening of $n_{\leq \mathrm{th}}$.

Finally, we discuss the effectiveness of forecasting in terms of the duration time ($\tau_{\leq \mathrm{th}}$) during the $n_{\leq \mathrm{th}}$ updates. Figure 11(d) shows the average of the duration time ($\langle \tau_{\leq \mathrm{th}} \rangle$) and the average of its ratio to the actual occurrence time ($\langle \tau_{\leq \mathrm{th}}/\tau_M^* \rangle$) for each τ_M^* (the probability distribution of $\tau_{\leq \mathrm{th}}$ is shown in Figure S16(b)). Unlike $\langle n_{\leq \mathrm{th}} \rangle$ in Figure 11(c), $\langle \tau_{\leq \mathrm{th}} \rangle$ increases linearly as τ_M^* gets larger, and it is sufficiently long in regime (Ⅲ). On the other hand, $\tau_{\leq \mathrm{th}}$ is very short in regime (I); however, the ratio $\langle \tau_{\leq \mathrm{th}}/\tau_M^* \rangle$ is high (nearly 0.7). Therefore, from the perspective of the time interval, the forecasting is also considered to be effective immediately or long after the large event.

7 Discussion and Conclusions

The Bayes' theorem and Bayesian updating on the inter-event times at different magnitude thresholds in a marked point process are considered. The analytical results for the uncorrelated time series are used to apply Bayesian updating to the time series of the ETAS model for examining its utility toward forecasting a large event using the temporal pattern of the smaller events.

First, the Bayes' theorem is considered for the general marked point process. The Bayes' theorem provides the relationship between the conditional and inverse probability density functions for the lengths of one upper interval and one lower interval. The inverse probability density function is represented by the generalized forms of the probability density functions of the inter-event times and the conditional probability density function. This inverse probability density function is derived for the uncorrelated time series analytically, and the condition to have a peak is also found.

The Bayes' theorem is extended to Bayesian updating that yields the inverse probability density function between the lengths of multiple consecutive lower intervals and the upper interval that includes them. Although the inverse probability density function is different for the updating manner, we consider the updating without the restriction on the position of the lower intervals. For the uncorrelated time series, the inverse probability density function and its approximation function are derived, and the latter approximation is shown to be reasonable using the distances.

Bayesian updating is applied to the time series of the ETAS model. We numerically analyze the approximation function and its kernel part. We use the maximum point of the kernel part as the estimate of the occurrence time of the next large event because the maximum peaks of these two functions are shown to not be different drastically. The accuracy of the estimation at each update is evaluated by the relative error with the actual occurrence time of the next large event; the effectiveness of the forecasting throughout the series of updates is judged by the continuity of the plausible estimations prior to the large event.

Statistical analysis indicates that the accuracy of the estimation at the last update does not drastically depend on the occurrence time of the next large event. This suggests that the inverse probability density function can estimate the occurrence time of the next large event in response to the temporal pattern of minor events. However, the continuity of plausible estimation depends on the occurrence time of the next large event. This is because the dominance rate of the non-stationary time series in which the estimate becomes unstable varies with the elapsed time from the previous large event obeying the Omori–Utsu law. The stationarity is dominant either immediately after or long

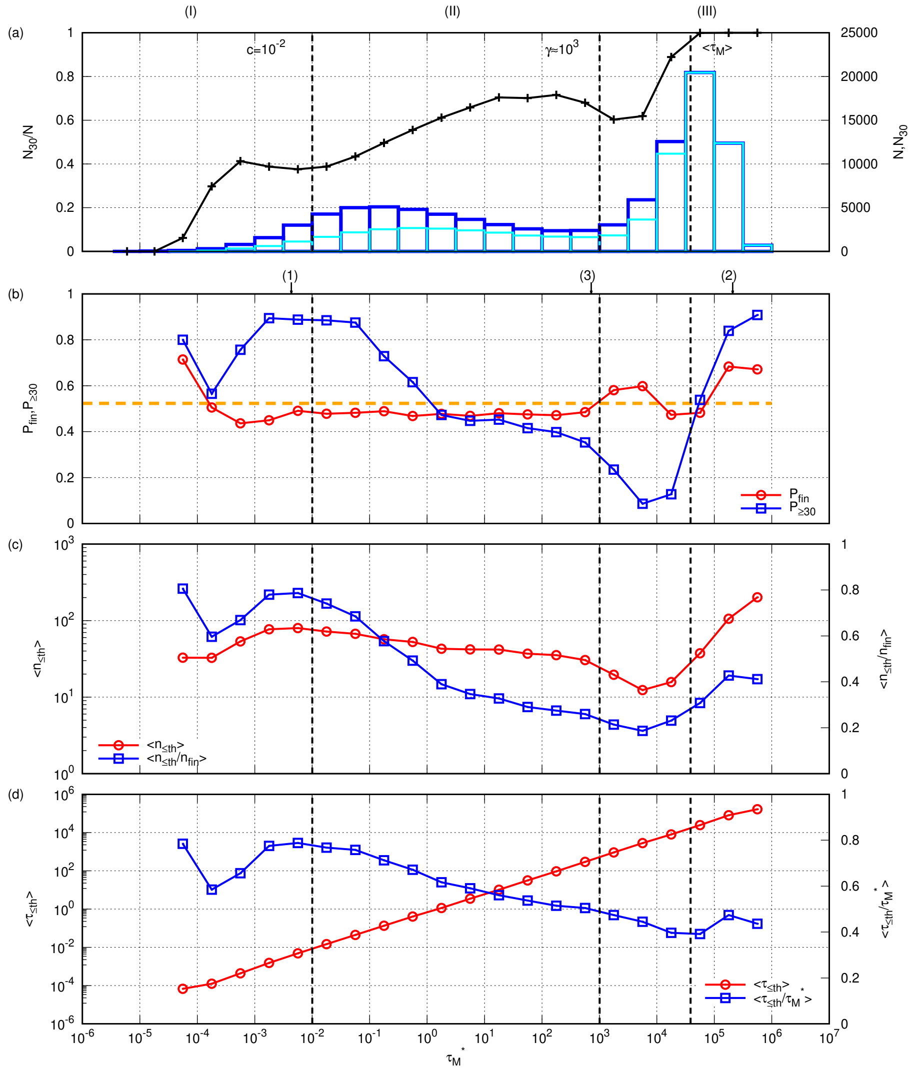

Figure 11. (a) (Blue) Number of upper intervals N, (Cyan) number of upper intervals that include at least 30 updates N_{30}, and (Black) their ratio N_{30}/N, for each τ_M^* in the test data. (b–d) Statistical results with $\delta_{\rm th} = 0.5$ for each τ_M^*. (1)–(3) indicates the τ_M^*'s of the examples in Figures S13–S15. (b) (Red) Probability $P_{\rm fin}$ that $n_{\leq\rm th} > 0$ holds (or the probability that Equation (49) is satisfied at the last ($n_{\rm fin}$-th) update), and (Orange dotted line) the average of $P_{\rm fin} \approx 0.52$ for the overall τ_M^*. (Blue) Proportion $P_{\geq 30}$ of such cases among them where $n_{\leq\rm th} \geq 30$. (c) (Red) Average of $n_{\leq\rm th}$, $\langle n_{\leq\rm th}\rangle$, and (Blue) the average of its proportion to the total number of updates, $\langle n_{\leq\rm th}/n_{\rm fin}\rangle$. (d) (Red) Average of the duration time $\tau_{\leq\rm th}$, $\langle\tau_{\leq\rm th}\rangle$ and (Blue) the average of its proportion to the actual occurrence time, $\langle\tau_{\leq\rm th}/\tau_M^*\rangle$.

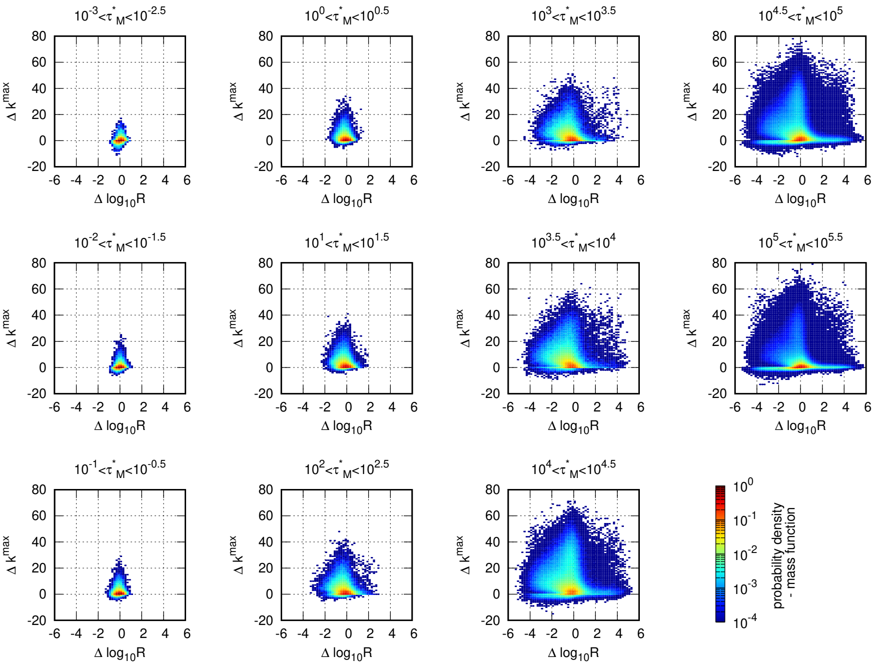

Figure 12. Joint probability density-mass functions of $\Delta \log_{10} R$ and $\Delta k^{\max}$ for each τ_M^*.

after the previous major event. Therefore, the forecasting by the Bayesian updating method can be effective for secondary disaster prevention in the former case, and for long-term risk assessment in the latter case.

The approximation function derived for the uncorrelated time series is applied in the Bayesian updating for the time series of the ETAS model. This allows us to perform the update in the convenient form of the product of the conditional probabilities. However, this implicitly assumes that there is no correlation between events and lower intervals; such an assumption can be reasonable for the stationary part of the time series, although it is not reasonable for non-stationary part. This probably is one of the reasons why forecasting is ineffective in the non-stationary regime.

Further, only the kernel part of the approximation function is used when estimating the occurrence time of the next large event for the time series of the ETAS model. The correction term needs to be investigated in detail to use the entire approximation function to perform point estimation by average, interval estimation, or probabilistic risk assessment using hazard rate. Comparison with the probabilistic evaluation using the inter-event time distribution becomes possible after clarifying the correction term.

Although the statistical property of Bayesian updating is examined for only one set of ETAS parameters, it is considered to be different for activities generated by other parameter values. For example, for the time series with the high background rate (λ_0) that corresponds to taking up a large spatial area, forecasting is considered to be less effective because in such time series, different mainshock-aftershocks sequences overlap (Touati et al., 2009) and the correlations between the upper and lower intervals are weakened. Further, if the background rate is low, forecasting is considered to be improved because a single mainshock-aftershocks sequence is exposed (Touati et al., 2009) and the correlation is easily reflected in the conditional probability. Forecasting is also considered to be improved for the time series with a large branching ratio (n_{br}). If the branching ratio is large, the number of aftershocks generated by an event increases (Helmstetter

& Sornette, 2003), which increases the number of updates in the Bayesian updating; this is advantageous for forecasting.

In this study, only one lower threshold magnitude (m) is set for a given upper threshold (M). One approach for performing Bayesian updating using more temporal information of the lower intervals is to set multiple lower thresholds ($m_1 < m_2 < \cdots (< M)$). When setting such lower thresholds, it is better to refer to the condition of $\Delta m > \log_{10} 3/b$ so that the inverse probability density function has a peak. This condition indicates that there is a trade-off between the b-value and Δm, and then, the range of lower thresholds that can be set varies with the b-value. However, the condition $\Delta m > \log_{10} 3/b$ is for the uncorrelated time series; finding the corresponding condition for the time series of the ETAS model is a future work. Considering such an extension is important for applying the Bayesian updating method to the real seismic catalogs in which the number of earthquakes is limited.

To utilize as much temporal information on small events as possible using the method described above is important for applying the Bayesian updating method to real seismic catalogs with a limited number of data. Another idea to apply the Bayesian updating method to seismic catalogs while compensating for the shortage of data is to use the ETAS model in combination. The ETAS model can be used to generate a sufficient amount of synthetic data with the parameter set determined for the past seismic activity. From such synthetic data, statistical amounts necessary in the numerical Bayesian updating method is obtained precisely. Moreover, it is necessary to develop further ingenuity by studying the properties of the conditional and inverse probability density functions through the analysis of seismic catalogs. With these auxiliaries, the application of the Bayesian updating method to real seismic activity is expected to proceed while solving the limitation of seismic data.

Appendix A Derivation of the conditional and inverse probability density functions for the uncorrelated time series

First, we derive the conditional probability density function (Equation (15)) by substituting Equations (12)–(14) into Equation (11). The denominator of Equation (11) is

$$\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M) = A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle} + 1,$$

and the numerator is

$$\begin{aligned}
&\sum_{i=1}^{\infty} i\rho_{mM}(\tau_m|i,\tau_M)\Psi_{mM}(i|\tau_M) = e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m) \\
&+ e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\frac{A_{\Delta m}}{\langle\tau_M\rangle}\sum_{i=0}^{\infty}(i+2)\frac{\left\{A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m}{\tau_M}\right)\right\}^i}{i!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m}{\tau_M}\right)}\theta(\tau_M-\tau_m) \\
&= e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m) + e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left\{A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m}{\tau_M}\right)+2\right\}\theta(\tau_M-\tau_m).
\end{aligned}$$

Equation (15) is obtained by rearranging the above equations.

We confirm that Equation (2) with this conditional probability in its kernel has the exponential distribution (Equation (10)) as the solution. By dividing both sides of Equation (2) by N_m and rewriting it using $N_M/N_m = \langle\tau_m\rangle/\langle\tau_M\rangle$ as well as Equation (15)

$$\begin{aligned}
p_m(\tau_m) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\int_{\tau_m}^{\infty}\Big[&e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m) \\
&+\frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\left\{A_{\Delta m}\frac{\tau_M-\tau_m}{\langle\tau_M\rangle}+2\right\}\theta(\tau_M-\tau_m)\Big]p_M(\tau_M), \qquad \text{(A1)}
\end{aligned}$$

where the following general relation is used.

$$\frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} = \sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M).$$

We show that the r.h.s. of Equation (A1) is equivalent to the l.h.s., $p_m(\tau_m) = e^{-\frac{\tau_m}{\langle\tau_m\rangle}}/\langle\tau_m\rangle$. Substitute $p_M(\tau_M) = e^{-\frac{\tau_M}{\langle\tau_M\rangle}}/\langle\tau_M\rangle$ into the r.h.s. of Equation (A1) and note that $A_{\Delta m}+1 = \langle\tau_M\rangle/\langle\tau_m\rangle$; the integral involving the delta function (R_1) is

$$R_1 = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} e^{-\frac{\tau_m}{\langle\tau_m\rangle}}, \tag{A2}$$

and the integral involving the step function (R_2) is

$$\begin{aligned} R_2 &= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^3} A_{\Delta m} e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}} \int_{\tau_m}^{\infty} \left\{ A_{\Delta m}\frac{\tau_M - \tau_m}{\langle\tau_M\rangle} + 2 \right\} e^{-\frac{\tau_M}{\langle\tau_M\rangle}} d\tau_M \\ &= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} A_{\Delta m}(A_{\Delta m}+2) e^{-\frac{\tau_m}{\langle\tau_m\rangle}}. \end{aligned} \tag{A3}$$

Therefore, the r.h.s. of Equation (A1) is shown to be equivalent to the l.h.s. of Equation (A1) as follows:

$$\begin{aligned} R_1 + R_2 &= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} \left(1 + A_{\Delta m}\right)^2 e^{-\frac{\tau_m}{\langle\tau_m\rangle}} \\ &= \frac{1}{\langle\tau_m\rangle} e^{-\frac{\tau_m}{\langle\tau_m\rangle}}. \end{aligned} \tag{A4}$$

Second, we derive the inverse probability density function (Equation (16)). From Equation (15), the generalized probability density functions for the uncorrelated time series are derived as

$$\begin{aligned} z_m(\tau_m) &= \frac{\tau_m}{\langle\tau_m\rangle^2} e^{-\frac{\tau_m}{\langle\tau_m\rangle}}, \\ z_M(\tau_M) &= \frac{\tau_M}{\langle\tau_M\rangle^2} e^{-\frac{\tau_M}{\langle\tau_M\rangle}}, \\ z_{mM}(\tau_m|\tau_M) &= \frac{\tau_m}{\tau_M} e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}} \left[\delta(\tau_M - \tau_m) + \frac{A_{\Delta m}}{\langle\tau_M\rangle} \left\{ A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1 - \frac{\tau_m}{\tau_M}\right) + 2 \right\} \theta(\tau_M - \tau_m) \right]. \end{aligned}$$

Equation (16) is obtained by substituting the above equations in Equation (8).

Derivative of Equation (16) by τ_M is

$$\frac{\partial}{\partial\tau_M} p_{Mm}(\tau_M(>\tau_m)|\tau_m) = -\frac{\langle\tau_m\rangle^2}{\langle\tau_M\rangle^5} A_{\Delta m}^2 e^{\frac{\tau_m - \tau_M}{\langle\tau_M\rangle}} \left[\tau_M - \left\{ \tau_m + \langle\tau_M\rangle \left(1 - \frac{2}{A_{\Delta m}}\right) \right\} \right].$$

Therefore, the inverse probability density function has a peak at

$$\tau_M^{\max} = \tau_m + \langle\tau_M\rangle \left(1 - \frac{2}{A_{\Delta m}}\right),$$

under the condition of $\tau_M^{\max} > \tau_m$, which is equivalent to

$$\Delta m > \frac{\log_{10} 3}{b}.$$

Appendix B Derivation of Equation (24) from Equation (23)

The summation part in the r.h.s. of Equation (23) is

$$\sum_{i=n}^{\infty}(i-n+1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M)\prod_{j=2}^{n}\rho_{mM}\left(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right)$$
$$=\Psi_{mM}(n|\tau_M)\rho_{mM}(\tau_m^{(1)}|n,\tau_M)\prod_{j=2}^{n}\rho_{mM}\left(\tau_m^{(j)}|n-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right)$$
$$+\sum_{i=n+1}^{\infty}(i-n+1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M)\prod_{j=2}^{n}\rho_{mM}\left(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right).$$

The first term on the r.h.s. of the above equation is transformed by substituting Equations (12)–(14) as

$$\frac{\left[A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right]^{n-1}}{(n-1)!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\frac{(n-1)}{\tau_M}\left(\frac{\tau_M-\tau_m^{(1)}}{\tau_M}\right)^{n-2}\frac{(n-2)}{\tau_M-\tau_m^{(1)}}\left(\frac{\tau_M-\tau_m^{(1)}-\tau_m^{(2)}}{\tau_M-\tau_m^{(1)}}\right)^{n-3}$$
$$\cdots\frac{\delta(\tau_M-\sum_{i=1}^{n}\tau_m^{(i)})}{\tau_M-\sum_{i=1}^{n-2}\tau_m^{(i)}}=\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{n-1}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta\left(\tau_M-\sum_{i=1}^{n}\tau_m^{(i)}\right). \tag{B1}$$

The second term except the step function is also transformed as

$$\sum_{i=n+1}^{\infty}(i-n+1)\frac{\left[A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right]^{i-1}}{(i-1)!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\frac{(i-1)}{\tau_M}\left(\frac{\tau_M-\tau_m^{(1)}}{\tau_M}\right)^{i-2}$$
$$\times\prod_{j=2}^{n}\frac{(i-j)}{\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}}\left(\frac{\tau_M-\sum_{k=1}^{j}\tau_m^{(k)}}{\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}}\right)^{i-j-1}$$
$$=\sum_{i=n+1}^{\infty}\frac{(i-n+1)}{(i-n-1)!}\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{i-1}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\left(\tau_M-\sum_{k=1}^{n}\tau_m^{(k)}\right)^{i-n-1}$$
$$=\sum_{i=0}^{\infty}\frac{i+2}{i!}\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{i+n}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\left(\tau_M-\sum_{k=1}^{n}\tau_m^{(k)}\right)^{i}$$
$$=\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{n}e^{-A_{\Delta m}\frac{\sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}}\sum_{i=0}^{\infty}\frac{i+2}{i!}\left\{\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left(\tau_M-\sum_{k=1}^{n}\tau_m^{(k)}\right)\right\}^{i}e^{-A_{\Delta m}\frac{\tau_M-\sum_{k=1}^{n}\tau_m^{(k)}}{\langle\tau_M\rangle}}$$
$$=\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{n}e^{-A_{\Delta m}\frac{\sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}}\left(A_{\Delta m}\frac{\tau_M-\sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}+2\right). \tag{B2}$$

Finally, Equation (24) is obtained by substituting Equations (B1) and (B2) in Equation (23), with the denominator of the r.h.s. of Equation (23)

$$\prod_{i=1}^{n}p_m(\tau_m^{(i)})=\frac{1}{\langle\tau_m\rangle^n}e^{-\frac{\sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_m\rangle}}.$$

Appendix C Another Bayesian updating method

In this appendix, we consider another method of Bayesian updating from the one introduced in §4; this method considers the consecutive lower intervals in the order of the appearance from the last event with magnitude greater than M. We derive the inverse probability density function for this updating method in the uncorrelated time series.

Let $N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ be the total number of such upper intervals of length τ_M that include the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ start from the leftmost one in the upper interval. Further, we denote the inverse probability density function for this updating by $p^*_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$. We derive it by representing $N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ in two ways as follows:

First, we derive $N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ by counting the total number of the upper intervals of length τ_M that include the leftmost consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$. The position of the first interval in the sequence of the consecutive lower intervals is fixed at the leftmost one in an upper interval, and therefore, the number of the sequence $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ in the time series is

$$N_M \prod_{i=1}^{n} p_m(\tau_m^{(i)}) d\tau_m^n.$$

Among them, the number of sequences that belong to the same upper interval is

$$N_M \left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1} \prod_{i=1}^{n} p_m(\tau_m^{(i)}) d\tau_m^n.$$

Therefore, the first representation is obtained as

$$\begin{aligned} &N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\ &\quad = N_M \left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1} \left\{\prod_{i=1}^{n} p_m(\tau_m^{(i)})\right\} p^*_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) d\tau_M d\tau_m^n. \end{aligned}$$

This equation is rewritten using Equation (10) in the explicit form as

$$\begin{aligned} &N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\ &\quad = N_M \left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1} \frac{1}{\langle \tau_m \rangle^n} e^{-\frac{\sum_{i=1}^{n} \tau_m^{(i)}}{\langle \tau_m \rangle}} p^*_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) d\tau_M d\tau_m^n. \end{aligned} \tag{C1}$$

Second, we derive $N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ by counting the total number of consecutive lower intervals that start from the leftmost one in the upper intervals of length τ_M. There is only one way for the sequence of consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ to be involved in each of the $N_M p_M(\tau_M) d\tau_M$ upper intervals of length τ_M. The probability of the occurrence of that sequence in the upper interval is, when $i(\geq n)$-lower intervals are included in it

$$\rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right) d\tau_m^n.$$

Therefore, the second representation is obtained as

$$\begin{aligned} &N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\ &= N_M p_M(\tau_M) \sum_{i=n}^{\infty} \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right) d\tau_M d\tau_m^n. \end{aligned}$$

This equation is rewritten in the explicit form using Equations (12)–(14) in the same way as in Appendix B.

$$\begin{aligned} &N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) = N_M \frac{1}{\langle \tau_M \rangle} e^{-\frac{\tau_M}{\langle \tau_M \rangle}} d\tau_M d\tau_m^n \left(\frac{A_{\Delta m}}{\langle \tau_M \rangle}\right)^{n-1} \\ &\times \left\{ e^{-A_{\Delta m} \frac{\tau_M}{\langle \tau_M \rangle}} \delta\left(\tau_M - \sum_{i=1}^{n} \tau_m^{(i)}\right) + \left(\frac{A_{\Delta m}}{\langle \tau_M \rangle}\right) e^{-A_{\Delta m} \frac{\sum_{i=1}^{n} \tau_m^{(i)}}{\langle \tau_M \rangle}} \theta\left(\tau_M - \sum_{i=1}^{n} \tau_m^{(i)}\right) \right\}. \end{aligned} \tag{C2}$$

Finally, $p^*_{Mm}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)})$ is derived from Equations (C1) and (C2) as

$$p^*_{Mm}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)}) \\ = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\left\{\frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-\frac{\tau_M-\sum_{i=1}^n\tau_m^{(i)}}{\langle\tau_M\rangle}}\theta\left(\tau_M-\sum_{i=1}^n\tau_m^{(i)}\right)+e^{-\frac{\tau_M-\sum_{i=1}^n\tau_m^{(i)}}{\langle\tau_m\rangle}}\delta\left(\tau_M-\sum_{i=1}^n\tau_m^{(i)}\right)\right\}. \tag{C3}$$

This is different from Equation (24), which reflects the difference whether the position of lower intervals is specified.

Appendix D Derivation of Equation (29)

First, we substitute Equations (12)–(14) into Equation (28)

$$\begin{aligned}P^{\mathrm{R}}(\tau_m|\tau_M) &= P^{\mathrm{L}}(\tau_m|\tau_M)\\ &= e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m)\\ &+\sum_{i=2}^{\infty}\frac{(i-1)}{\tau_M}\left(1-\frac{\tau_m}{\tau_M}\right)^{i-2}\frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right)^{i-1}}{(i-1)!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\theta(\tau_M-\tau_m).\end{aligned}$$

In the above equation, the summation part of the term including the step function can be transformed as

$$\begin{aligned}&\sum_{i=2}^{\infty}\frac{(i-1)}{\tau_M}\left(1-\frac{\tau_m}{\tau_M}\right)^{i-2}\frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right)^{i-1}}{(i-1)!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\\ &= \frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\sum_{i=0}^{\infty}\frac{\left\{A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m}{\tau_M}\right)\right\}^i}{i!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m}{\tau_M}\right)}\\ &= \frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}.\end{aligned}$$

Finally, Equation (29) is obtained by rearranging the above equations.

Appendix E Derivation of Equation (34)

In this appendix, $P(\tau_m^{(1)},\cdots,\tau_m^{(l)}|\tau_M)$ is derived for the uncorrelated time series. First, we divide the summation in Equation (33) into two parts:

$$\begin{aligned}P(\tau_m^{(1)},\cdots,\tau_m^{(l)}|\tau_M) &= \sum_{i=l}^{\infty}\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M)\prod_{j=2}^{l}\rho_{mM}\left(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right)\\ &= \Psi_{mM}(l|\tau_M)\rho_{mM}(\tau_m^{(1)}|l,\tau_M)\prod_{j=2}^{l}\rho_{mM}\left(\tau_m^{(j)}|l-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(j)}\right)\\ &+\sum_{i=l+1}^{\infty}\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M)\prod_{j=2}^{l}\rho_{mM}\left(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right)\end{aligned}$$

This equation is further rewritten by substituting Equations (12)–(14) in the same way as in Appendix B. The second term on the r.h.s. except for the step function is

$$\begin{aligned}
&\sum_{i=l+1}^{\infty} \frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right)^{i-1}}{(i-1)!} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}} \frac{i-1}{\tau_M} \left(\frac{\tau_M-\tau_m^{(1)}}{\tau_M}\right)^{i-2} \\
&\times \prod_{j=2}^{l} \frac{(i-j)}{\tau_M - \sum_{k=1}^{j-1}\tau_m^{(k)}} \left(\frac{\tau_M - \sum_{k=1}^{j}\tau_m^{(k)}}{\tau_M - \sum_{k=1}^{j-1}\tau_m^{(k)}}\right)^{i-j-1} \\
&= \sum_{i=l+1}^{\infty} \frac{1}{(i-l-1)!} \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{i-1} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}} \left(\tau_M - \sum_{k=1}^{l}\tau_m^{(k)}\right)^{i-l-1} \\
&= \sum_{i=0}^{\infty} \frac{1}{i!} \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{i+l} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}} \left(\tau_M - \sum_{k=1}^{l}\tau_m^{(k)}\right)^{i} \\
&= \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{l} e^{-A_{\Delta m}\frac{\sum_{i=1}^{l}\tau_m^{(i)}}{\langle\tau_M\rangle}} \sum_{i=0}^{\infty} \frac{\left\{\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left(\tau_M - \sum_{k=1}^{l}\tau_m^{(k)}\right)\right\}^{i}}{i!} e^{-\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left(\tau_M - \sum_{k=1}^{l}\tau_m^{(k)}\right)} \\
&= \prod_{i=1}^{l} P_i(\tau_m^{(i)}|\tau_M),
\end{aligned}$$

563 where $P_i(\tau_m^{(i)}|\tau_M) := \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right) e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}$.

Therefore

$$P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M) = \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{l-1} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}} \delta\left(\tau_M - \sum_{i=1}^{l}\tau_m^{(i)}\right) + \prod_{i=1}^{l} P_i(\tau_m^{(i)}|\tau_M)\theta\left(\tau_M - \sum_{i=1}^{l}\tau_m^{(i)}\right).$$

Finally, because $\tau_M \gneqq \sum_{i=1}^{l}\tau_m^{(i)}$ holds for $l < n$ by the condition of Equation (26)

$$P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M) = \prod_{i=1}^{l} P_i(\tau_m^{(i)}|\tau_M).$$

564 Appendix F Derivation of Equation (39)

In this appendix, the approximation function of the inverse probability density function for the uncorrelated time series (Equation (39)) is derived. By substituting Equations (10), (15), and (34) into Equation (36)

$$\begin{aligned}
&p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\
&= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}+1\right)}{\left(A_{\Delta m}\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}} \left[\prod_{i=1}^{n} \frac{e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}} \frac{A_{\Delta m}}{\langle\tau_M\rangle}\left\{A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m^{(i)}}{\tau_M}\right)+2\right\}}{\frac{1}{\langle\tau_m\rangle}e^{-\frac{\tau_m^{(i)}}{\langle\tau_m\rangle}}\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}+1\right)}\right] \frac{1}{\langle\tau_M\rangle} e^{-\frac{\tau_M}{\langle\tau_M\rangle}} \\
&- \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{(n-1)}{\left(A_{\Delta m}\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}} \left[\prod_{i=1}^{n} \frac{\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}}{\frac{1}{\langle\tau_m\rangle}e^{-\frac{\tau_m^{(i)}}{\langle\tau_m\rangle}}}\right] \frac{1}{\langle\tau_M\rangle} e^{-\frac{\tau_M}{\langle\tau_M\rangle}},
\end{aligned} \tag{F1}$$

where the following relation is used.

$$\begin{aligned}
\frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} &= \sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M) \\
&= A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle} + 1.
\end{aligned}$$

The two products ($[\cdots]$'s) in Equation (F1) are respectively transformed as

$$\prod_{i=1}^{n} \frac{\left(\frac{A_{\Delta m}}{\langle \tau_M \rangle}\right) e^{-A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle}}}{\frac{1}{\langle \tau_m \rangle} e^{-\frac{\tau_m^{(i)}}{\langle \tau_m \rangle}}} = \left(A_{\Delta m} \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^n \prod_{i=1}^{n} e^{-A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle} + \frac{\tau_m^{(i)}}{\langle \tau_m \rangle}}$$
$$= \left(A_{\Delta m} \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^n \prod_{i=1}^{n} e^{\frac{\tau_m^{(i)}}{\langle \tau_M \rangle}}. \tag{F2}$$

$$\prod_{i=1}^{n} \frac{e^{-A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle}} \frac{A_{\Delta m}}{\langle \tau_M \rangle} \left\{ A_{\Delta m} \frac{\tau_M}{\langle \tau_M \rangle} \left(1 - \frac{\tau_m^{(i)}}{\tau_M}\right) + 2 \right\}}{\frac{1}{\langle \tau_m \rangle} e^{-\frac{\tau_m^{(i)}}{\langle \tau_m \rangle}} \left(A_{\Delta m} \frac{\tau_M}{\langle \tau_M \rangle} + 1\right)}$$
$$= \left(A_{\Delta m} \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^n \left(\prod_{i=1}^{n} e^{-A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle} + \frac{\tau_m^{(i)}}{\langle \tau_m \rangle}}\right) \left\{ \prod_{i=1}^{n} \frac{A_{\Delta m} \frac{\tau_M}{\langle \tau_M \rangle} + 1 - \left(A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle} - 1\right)}{A_{\Delta m} \frac{\tau_M}{\langle \tau_M \rangle} + 1} \right\}$$
$$= \left(A_{\Delta m} \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^n \left(\prod_{i=1}^{n} e^{\frac{\tau_m^{(i)}}{\langle \tau_M \rangle}}\right) \left\{ \prod_{i=1}^{n} \left(1 - \frac{\tau_m^{(i)} - \frac{\langle \tau_M \rangle}{A_{\Delta m}}}{\tau_M + \frac{\langle \tau_M \rangle}{A_{\Delta m}}}\right) \right\}. \tag{F3}$$

Finally, Equation (39) is derived by substituting Equations (F2) and (F3) into Equation (F1).

Appendix G Relation between the inverse probability density function and its approximation function in the uncorrelated time series

In this appendix, we discuss the relation between the inverse probability density function (Equation (23)) and Equation (36), i.e., the approximations made on Equation (23) that correspond to the assumptions made in §4.2 to derive Equation (36).

The summation in Equation (23) can be decomposed into

$$\sum_{i=n}^{\infty} (i - n + 1) \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM} \left(\tau_m^{(j)}|i - j + 1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right)$$
$$= -(n-1) \sum_{i=n}^{\infty} \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM} \left(\tau_m^{(j)}|i - j + 1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right)$$
$$+ \sum_{i=n}^{\infty} i \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM} \left(\tau_m^{(j)}|i - j + 1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right). \tag{G1}$$

The first term on the r.h.s. of Equation (G1) is equivalent to $-(n-1) \prod_{i=1}^{n} P_i$ (Appendix E), and then, this term formally coincides with the correction term in Equation (36). Therefore, the second term corresponds to the kernel part (Equation (37)).

n-consecutive lower intervals must be included in only one upper interval. Under this condition, three constraints are imposed on the lower intervals, which appear on the l.h.s. of Equation (G1) as follows: (1) The number of lower intervals included in the upper interval must be larger than or equal to n. Then, the summation is taken in the range of $i \geq n$. (2) The way to choose the n-consecutive intervals from the i-lower intervals in an upper interval is only $(i - n + 1)$. If the first lower interval (or the leftmost one in the sequence of the consecutive lower intervals) is in either remaining $(n-1)$ ways, the sequence overflows from the upper interval. (3) The probability of the length of the j-th interval in the consecutive lower intervals depends on the way other (k-th, $1 \leq k <$

j) lower intervals appear, i.e., it is dependent on the remained time $\tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}$ and number of pieces of lower intervals $(i-j+1)$, $\rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right)$.

These constraints are relaxed in the derivation in §4.2. In the view in §4.2, the upper intervals of length τ_M are collected and the new time series is generated as shown in Figure 5. For this new time series, the only constraint imposed on the lower intervals is that they are included in the upper interval of length τ_M; each interval is assumed to occur independently. Therefore, the three constraints are changed in the following manner: (1) The new time series is generated by gathering all upper intervals of length τ_M, regardless of the number of lower intervals included in it. In addition, the restriction on the range of the summation ($i \geq n$) does not make much sense because the consecutive lower intervals are not assumed to be within only one upper interval, i.e., it is expanded to $i \geq 1$. (2) The number of ways to choose the n-consecutive intervals from i-lower intervals is unchanged; this exceeds the above mentioned upper limit $(i-n+1)$ although such cases are subtracted by the first term on the r.h.s. of Equation (G1), i.e., the correction term in Equation (36). (3) The constraints imposed on the condition in ρ_{mM} are removed; because the probability of the length of j-th interval is only not affected by other lower intervals, the temporal part of ρ_{mM} is replaced by τ_M (Equation (G2)). In addition, the constraint on the number of division can be eliminated by taking the average (Equation (G3)).

$$\rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right) \approx \rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M\right) \tag{G2}$$

$$\approx \frac{\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(j)}|i, \tau_M)}{\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M)} \tag{G3}$$

$$= p_{mM}(\tau_m^{(j)}|\tau_M).$$

Thus, ρ_{mM}'s are simply replaced by the conditional probability density functions.

In this way, the approximate view in §4.2 implies the following replacement in the exact inverse probability density function.

$$\begin{aligned}
&\sum_{i=n}^{\infty} i\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i, \tau_M)\prod_{j=2}^{n}\rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{n}\tau_m^{(k)}\right)\\
&\quad\approx \sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M)\prod_{j=1}^{n} p_{mM}(\tau_m^{(j)}|\tau_M)\\
&\quad= \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}}\prod_{j=1}^{n} p_{mM}(\tau_m^{(j)}|\tau_M).
\end{aligned}$$

Appendix H Distance between the inverse probability density function and the inter-event time distribution

In this Appendix, we derive the distance between the inverse probability density function (Equation (24)) and the inter-event time distribution ($p_M(\tau_M)$)

$$D(p_{Mm}||p_M) := \int_T^{\infty} |p_\theta(\tau_M, T) - p_M(\tau_M)|^2 \, d\tau_M, \tag{H1}$$

where

$$\begin{aligned}
p_\theta(\tau_M, T) &= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2}\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)e^{-\frac{\tau_M - T}{\langle\tau_M\rangle}}\left(A_{\Delta m}\frac{\tau_M - T}{\langle\tau_M\rangle} + 2\right),\\
p_M(\tau_M) &= \frac{1}{\langle\tau_M\rangle}e^{-\frac{\tau_M}{\langle\tau_M\rangle}}.
\end{aligned}$$

By substituting these functions in Equation (H1), the distance is derived as

$$D(p_{Mm}||p_M) = \frac{\langle\tau_M\rangle C_1^2}{4} + \frac{C_1 C_2(T)}{2} + \frac{C_2(T)^2}{2\langle\tau_M\rangle}, \tag{H2}$$

where

$$\begin{aligned} C_1 &= \frac{1}{\langle\tau_M\rangle}\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^2, \\ C_2(T) &= 2\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) - e^{-\frac{T}{\langle\tau_M\rangle}}. \end{aligned}$$

Appendix I On the cause of the separation of $\langle D'(p_{Mm}^{\mathrm{approx}}||p_M)\rangle$ and $\langle D''(p_{Mm}^{\mathrm{approx}}||p_M)\rangle$ at large T

In this appendix, we examine the cause of the separation between $\langle D'(p_{Mm}^{\mathrm{approx}}||p_M)\rangle$ and $\langle D''(p_{Mm}^{\mathrm{approx}}||p_M)\rangle$ at long-elapsed time T. Let us compare Figure 7 to Figures S6(a) and (c) for $\mathcal{N} = 10^5$. The separation is suppressed compared to that shown in Figure 7, which indicates that the fluctuations in the spline functions of $\boldsymbol{P}_1$ caused by a relatively small number of samples in the calculation of $\boldsymbol{P}_1$ are suppressed by increasing the sample data. This leads to the reduction of errors in the calculations (44), and to the improvement of the calculation of distance in Equation (46).

In addition, we tested numerical updating with $\mathcal{N} = 10^5$ by excluding some larger columns of the matrix $\boldsymbol{P}_1$, i.e., by using the following matrix $\boldsymbol{P}_1'$ with an integer l_c

$$\boldsymbol{P}_1' = \left[[P_{1,jk}]_{j=j_{\min}^{(k)},\cdots,j_{\max}^{(k)}}\right]_{k=k_{\min},\cdots,k_{\max}-l_c}. \tag{I1}$$

For this $\boldsymbol{P}_1'$, the interpolation and extrapolation procedures are conducted in the same way as in §5.1, and the numerical updating is executed.

Figures S6(b) and (d) show the results of the distance for such updating with (b) $\Delta\tau_M = 0.1$ and $l_c = 5$, and (d) $\Delta\tau_M = 0.025$ and $l_c = 20$. Compared to the results obtained using $\boldsymbol{P}_1$ in Figures S6(a) and (c), the separation is suppressed further. Combined with the results for the kernel part, these results suggest the following; the number of samples to calculate $\boldsymbol{P}_1$ is so small compared to that of the conditional probability (the number of sample is only one for an upper interval for $\boldsymbol{P}_1$ whereas all the lower intervals included in an upper interval are used as a sample to calculate the conditional probability), in particular for a large k, that its fluctuation becomes too large to compute the correction term precisely.

Acknowledgments

The idea of generalized probability density functions and the simplified relation (7) are by Y. Aizawa. We would like to thank Editage (www.editage.com) for English language editing. This work was supported by JST SPRING, Grant Number JPMJSP2110.

References

Aizawa, Y., Hasumi, T., & Tsugawa, S. (2013). Seismic statistics. *Nonlinear Phenomena in Complex Systems*, *16*(2), 116–130.

Aizawa, Y., & Tsugawa, S. (2014). Aftershock cascade of the 3.11 earthquake (2011) in fukushima-miyagi area. In D. Matrasulov & H. E. Stanley (Eds.), *Nonlinear phenomena in complex systems: From nano to macro scale* (pp. 21–33). Dordrecht: Springer Netherlands.

Bak, P., Christensen, K., Danon, L., & Scanlon, T. (2002). Unified scaling law for earthquakes. *Physical Review Letters*, *88*(17), 178501.

Bottiglieri, M., de Arcangelis, L., Godano, C., & Lippiello, E. (2010). Multiple-time scaling and universal behavior of the earthquake interevent time distribution. *Physical Review Letters*, *104*(15), 158501.

Corral, A. (2004). Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. *Physical Review Letters*, *92*(10), 108501.

Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in california. *Bulletin of the Seismological Society of America*, *34*(4), 185–188.

Helmstetter, A., & Sornette, D. (2002). Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research: Solid Earth*, *107*(B10), ESE 10-1-ESE 10-21. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001JB001580` doi: https://doi.org/10.1029/2001JB001580

Helmstetter, A., & Sornette, D. (2003). Importance of direct and indirect triggered seismicity in the etas model of seismicity. *Geophysical Research Letters*, *30*(11). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003GL017670` doi: https://doi.org/10.1029/2003GL017670

Lippiello, E., Corral, Á., Bottiglieri, M., Godano, C., & Arcangelis, L. (2012). Scaling behavior of the earthquake intertime distribution: Influence of large shockss and time scales in the omori law. *Physical Review E*, *86*(6), 066119.

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

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*(2), 379–402.

Omori, F. (1894). On aftershocks of earthquakes. *The journal of the College of Science, Imperial University of Tokyo, Japan*, *7*, 111–200.

Saichev, A., & Sornette, D. (2006). "universal" distribution of interearthquake times explained. *Phys. Rev. Lett.*, *97*, 078501. Retrieved from `https://link.aps.org/doi/10.1103/PhysRevLett.97.078501` doi: 10.1103/PhysRevLett.97.078501

Scholz, C. H. (2002). *The mechanics of earthquakes and faulting.* Cambridge university press.

Tanaka, H., & Aizawa, Y. (2017). Detailed analysis of the interoccurrence time statistics in seismic activity. *Journal of the Physical Society of Japan*, *86*(2), 024004.

Tanaka, H., & Umeno, K. (2021). Bayesian updating for time-intervals of different magnitude thresholds in marked point process and its application to time-series of ETAS model. *ESS Open Archive*. Retrieved from `https://doi.org/10.1002%2Fessoar.10509635.1` doi: 10.1002/essoar.10509635.1

Touati, S., Naylor, M., & Main, I. G. (2009). Origin and nonuniversality of the earthquake interevent time distribution. *Physical Review Letters*, *102*(16), 168501.

Utsu, T. (1961). A statistical study on the occurrence of aftershocks. *Geophysical Magazine*, *30*, 521–605.

Webb, D. J. (1974). The statistics of relative abundance and diversity. *Journal of Theoretical Biology*, *43*(2), 277–291.

Bayesian updating on time intervals at different magnitude thresholds in a marked point process and its application to synthetic seismic activity

Hiroki Tanaka [1] and Ken Umeno [1]

[1]Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto-shi, JAPAN

Key Points:

- Bayesian updating for time intervals at different magnitude thresholds in marked point process is discussed for large event forecasting
- Inverse probability density and its approximation functions are derived analytically for a time series with no correlation between events
- Forecasting of large events by Bayesian method is examined for the ETAS model time series and its effectiveness is discussed statistically

Corresponding author: Hiroki Tanaka, tanaka.hiroki.43s@st.kyoto-u.ac.jp

Abstract

We present a Bayesian updating method on the inter-event times at different magnitude thresholds in a marked point process, toward the probabilistic forecasting of an upcoming large event using temporal information on smaller events. Bayes' theorem in a marked point process that yields the one-to-one relationship between intervals at lower and upper magnitude thresholds is presented. This theorem is extended to Bayesian updating for an uncorrelated marked point process that yields the relationship between multiple consecutive lower intervals and one upper interval. The inverse probability density function and its approximation function are derived. For the former, the condition for having a peak is shown. The latter is easier to apply to the time series of the ETAS model, and it consists of the kernel part, which includes the product of the conditional probabilities, and the correction term. The maximum point of the kernel part is shown to be not significantly affected by the correction term when applying the Bayesian updating to the ETAS model time series numerically. The occurrence time of the upcoming large event is estimated using this maximum point, and its accuracy is evaluated considering the relative error with the actual occurrence time. Moreover, forecasting is evaluated to be effective by the continuity of the updates with the accuracy within an acceptable range prior to the upcoming large event. Under these conditions, the statistical analysis indicates that forecasting is relatively effective immediately or long after the last major event in which stationarity is dominant in the time series.

Plain Language Summary

In order to forecast future large earthquakes, it is important to use as much information as possible on the seismic activity at hand. The number of small earthquakes is much larger than that of large earthquakes, and we propose a method to use this information to forecast probabilistically the timing of future large earthquakes. Theoretical analysis is performed on a simple time series. The theoretical results are applied to a seismic activity model, and it is shown that this method is relatively effective in forecasting the timing of future large events when stationary activity is dominant; in this model, either immediately after a large event or after sufficient time has passed since the last large event. Therefore, this method can be applied to reduce secondary disasters after a major earthquake and to evaluate the risk of earthquake occurrence over a long period of time.

1 Introduction

The probabilistic forecasting of the timing of future major earthquakes is important for seismic risk assessment. Therefore, it is necessary to effectively use the information on the temporal properties of seismic activity represented by a marked point process with the magnitude as the mark as indicated in Figure 1. A basic approach involves using the hazard rate based on the inter-event time distribution of earthquakes (Scholz, 2002). The inter-event time distribution is defined as a probability density function of the length of the interval between adjacent points in the point process determined by setting a magnitude threshold for the marked point process. For the magnitude threshold M (m), we denote the inter-event times using a variable τ_M (τ_m), and the inter-event time distribution it follows by $p_M(\tau_M)$ ($p_m(\tau_m)$).

The inter-event time distribution of earthquakes has been studied for not only risk assessment but also to understand the statistical nature of seismicity. For example, it has been studied with the aim to unify it with other established laws (Bak et al., 2002; Aizawa et al., 2013; Aizawa & Tsugawa, 2014) such as the GR law (Gutenberg & Richter, 1944) and the Omori–Utsu law (Omori, 1894; Utsu, 1961); further, its scaling universality (Corral, 2004) has been discussed using the ETAS model (Saichev & Sornette, 2006; Touati et al., 2009; Bottiglieri et al., 2010; Lippiello et al., 2012).

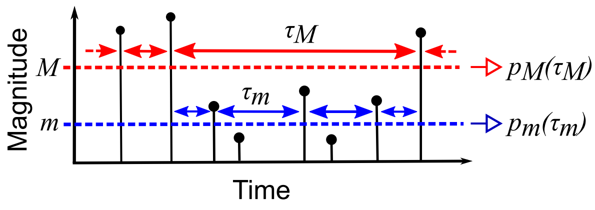

Figure 1. Schematic of a marked point process with the magnitude as a mark.

The ETAS model is a stochastic model that combines the GR and Omori–Utsu laws (Ogata, 1988, 1998); this model can generate a time series like that of the seismic activity. The ETAS model generates an inhomogeneous Poisson process with a history-dependent occurrence rate. Let t_j and M_j ($j \in \mathbb{N}$) represent the occurrence time and magnitude of the j-th event before time t; then, the occurrence rate ($\lambda(t)$) at time t is given by

$$\lambda(t) = \lambda_0 + \sum_{j:t_j<t} \frac{K10^{\alpha(M_j-M_0)}}{(t-t_j+c)^{\theta+1}}. \tag{1}$$

The magnitude is generated randomly and independently obeying the GR law, $P(M) \propto 10^{-bM}$. Here, M_0 represents the minimum magnitude and $(\lambda_0, K, \alpha, c, \theta, b)$ represent the parameters that characterize the activity. In particular, λ_0 represents the constant rate for background seismicity. The combination of the remaining parameters yields the branching ratio $n_{\mathrm{br}} = \dfrac{K}{\theta c^\theta}\dfrac{b}{b-\alpha}$ (when $\theta > 0$) (Helmstetter & Sornette, 2002) that determines the stationarity of the time series as well as the average number of aftershocks generated by a mainshock (Helmstetter & Sornette, 2003).

The ETAS model provides a standard seismicity for detecting anomalous activity (Ogata, 1988). This model has been extended to a spatio-temporal version (Ogata, 1998), and its application to the evaluation of seismic risk has been studied actively. The conditional intensity function provides the risk of an event at a given time, space, and magnitude based on the history of seismic activity, which includes small earthquakes.

In the aforementioned probabilistic evaluation using the inter-event time distribution, temporal information on events smaller than the magnitude threshold set on the marked point process is not utilized. Therefore, in this paper, we propose another approach to probabilistically forecast major earthquakes based on the inter-event time distribution while considering the temporal information on smaller events. This is achieved by utilizing a conditional probability that yields the statistical relationship between the inter-event times at different two magnitude thresholds (Tanaka & Aizawa, 2017).

For two magnitude thresholds m and $M (= m + \Delta m, \Delta m > 0)$ set in the time series, the conditional probability ($p_{mM}(\tau_m|\tau_M)$, hereafter referred to as the conditional probability density function) is defined as the probability density function of the length of a lower interval (τ_m) under the condition that it is included in the upper interval of length τ_M (Figure 1). Inter-event time distributions at magnitude thresholds m and M are connected by an integral equation with this conditional probability density function in its kernel (Tanaka & Aizawa, 2017). This integral equation is given as

$$N_m p_m(\tau_m) = N_M \int_{\tau_m}^{\infty} \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M) p_M(\tau_M) d\tau_M, \tag{2}$$

where N_m and N_M represent the total number of intervals at magnitude thresholds m and M, respectively. Further, $\langle\langle \tau_m \rangle\rangle_{\tau_M}$ represents the average of the conditional probability density function, $\langle\langle \tau_m \rangle\rangle_{\tau_M} := \int_0^\infty \tau_m p_{mM}(\tau_m|\tau_M) d\tau_m$.

Thus, the conditional probability density function yields the statistical relation between inter-event times at different magnitude thresholds. This suggests that the information on the lower intervals can be utilized for estimating the length of the upper interval through the conditional probability density function. Given this context, we consider the Bayes' theorem and the Bayesian updating on the intervals at different magnitude thresholds in this paper; further, we report the results of the numerical analysis of the properties of the inverse probability density function (Tanaka & Umeno, 2021).

In §2, we derive the Bayes' theorem for intervals at different magnitude thresholds in the marked point process. In §3, the inverse probability density function is derived for the uncorrelated time series that corresponds to the background seismicity of the ETAS model. In §4, the Bayesian updating method is considered for the uncorrelated time series, and the inverse probability density function and its approximation function are derived. These functions are calculated numerically and compared in §5. In §6, Bayesian updating is applied to the time series of the ETAS model. The approximation function is examined numerically, and the property of the maximum point of its kernel part is analyzed statistically considering the effectiveness for forecasting. Finally, §7 presents additional discussions and conclusions.

2 Bayes' theorem for inter-event times at different magnitude thresholds

We consider the Bayes' theorem between the inter-event times at different magnitude thresholds (m and M) in a marked point process, and we derive the general relationship between the conditional probability density function $p_{mM}(\tau_m|\tau_M)$ and the inverse probability density function $p_{Mm}(\tau_M|\tau_m)$ (Tanaka & Umeno, 2021). Here $p_{Mm}(\tau_M|\tau_m)$ represents the probability density function of the upper interval under the condition that it includes a lower interval of length τ_m.

Let the total number of the pairs of the upper interval of length within $[\tau_M, \tau_M + d\tau_M)$ and the lower interval of length within $[\tau_m, \tau_m + d\tau_m)$ by $N_{mM}(\tau_M, \tau_m)$ (Figure 2). Hereafter, we express this $N_{mM}(\tau_M, \tau_m)$ as the number of the pairs of the intervals such that the length of the upper interval *is* τ_M and the length of the lower interval *is* τ_m, for simplicity, and other numbers of the intervals are expressed in the same way. $N_{mM}(\tau_M, \tau_m)$ can be represented in two ways:

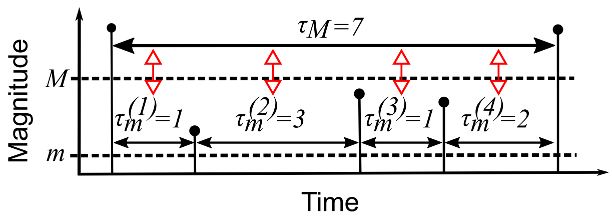

Figure 2. Schematic of the approach to count the number of pairs of upper and lower intervals whose lengths are τ_M and τ_m, respectively. Four pairs are shown in the figure, and $N_{mM}(7, 1) = 2$, $N_{mM}(7, 2) = 1$, and $N_{mM}(7, 3) = 1$.

1) Derive $N_{mM}(\tau_M, \tau_m)$ by counting the cumulative total number of the upper intervals of length τ_M that include the lower interval of length τ_m (Figure 3(a)). Among the N_m lower intervals in the time series, there are $N_m p_m(\tau_m) d\tau_m$ intervals of length τ_m. There exists only one upper interval that includes each of such lower intervals. The probability that the length of that upper interval is τ_M is given by $p_{Mm}(\tau_M|\tau_m) d\tau_M$. There-

fore

$$N_{mM}(\tau_M, \tau_m) = N_m p_m(\tau_m) p_{Mm}(\tau_M|\tau_m) d\tau_m d\tau_M. \tag{3}$$

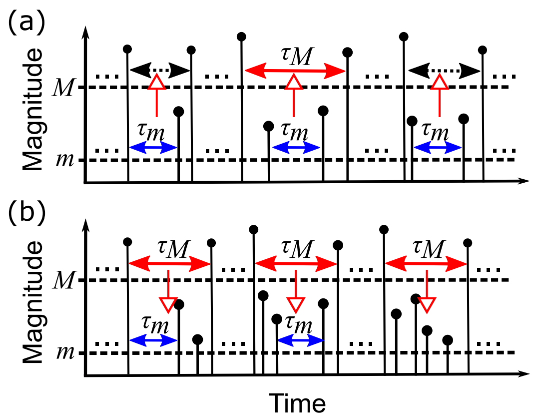

Figure 3. Schematic of the two approaches for calculating $N_{mM}(\tau_M, \tau_m)$. (a) The first approach involves counting the cumulative total number of the upper intervals of length τ_M that include the lower interval of length τ_m. (b) The second approach involves counting the number of the lower intervals of length τ_m included in the upper interval of length τ_M.

2) Derive $N_{mM}(\tau_M, \tau_m)$ by counting the total number of the lower intervals of length τ_m included in the upper interval of length τ_M (Figure 3(b)). The number of the upper intervals of length τ_M in the time series is $N_M p_M(\tau_M) d\tau_M$. Therefore, the number of the lower intervals included in these upper intervals is

$$N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} d\tau_M.$$

Among them, the proportion of the lower intervals whose length is τ_m is $p_{mM}(\tau_m|\tau_M) d\tau_m$. Therefore

$$N_{mM}(\tau_M, \tau_m) = N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M) d\tau_m d\tau_M. \tag{4}$$

From Equations (3) and (4)

$$N_m p_m(\tau_m) p_{Mm}(\tau_M|\tau_m) = N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M). \tag{5}$$

By using the average intervals at each magnitude threshold

$$\langle\tau_m\rangle := \int_0^\infty \tau_m p_m(\tau_m) d\tau_m,$$
$$\langle\tau_M\rangle := \int_0^\infty \tau_M p_M(\tau_M) d\tau_M,$$

and $N_M/N_m = \langle\tau_m\rangle/\langle\tau_M\rangle$, Equation (5) is rewritten as

$$p_{Mm}(\tau_M|\tau_m) = \left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} \right) \frac{p_{mM}(\tau_m|\tau_M) p_M(\tau_M)}{p_m(\tau_m)}. \tag{6}$$

Equation (6) can be considered as the Bayes' theorem for a marked point process. The parenthesized part is from the difference in the number of intervals for each magnitude threshold ($\langle\tau_m\rangle/\langle\tau_M\rangle$) and the inclusion relationship between the upper and lower intervals $(\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M})$, i.e., a lower interval is always included in only one upper interval, whereas an upper interval includes $\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M}$ lower intervals on average. This part disappears by using generalized probability density functions

$$z_m(\tau_m) := \frac{\tau_m}{\langle\tau_m\rangle} p_m(\tau_m),$$
$$z_M(\tau_M) := \frac{\tau_M}{\langle\tau_M\rangle} p_M(\tau_M),$$
$$z_{mM}(\tau_m|\tau_M) := \frac{\tau_m}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M).$$

These functions satisfy the normalization condition of the probability density function. Equations (2) and (6) are simplified as

$$z_m(\tau_m) = \int_0^\infty z_{mM}(\tau_m|\tau_M) z_M(\tau_M) d\tau_M, \tag{7}$$
$$p_{Mm}(\tau_M|\tau_m) = \frac{z_{mM}(\tau_m|\tau_M) z_M(\tau_M)}{z_m(\tau_m)}. \tag{8}$$

These equations indicate that $p_{Mm}(\tau_M|\tau_m)$ satisfies the normalization condition.

3 Bayes' theorem for uncorrelated time series

In this section, we derive $p_{Mm}(\tau_M|\tau_m)$ for an uncorrelated time series generated by the ETAS model with $\lambda(t) \equiv \lambda_0$ (Tanaka & Umeno, 2021). In this case, the magnitudes and inter-event times obey the following probability density functions independently.

$$P(m) \propto 10^{-bm}, \tag{9}$$
$$p_m(\tau_m) = \frac{1}{\langle\tau_m\rangle} e^{-\frac{\tau_m}{\langle\tau_m\rangle}}. \tag{10}$$

First, we derive $p_{mM}(\tau_m|\tau_M)$, which can be expressed generally as

$$p_{mM}(\tau_m|\tau_M) = \frac{\sum_{i=1}^{\infty} i\ \rho_{mM}(\tau_m|i,\tau_M)\Psi_{mM}(i|\tau_M)}{\sum_{i=1}^{\infty} i\ \Psi_{mM}(i|\tau_M)}, \tag{11}$$

where $i(\in \mathbb{Z})$ represents the number of lower intervals included in the upper interval of length τ_M; $\Psi_{mM}(i|\tau_M)$ represents the probability mass function of such i under the condition that the length of the upper interval is τ_M; and $\rho_{mM}(\tau_m|i,\tau_M)$ represents the probability density function of the length of a lower interval given that the length of the upper interval is τ_M and the number of the lower intervals in it is i. We can calculate the conditional probability density function and other related amounts when we know these functions.

In the case of the uncorrelated time series, these functions can be obtained as follows. For the selected stationary Poisson process, the average number of events included in the upper interval of length τ_M is $(\tau_M/\langle\tau_m\rangle - \tau_M/\langle\tau_M\rangle)$, because no event greater than M occurs in the interval considered, and therefore, $\tau_M/\langle\tau_M\rangle$-events larger than M occurring in the interval of length τ_M on average must be excluded from the average number $\tau_M/\langle\tau_m\rangle$ of events occurring in the interval of length τ_M. Then, the average occurrence rate in the upper interval of length τ_M is $(1/\langle\tau_m\rangle - 1/\langle\tau_M\rangle)$. The number of events with $m <$ magnitude $\leq M$ in τ_M is one less than that of the lower intervals, and therefore, the probability of including i lower intervals is equal to the probability of including $(i-1)$ events with an average occurrence rate $(1/\langle\tau_m\rangle - 1/\langle\tau_M\rangle)$. Therefore

$$\Psi_{mM}(i|\tau_M) = \frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right)^{i-1}}{(i-1)!} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}, \tag{12}$$

where

$$A_{\Delta m} := \frac{\langle \tau_M \rangle}{\langle \tau_m \rangle} - 1$$
$$= 10^{b\Delta m} - 1.$$

The second transformation in the above equation does not strictly hold for a time series with a finite number of events because the number of the events is different from that of the intervals by 1. However, we consider that the statistical properties are for infinite samples, and in a time series containing an infinite number of events, the two are equivalent and the equality holds.

The other function $\rho_{mM}(\tau_m|i,\tau_M)$ is obtained as follows. For $i = 1$

$$\rho_{mM}(\tau_m|1,\tau_M) = \delta(\tau_M - \tau_m), \tag{13}$$

where $\delta(x)$ represents the Dirac's delta function. For $i \geq 2$ (Webb, 1974)

$$\rho_{mM}(\tau_m|i,\tau_M) = \frac{(i-1)}{\tau_M}\left(1 - \frac{\tau_m}{\tau_M}\right)^{i-2} \theta(\tau_M - \tau_m), \tag{14}$$

where $\theta(x)$ represents the unit step function that returns 1 for $x > 0$ and 0 for $x \leq 0$.

From Equations (12)–(14), $p_{mM}(\tau_m|\tau_M)$ is derived as (Appendix A)

$$p_{mM}(\tau_m|\tau_M) = \frac{e^{-A_{\Delta m}\frac{\tau_M}{\langle \tau_M \rangle}}\delta(\tau_M - \tau_m) + \frac{A_{\Delta m}}{\langle \tau_M \rangle} e^{-A_{\Delta m}\frac{\tau_m}{\langle \tau_M \rangle}}\left\{A_{\Delta m}\frac{\tau_M - \tau_m}{\langle \tau_M \rangle} + 2\right\}\theta(\tau_M - \tau_m)}{\left(A_{\Delta m}\frac{\tau_M}{\langle \tau_M \rangle} + 1\right)}. \tag{15}$$

This conditional probability composed of Equations (12)–(14) certainly has exponential distributions as the solution of Equation (2) (Appendix A).

Second, we derive $p_{Mm}(\tau_M|\tau_m)$. From Equations (10) and (15), $p_{Mm}(\tau_M|\tau_m)$ is obtained as (Appendix A)

$$p_{Mm}(\tau_M|\tau_m) = \frac{e^{-\frac{\tau_M - \tau_m}{\langle \tau_m \rangle}}\delta(\tau_M - \tau_m) + \frac{A_{\Delta m}}{\langle \tau_M \rangle} e^{-\frac{\tau_M - \tau_m}{\langle \tau_M \rangle}}\left\{A_{\Delta m}\frac{\tau_M - \tau_m}{\langle \tau_M \rangle} + 2\right\}\theta(\tau_M - \tau_m)}{\left(A_{\Delta m} + 1\right)^2}. \tag{16}$$

We emphasize that $p_{Mm}(\tau_M|\tau_m)$ has a peak at

$$\tau_M^{\max} = \tau_m + \langle \tau_M \rangle \left(1 - \frac{2}{A_{\Delta m}}\right), \tag{17}$$

when the next condition is satisfied (Appendix A).

$$\Delta m > \frac{\log_{10} 3}{b}. \tag{18}$$

4 Bayesian updating for uncorrelated time series

The Bayes' theorem shows a one-to-one relationship between an upper and a lower interval. In this section, we extend it to the relationship between an upper interval and multiple consecutive lower intervals by considering Bayesian updating for the uncorrelated time series (Tanaka & Umeno, 2021). We derive the inverse probability density function $p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$, as well as its approximation function, for the upper interval under the condition that it includes the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$.

4.1 Inverse probability density function

As in §2, we derive the inverse probability density function by expressing the total number of combinations of the upper interval of length τ_M and the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ included in it denoted by $N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ in two ways.

First, we derive $N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ by counting the cumulative total number of the upper intervals of length τ_M that include the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ (Figure 4(a)). We begin with the case $n = 2$. The intervals in the uncorrelated time series emerge independently, and therefore, the total number of the two consecutive lower intervals of lengths $\tau_m^{(1)}$ and $\tau_m^{(2)}$ is

$$N_m p_m(\tau_m^{(1)}) p_m(\tau_m^{(2)}) d\tau_m^2.$$

Among them, some pairs do not belong to the same upper interval (the case of (3) in Figure 4(a)). In that case, the magnitude of the event sandwiched between the two lower intervals is larger than M. In the uncorrelated time series, the proportion that the consecutive lower intervals belong to the same upper interval equals to the probability that the magnitude of the event sandwiched between the two lower intervals is smaller than M. It is given by the GR law as

$$\begin{aligned} 1 - \frac{P(M)}{P(m)} &= 1 - 10^{-b\Delta m} \\ &= 1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}. \end{aligned}$$

Therefore

$$N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)}) = N_m \left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right) p_m(\tau_m^{(1)}) p_m(\tau_m^{(2)}) p_{Mm}(\tau_M | \tau_m^{(1)}, \tau_m^{(2)}) d\tau_m^2 d\tau_M. \quad (19)$$

Equation (19) is generalized for $n(\geq 2)$ consecutive lower intervals.

$$\begin{aligned} &N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\ &= N_m \left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1} \left(\prod_{i=1}^{n} p_m(\tau_m^{(i)})\right) p_{Mm}(\tau_M | \tau_m^{(1)}, \cdots, \tau_m^{(n)}) d\tau_m^n d\tau_M. \end{aligned} \quad (20)$$

Second, we derive $N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ by counting the total number of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ included in the upper interval of length τ_M (Figure 4(b)). To this end, we start with the case $n = 2$ again (Figure 4(b)). When the upper interval of length τ_M includes $i(\geq 2)$ lower intervals, the first interval of the two consecutive lower intervals is selected from $(i-1)$ intervals except for the rightmost one. The probability that this first interval has length $\tau_m^{(1)}$ is $\rho_{mM}(\tau_m^{(1)} | i, \tau_M) d\tau_m$. The second lower interval is fixed at adjacent to the first one. This second interval is one of the $(i-1)$ intervals that divide the remaining length $\tau_M - \tau_m^{(1)}$, and therefore, the probability that the second interval has length $\tau_m^{(2)}$ is $\rho_{mM}(\tau_m^{(2)} | i-1, \tau_M - \tau_m^{(1)}) d\tau_m$. Thus, considering all $i(\geq 2)$

$$\begin{aligned} &N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)}) \\ &= N_M p_M(\tau_M) d\tau_M \sum_{i=2}^{\infty} (i-1) \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)} | i, \tau_M) \rho_{mM}(\tau_m^{(2)} | i-1, \tau_M - \tau_m^{(1)}) d\tau_m^2. \end{aligned} \quad (21)$$

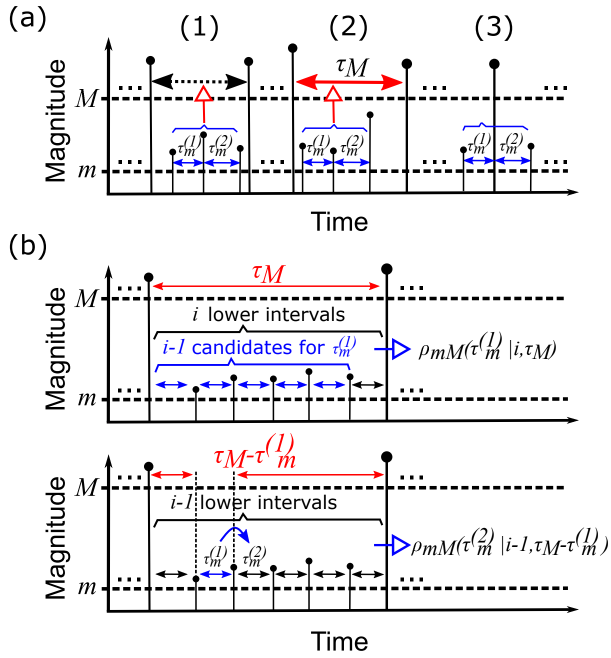

Figure 4. Schematic of the two approaches to calculate $N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)})$. (a) The first approach involves counting the cumulative total number of the upper intervals of length τ_M that include the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \tau_m^{(2)}\}$. (b) The second approach involves counting the number of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \tau_m^{(2)}\}$ included in the upper interval of length τ_M.

Equation (21) is generalized for the case $n(\geq 2)$ lower intervals as

$$N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) = N_M p_M(\tau_M) d\tau_M \\ \times \sum_{i=n}^{\infty}(i-n+1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M)\prod_{j=2}^{n}\rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1}\tau_m^{(k)}\right) d\tau_m^n. \tag{22}$$

From Equations (20) and (22), $p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$ is derived as

$$p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\frac{1}{\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}}\frac{p_M(\tau_M)}{\prod_{i=1}^{n}p_m(\tau_m^{(i)})} \\ \times \sum_{i=n}^{\infty}(i-n+1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M)\prod_{j=2}^{n}\rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1}\tau_m^{(k)}\right). \tag{23}$$

Furthermore, the explicit form of the inverse probability density function is derived by substituting Equations (10) and (12)–(14) into Equation (23) as (Appendix B)

$$p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^2\left[e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_m\rangle}}\delta\left(\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}\right)\right. \\ \left. + \frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}}\left\{\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left(\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}\right) + 2\right\}\theta\left(\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}\right)\right]. \tag{24}$$

Equation (24) includes the case $n = 1$ (Equation (16)). In addition, Equation (24) is identical to Equation (16) when τ_m is replaced with $\mathcal{T} := \sum_{i=1}^{n} \tau_m^{(i)}$; this implies that the occurrence pattern of small events does not affect that of upper intervals. This seems natural for the uncorrelated time series.

The same property as Equations (17) and (18) holds for $p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$; it has a peak at

$$\tau_M^{\max} = \mathcal{T} + \langle\tau_M\rangle \frac{\frac{\langle\tau_M\rangle}{\langle\tau_m\rangle} - 3}{\frac{\langle\tau_M\rangle}{\langle\tau_m\rangle} - 1} \; (> \mathcal{T}),$$

under the condition

$$\Delta m > \frac{\log_{10} 3}{b}. \tag{25}$$

In the above mentioned Bayesian updating, the position of the consecutive lower intervals in an upper interval is not restricted. However, update can be started only from the lower interval immediately after the event with the magnitude above M. In such a method, the inverse probability density function is different from Equation (24) (Appendix C). At a glance, this updating method seems suitable under the situation wherein the information on the lower intervals observed one after another is imported sequentially; however, seismic catalogs are known to be incomplete immediately after a large earthquake. In that case, the lower intervals should be considered not from the leftmost one but from somewhere else. Therefore, in the present paper, we limit ourselves to examine the property of the inverse probability density function of the unrestricted updating method that is more appropriate for application to earthquake catalogs.

4.2 Approximation function of inverse probability density function

Equation (23) indicates that new information on the lower intervals cannot be added by the product of the conditional probabilities as is usual in Bayesian updating. In this sub-section, we derive its approximation function with a convenient form applicable to the time series with correlations between events.

To this end, we use the approximate derivation of $N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ described below instead of the second approach for deriving Equation (22). In the following, the upper and the lower consecutive intervals are assumed to satisfy

$$\tau_M \geq \sum_{i=1}^{n} \tau_m^{(i)}. \tag{26}$$

First, consider the case $n = 2$. There are $N_M p_M(\tau_M) d\tau_M$ upper intervals of length τ_M in the time series. These upper intervals are as shown in Figure 5(a), and we use them to generate a new time series by connecting them in the order of appearance as in Figure 5(b). Let the number of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \tau_m^{(2)}\}$ in this new time series be denoted by $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$. The total number of the lower intervals in this new time series is given as

$$N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} d\tau_M.$$

Therefore, based on the assumption that $\tau_m^{(1)}$ and $\tau_m^{(2)}$ emerge independently, $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$ is approximately calculated as

$$N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M) \approx N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} p_{mM}(\tau_m^{(1)}|\tau_M) p_{mM}(\tau_m^{(2)}|\tau_M) d\tau_m^2 d\tau_M. \tag{27}$$

$N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$ is not equivalent to $N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)})$ because $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$ includes cases where the two consecutive lower intervals do not belong to the same up-

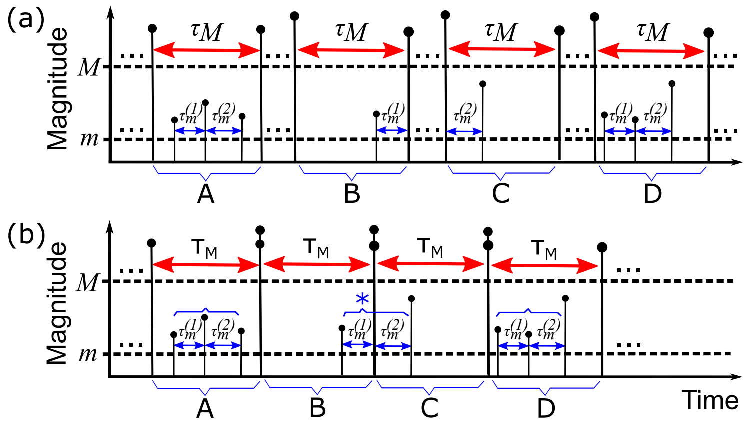

Figure 5. Schematic of another approach to count the total number of consecutive lower intervals of lengths $\tau_m^{(1)}$ and $\tau_m^{(2)}$ included in the upper interval of length τ_M. (a) First, pick up all upper intervals of length τ_M from the time series. (b) Second, generate new time series by connecting these upper intervals in the order of appearance. Third, $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$ is calculated by counting the total number of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \tau_m^{(2)}\}$ in this new time series. In this counting process, an approximate calculation using the product of the conditional probability is conducted. Finally, $N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)})$ is obtained by excluding such pairs where the two consecutive lower intervals are not included in the same upper interval (the cases indicated with $*$) from $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$.

per interval (the case indicated by $*$ in Figure 5(b)). Therefore, it is necessary to count such cases in the time series, and subtract them from $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$.

These cases to exclude occur when an upper interval of length τ_M whose rightmost lower interval has length $\tau_m^{(1)}$ is adjacent to the left of another upper interval whose leftmost lower interval has length $\tau_m^{(2)}$. The probability density that the length of the rightmost or leftmost lower interval of the upper interval of length τ_M is τ_m is, because the position of the rightmost or leftmost interval is confirmed among the i-lower intervals, calculated as

$$\begin{aligned} P^{\mathrm{R}}(\tau_m|\tau_M) &= P^{\mathrm{L}}(\tau_m|\tau_M) \\ &= \sum_{i=1}^{\infty} \Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m|i,\tau_M). \end{aligned} \tag{28}$$

Here, the probability density for the rightmost lower interval is denoted by $P^{\mathrm{R}}(\tau_m|\tau_M)$, and the leftmost by $P^{\mathrm{L}}(\tau_m|\tau_M)$. Equation (28) can be explicitly written using Equations (12)–(14) as (Appendix D)

$$\begin{aligned} P^{\mathrm{R}}(\tau_m|\tau_M) &= P^{\mathrm{L}}(\tau_m|\tau_M) \\ &= e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m) + \frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\theta(\tau_M-\tau_m). \end{aligned} \tag{29}$$

By using $P^{\mathrm{L}}(\tau_m|\tau_M)$ and $P^{\mathrm{R}}(\tau_m|\tau_M)$, the number of cases to exclude can be expressed for a sufficiently large N_M (because $N_M p_M(\tau_M)d\tau_M$ in Equation (30) is precisely $N_M p_M(\tau_M)d\tau_M - 1$) as

$$N_M p_M(\tau_M) P^{\mathrm{R}}(\tau_m^{(1)}|\tau_M) P^{\mathrm{L}}(\tau_m^{(2)}|\tau_M) d\tau_m^2 d\tau_M. \tag{30}$$

Therefore, $N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)})$ is approximately derived as

$$N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)}) \approx N_M p_M(\tau_M) \times \left(\frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m^{(1)}|\tau_M) p_{mM}(\tau_m^{(2)}|\tau_M) - P^{\mathrm{R}}(\tau_m^{(1)}|\tau_M) P^{\mathrm{L}}(\tau_m^{(2)}|\tau_M) \right) d\tau_m^2 d\tau_M. \quad (31)$$

Next, we consider the case $n(\geq 3)$. Equation (27) is generalized as

$$N'_{mM}(\tau_m^{(1)}, \cdots, \tau_m^{(n)}|\tau_M) \approx N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \left(\prod_{i=1}^{n} p_{mM}(\tau_m^{(i)}|\tau_M) \right) d\tau_m^n d\tau_M. \quad (32)$$

From this $N'_{mM}(\tau_m^{(1)}, \cdots, \tau_m^{(n)}|\tau_M)$, the cases wherein the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ are not included in the same upper interval need to be excluded. Considering the condition of Equation (26), a sequence of consecutive lower intervals is divided by only one boundary event with a magnitude above M (Figure 6). Let the probability that the rightmost or leftmost lower intervals of the upper interval of length τ_M is $\{\tau_m^{(1)}, \cdots, \tau_m^{(l)}\}$ ($l \geq 2$) be $P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M)$. Then, as the position of the rightmost or leftmost lower intervals is confirmed among the $i(\geq l)$ lower intervals, $P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M)$ is

$$P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M) = \sum_{i=l}^{\infty} \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{l} \rho_{mM} \left(\tau_m^{(j)} | i - j + 1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)} \right). \quad (33)$$

By substituting Equations (12)–(14) into Equation (33) (Appendix E)

$$P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M) = \prod_{i=1}^{l} P_i(\tau_m^{(i)}|\tau_M), \text{ where } P_i(\tau_m^{(i)}|\tau_M) := \left(\frac{A_{\Delta m}}{\langle \tau_M \rangle} \right) e^{-A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle}}. \quad (34)$$

There are $(n - 1)$ possible choices for the boundary position of the consecutive lower intervals (Figure 6(a)), each with an equal probability $\prod_{i=1}^{n} P_i$. The number of consecutive upper intervals in the new time series is almost $N_M p_M(\tau_M) d\tau_M$, and therefore, the number of cases to be excluded is

$$N_M p_M(\tau_M)(n - 1) \left(\prod_{i=1}^{n} P_i(\tau_m^{(i)}|\tau_M) \right) d\tau_m^n d\tau_M.$$

Then

$$N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) \approx N_M p_M(\tau_M) \left\{ \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \prod_{i=1}^{n} p_{mM}(\tau_m^{(i)}|\tau_M) - (n - 1) \prod_{i=1}^{n} P_i(\tau_m^{(i)}|\tau_M) \right\} d\tau_m^n d\tau_M. \quad (35)$$

Therefore, from Equations (20) and (35), the approximation function ($p_{Mm}^{\mathrm{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$) of the inverse probability density function is derived as

$$p_{Mm}^{\mathrm{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle} \frac{1}{\left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1}} \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \left(\prod_{i=1}^{n} \frac{p_{mM}(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})} \right) p_M(\tau_M) - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle} \frac{(n - 1)}{\left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1}} \left(\prod_{i=1}^{n} \frac{P_i(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})} \right) p_M(\tau_M). \quad (36)$$

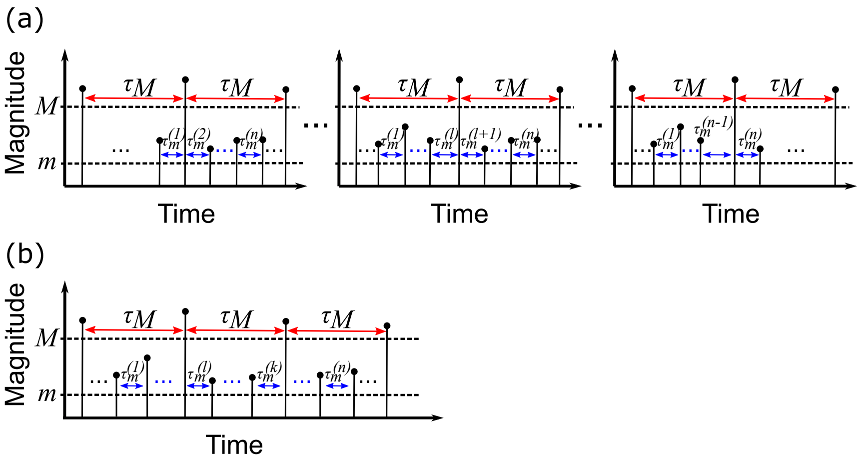

Figure 6. Schematic of the patterns of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ excluded from $N'_{mM}(\tau_m^{(1)}, \cdots, \tau_m^{(n)}|\tau_M)$. (a) There are $(n - 1)$ ways to divide the sequence of lower intervals by the event with a magnitude greater than M at the boundary of the upper intervals of length τ_M. (b) The sequence can not be divided by more than one boundary according to condition (26).

Equation (36) is composed of two parts: the first term of the r.h.s. involves the product of the conditional probability density functions, and we refer to this part as the kernel part of the approximation function ($p_{Mm}^{\text{kernel}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$) hereafter.

$$p_{Mm}^{\text{kernel}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{1}{\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}} \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} \left(\prod_{i=1}^{n} \frac{p_{mM}(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})}\right) p_M(\tau_M). \tag{37}$$

The second term of the r.h.s. is referred to as the correction term, and we denote the part other than $(n-1)$ by $p_{Mm}^{\text{correct}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$ as

$$\text{correction term} = (n-1)p_{Mm}^{\text{correct}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}), \tag{38}$$

$$\text{where } p_{Mm}^{\text{correct}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{1}{\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}} \left(\prod_{i=1}^{n} \frac{P_i(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})}\right) p_M(\tau_M).$$

Equation (36) can be explicitly written as (Appendix F)

$$p_{Mm}^{\text{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} \left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) \left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle} + 1\right) e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}}$$
$$\times \prod_{i=1}^{n} \left\{1 - \left(\frac{\tau_m^{(i)} - \frac{\langle\tau_M\rangle}{A_{\Delta m}}}{\tau_M + \frac{\langle\tau_M\rangle}{A_{\Delta m}}}\right)\right\} - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} \left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) (n-1) e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}}. \tag{39}$$

The kernel part is explicitly expressed as

$$p_{Mm}^{\text{kernel}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} \left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) \left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle} + 1\right)$$
$$\times e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}} \prod_{i=1}^{n} \left\{1 - \left(\frac{\tau_m^{(i)} - \frac{\langle\tau_M\rangle}{A_{\Delta m}}}{\tau_M + \frac{\langle\tau_M\rangle}{A_{\Delta m}}}\right)\right\}. \tag{40}$$

Note that functions (36)–(40) do not satisfy the normalization condition. Furthermore, in some cases, $p_{Mm}^{\text{approx}}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)})$ in Equations (36) and (39) may take negative values when the correction term is larger than the kernel part. The relationship between the inverse probability density function and its approximation function is discussed in Appendix G.

5 Examination of Bayesian updating method in uncorrelated time series

In this section, we compute the inverse probability density function given by Equation (24) and the (part of) approximation function (Equations (36)–(40)) for the numerically generated uncorrelated time series, and we compare their properties (Tanaka & Umeno, 2021). We examine the numerical method of the Bayesian updating by changing some conditions to see its utility.

5.1 Time series generation and Bayesian updating methods

The uncorrelated time series can be numerically generated by setting $\lambda(t) \equiv \lambda_0$ in Equation (1). In fact, it is numerically generated as the renewal process in which magnitudes and time intervals are generated randomly obeying Equations (9) and (10), respectively. We set the parameter values to be $b = 1$ and $\lambda_0 = 0.0007$. The magnitude thresholds are set to $(M, m) = (5.0, 3.0)$. The b-value condition of Equation (25) is satisfied for these settings. The occurrence time of each event is recorded to 20 decimal places. For such time series, Bayesian updating is applied as explained below.

Bayesian updating is executed for each lower interval in the order of appearance starting from the one immediately after the event with a magnitude above M by substituting their lengths $\{\tau_m^{(1)},\tau_m^{(2)},\cdots,\tau_m^{(n)}\}$ into Equations (24), (39), and (40). The summation of the lower intervals at the n-th update $\sum_{i=1}^{n}\tau_m^{(i)}$ is equivalent to the elapsed time T from the previous event with a magnitude above M. Further, the updating is performed until the event immediately before the next large event with a magnitude above M (i.e., the rightmost lower interval in an upper interval is not used). Therefore, we consider only cases where at least one event is (or two lower intervals are) included in an upper interval.

In addition, we use the following numerical method based on Equations (36) and (37). First, we generate $\mathcal{N}$ time series each contains 10^5 events as sample data. From these sample data, numerically obtain the statistics required for the calculating Equations (36) and (37), i.e., $p_m(\tau_m)$, $p_M(\tau_M)$, $p_{mM}(\tau_m|\tau_M)$ and $P_i(\tau_m|\tau_M)$, and the average number of lower intervals included in the upper interval of length τ_M, $\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M}$. Although the last one is a quantity related to the conditional probability, we calculate it separately. Moreover, we calculate only $P_1(\tau_m|\tau_M)$ and use it instead of $P_i(\tau_m|\tau_M)$ for $i \geq 2$.

These statistics are obtained as a vector or a matrix on discretized intervals as

$$\begin{aligned}\tau_{m,j} &:= 10^{(j+0.5)\Delta\tau_m},\\ \tau_{M,k} &:= 10^{(k+0.5)\Delta\tau_M},\end{aligned} \tag{41}$$

where $j, k \in \mathbb{Z}$, such that

$$\begin{aligned}\boldsymbol{p}_m &= [p_{m,j}]_{j=j_{\min},\cdots,j_{\max}},\\ \boldsymbol{p}_M &= [p_{M,k}]_{k=k_{\min},\cdots,k_{\max}},\\ \boldsymbol{p}_{mM} &= \left[[p_{mM,jk}]_{j=j_{\min}^{(k)},\cdots,j_{\max}^{(k)}}\right]_{k=k_{\min},\cdots,k_{\max}},\\ \boldsymbol{P}_1 &= \left[[P_{1,jk}]_{j=j_{\min}^{(k)},\cdots,j_{\max}^{(k)}}\right]_{k=k_{\min},\cdots,k_{\max}}.\end{aligned} \tag{42}$$

In Equation (42), $j_{\min}, j_{\max}, k_{\min}$, and $k_{\max}$ represent the smallest and largest bin numbers of each distribution. For the statistics obtained as a matrix, the range of j depends on k, and this is indicated as $j^{(k)}_{\min}$ and $j^{(k)}_{\max}$. The ranges of j and k are different for distribution; however, the same symbol is used in Equation (42). In this paper, we fix $\Delta\tau_m = 0.1$, and in this section, we examine the cases $\mathcal{N} = 10^3, 10^5$ and $\Delta\tau_M = 0.1, 0.025$. In the case $\mathcal{N} = 10^5$, they are fully used only for $p_{mM}(\tau_m|\tau_M)$ and $P_1(\tau_m|\tau_M)$, and only 10^3 of them are used for $p_m(\tau_m)$, $p_M(\tau_M)$, and $\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M}$.

To use these amounts in numerical Bayesian updating, we perform the following interpolations between the data points and extrapolations outside the data range. We describe these procedures using the example of the case $\mathcal{N} = 10^3$ and $\Delta\tau_M = 0.1$.

First, for the inter-event time distributions ($\boldsymbol{p}_m$ and $\boldsymbol{p}_M$), we interpolate between the data points of each distribution (between $\tau_{m,j}$ and $\tau_{M,k}$, respectively) using cubic spline functions. Outside the data range (i.e., $\tau_m < \tau_{m,j_{\min}}$, $\tau_m > \tau_{m,j_{\max}}$ and $\tau_M < \tau_{M,k_{\min}}$, $\tau_M > \tau_{M,k_{\max}}$), we extrapolate the fitting curve for the edge 10 points (Figure S1). The distributions are defined for all continuous τ_m values and for all $\tau_{M,k}$ using this process.

Second, for the bivariate distributions ($\boldsymbol{p}_{mM}$ and $\boldsymbol{P}_1$), we perform the same interpolations and extrapolations for $\tau_{m,j}$ (Figures S2 and S3). Meanwhile, for $\tau_{M,k}$, the domain is extended using the average of the functions at $\{\tau_{M,k_{\min}}, \cdots, \tau_{M,k_{\min}+l_e-1}\}$ as the substitute for $\tau_{M,k}$ with $k < k_{\min}$, whereas using the functions at $\{\tau_{M,k_{\max}-l_e+1}, \cdots, \tau_{M,k_{\max}}\}$ as the substitute for $\tau_{M,k}$ with $k > k_{\max}$. We set $l_e = 5$ for $\Delta\tau_M = 0.1$ and $l_e = 20$ for $\Delta\tau_M = 0.025$.

Finally, for $\tau_{M,k}/\langle\langle\tau_m\rangle\rangle_{\tau_{M,k}}$, the interpolation and extrapolation procedures are conducted in the same way as $\boldsymbol{p}_M$, although the extrapolation functions are different (Figure S4).

Thus, the discrete variable $\tau_{m,j}$ becomes continuous as τ_m and the distribution functions are defined for all τ_m larger than 0. This makes it possible to return a value for any input of the length of a lower interval when performing Bayesian updating. Further, the distribution functions are defined for any k in Equation (41). We set the range of k to be $-120 \le k \le 70$ for $\Delta\tau_M = 0.1$, and $-480 \le k \le 280$ for $\Delta\tau_M = 0.025$. Although this yields the maximum range of the Bayesian updating, the updating at the n-th step is performed within the range $\max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\} < \tau_M$. The properties of the inverse probability density function and the (part of) approximation function are examined within this range.

The kernel parts of the approximation functions are computed by calculating Equation (37) in a step-by-step manner as

$$
\begin{aligned}
\ln p^{\text{kernel}}_{Mm}(\tau_{M,k}|\tau_m^{(1)}) &= \ln\left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\frac{\tau_{M,k}}{\langle\langle\tau_m\rangle\rangle_{\tau_{M,k}}}\right) + \ln p_{mM}(\tau_m^{(1)}|\tau_{M,k}) - \ln p_m(\tau_m^{(1)}) + \ln p_{M,k},\\
\ln p^{\text{kernel}}_{Mm}(\tau_{M,k}|\tau_m^{(1)}, \tau_m^{(2)}) &= -\ln\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln p_{mM}(\tau_m^{(2)}|\tau_{M,k}) - \ln p_m(\tau_m^{(2)}) + \ln p^{\text{kernel}}_{Mm}(\tau_{M,k}|\tau_m^{(1)}),\\
\ln p^{\text{kernel}}_{Mm}(\tau_{M,k}|\tau_m^{(1)}, \tau_m^{(2)}, \tau_m^{(3)}) &= -\ln\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln p_{mM}(\tau_m^{(3)}|\tau_{M,k}) - \ln p_m(\tau_m^{(3)}) + \ln p^{\text{kernel}}_{Mm}(\tau_{M,k}|\tau_m^{(1)}, \tau_m^{(2)}),\\
&\vdots
\end{aligned}
\tag{43}
$$

The correction terms of the approximation functions are calculated by first update as

$$\begin{aligned}
\ln p_{Mm}^{\mathrm{correct}}(\tau_{M,k}|\tau_m^{(1)}) &= \ln\left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(1)}|\tau_{M,k}) - \ln p_m(\tau_m^{(1)}) + \ln p_{M,k},\\
\ln p_{Mm}^{\mathrm{correct}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)}) &= -\ln\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(2)}|\tau_{M,k}) - \ln p_m(\tau_m^{(2)}) + \ln p_{Mm}^{\mathrm{correct}}(\tau_{M,k}|\tau_m^{(1)}),\\
\ln p_{Mm}^{\mathrm{correct}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)},\tau_m^{(3)}) &= -\ln\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(3)}|\tau_{M,k}) - \ln p_m(\tau_m^{(3)}) + \ln p_{Mm}^{\mathrm{correct}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)}),\\
&\vdots
\end{aligned} \tag{44}$$

and then, we add $\ln(n-1)$ for each $\ln p_{Mm}^{\mathrm{correct}}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)})$.

The approximation functions are obtained by adding together the kernel part and the correction term calculated by these separate updates. The approximation functions are calculated only for such k's that $p_{\sup} > \ln p_{Mm}^{\mathrm{kernel}}, \ln p_{Mm}^{\mathrm{correct}} > p_{\inf}$. Here, $p_{\sup}(=600)$ and $p_{\inf}(=-600)$ yield the upper and lower limits of p_{Mm}^{kernel} and $p_{Mm}^{\mathrm{correct}}$ to ensure that these are within the range of the computer capacity. In addition, such k's for which the correction term is so large that Equation (36) becomes negative are excluded.

Figure S5 shows an example of Bayesian updating for the uncorrelated time series. The inverse probability density function given by Equation (24) has a characteristic peak that is not observed in $p_M(\tau_M)$. The correction term makes the kernel part obtained from Equation (40) closer to the inverse probability density function. Moreover, the numerical calculations based on Equations (36) and (37) with $\mathcal{N} = 10^3$ and $\Delta\tau_M = 0.1$ appear to be consistent with these results.

In the next subsection, we compare these functions statistically to examine numerical Bayesian updating method.

5.2 Examination of numerical Bayesian updating method

In this subsection, we compare the probability density functions and the (part of) approximation functions statistically. The Bayesian updating method described in the previous subsection is applied to 100 test data time series, each containing 10^5 events prepared separately from the sample data.

5.2.1 Comparison by distance

We define the distance for two square-integrable functions $f(\cdot)$ and $g(\cdot)$ as

$$D(f||g) := \int_T^\infty |f(\tau_M) - g(\tau_M)|^2 \, d\tau_M. \tag{45}$$

The range of the integral is set to (T,∞) to exclude the Dirac's delta function at $\tau_M = T$ in the inverse probability density function. For $f = p_{Mm}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)})$ and $g = p_M(\tau_M)$, the distance can be analytically derived (Appendix H), whereas when $f(\cdot)$ or $g(\cdot)$ is the (part of) approximation function, the distance is calculated numerically as

$$D(f||g) \simeq \sum_{\substack{k;\tau_{M,k}>T\\ p_{\sup}>\ln f,\ln g>p_{\inf}}} |f(\tau_{M,k}) - g(\tau_{M,k})|^2 (\ln 10)\tau_{M,k}\Delta\tau_M. \tag{46}$$

$D(f||g)$ is calculated for each update throughout the 100 test data time series. If no k's satisfy $p_{\sup} > \ln f, \ln g > p_{\inf}$, it is not included in the following calculation. The average distance $\langle D(f||g)\rangle$ is calculated by averaging these distances for each elapsed time $T \in [10^{0.1l}, 10^{0.1(l+1)})$ with $l \in \mathbb{Z}$ from the previous event larger than M.

Figure 7(a) shows the average distance for the cases $f = p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$, $p_{Mm}^{\text{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$, $p_{Mm}^{\text{kernel}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$, and $g = p_M(\tau_M)$. In addition to the analytical calculation in Equation (45) for $D(p_{Mm}||p_M)$, the results of the numerical integration of Equation (46) are presented; the calculations using Equations (39) and (40) are indicated by $D'(\cdot||\cdot)$. The results of the calculation using Equations (36) and (37) with the numerical method in §5.1 with $\mathcal{N} = 10^3$ and $\Delta\tau_M = 0.1$ are presented by $D''(\cdot||\cdot)$. The results for $\mathcal{N} = 10^5$ with $\Delta\tau_M = 0.1$ and $\Delta\tau_M = 0.025$ are shown in Figure S6.

First, one can see that $\langle D'(p_{Mm}||p_M)\rangle$ is almost consistent with $\langle D'(p_{Mm}^{\text{approx}}||p_M)\rangle$, which indicates that $p_{Mm}^{\text{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$ derived in the previous section certainly approximates the inverse probability density function, regardless of the elapsed time (or regardless of the number of updates, because the occurrence rate is constant). However, these separate from $D(p_{Mm}||p_M)$ at around $T \sim 10^5$ and at a large T. As such separations disappear when $\Delta\tau_M = 0.025$ (Figure S6(c,d)), this is attributed to the coarseness of the numerical integration.

Second, $\langle D'(p_{Mm}^{\text{kernel}}||p_M)\rangle$ is nearly consistent with $\langle D''(p_{Mm}^{\text{kernel}}||p_M)\rangle$. This suggests that the numerical updating method in Equation (43) certainly calculates the kernel part. However, $\langle D''(p_{Mm}^{\text{approx}}||p_M)\rangle$ gradually separates from $\langle D'(p_{Mm}^{\text{approx}}||p_M)\rangle$ at a large T. This separation is more clearly illustrated in Figure 7(b), which shows the average distances between $f = p_{Mm}^{\text{approx}}, p_{Mm}^{\text{kernel}}$ and $g = p_{Mm}$ calculated by Equation (46). This separation can be attributed to the calculation of the correction term in Equation (44), in particular to the fluctuation in the numerically obtained $\boldsymbol{P}_1$ (Appendix I).

5.2.2 Comparison by maximum peak time

In the previous subsection, the approximation function calculated by the numerical Bayesian updating method is suggested to be separate from the inverse probability density function. However, we show that such a separation does not have a considerable effect around the maximum peak. To this end, we further compare the maximum points (hereafter, maximum peak time) of the inverse probability density function in Equation (23) and its approximation function in Equation (36) with the numerical updating method, each denoted by $\hat{\tau}_M^{\max}$ and $\tau_M^{\max,\text{approx}}$. Both functions are descretized as Equation (41); the corresponding k in Equation (41) is denoted by $\hat{k}^{\max}$ and $k^{\max,\text{approx}}$, respectively.

$\hat{k}^{\max}$ and $k^{\max,\text{approx}}$ are numerically searched for each update. These are determined as such k that the function takes the maximum value within the range for which the above mentioned numerical results are obtained, while excluding its edges. Thus, if $\hat{k}^{\max}$ or $k^{\max,\text{approx}}$ is located at such edges, it is not considered the peak and is set to $k = 80$ when $\Delta\tau_M = 0.1$ and $k = 320$ when $\Delta\tau_M = 0.025$. Further, when the numerical results of the approximation function are not obtained for any k (when the correction term exceeds the kernel part for all k), $k^{\max,\text{approx}}$ is set to be 80 or 320.

Figure 8 shows the joint probability mass function (p.m.f.) of $(\hat{k}^{\max}, k^{\max,\text{approx}})$ for $\mathcal{N} = 10^3$ and $\Delta\tau_M = 0.1$. Those for $\mathcal{N} = 10^5$ are presented in Figure S7. Here, the population is all the pairs of $(\hat{k}^{\max}, k^{\max,\text{approx}})$ obtained for each update throughout the test data. The maximum peak search is conducted in the two ranges; (a) $\tau_M > \max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$, and (b) $\tau_M > T$. In the former case, the p.m.f. is bimodal; the higher peak exists around $\hat{k}^{\max} = k^{\max,\text{approx}}$, and the other lower peak around $\hat{k}^{\max} > k^{\max,\text{approx}}$. The second peak disappears in the latter case, and the first peak is intrinsic, i.e., the positions of the maximum peak are close between the inverse probability density function and its approximation function. The situation is the same for other cases (Figure S7). These results indicate that it is the off-peak region of the approximation function that contributes to the separation of the average distances.

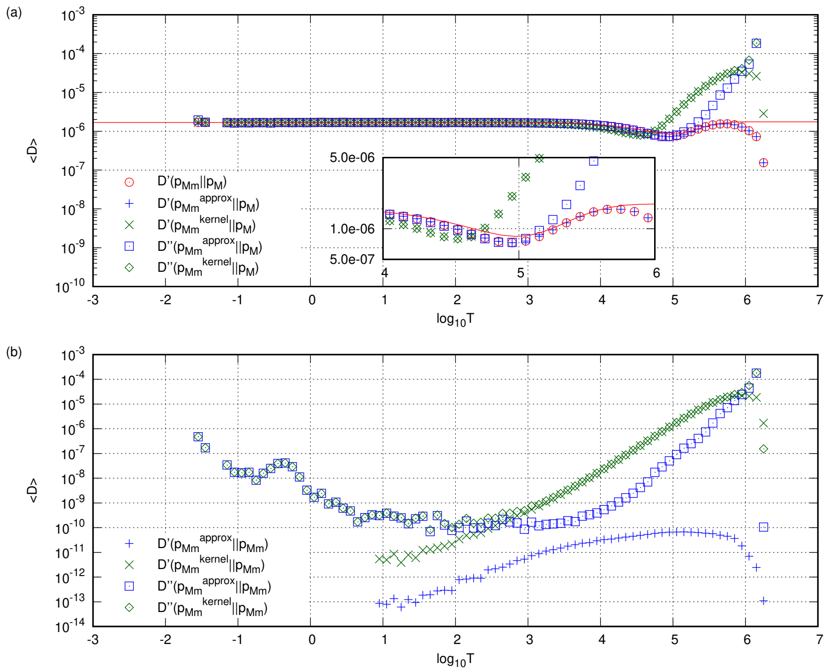

Figure 7. Average distances for each elapsed time (T) from the previous large event with a magnitude above M. (a) Distances between the inter-event time distribution and other function. $D(p_{Mm}||p_M)$ (Equation (H2) in Appendix H) is shown by the red curve, and the symbols are numerical results for Equation (46). (b) Distances between the inverse probability density function and other function numerically calculated by Equation (46).

The results obtained in this section indicate that the numerical method using 1100 time series (1000 for sample data and 100 for test data) is sufficient to calculate the kernel part as well as the maximum peak time of the approximation function that is important in the inference, and to examine their statistical property. Further, these results indicate that Bayesian updating can be applied with the numerical method even if the explicit functional forms of the inter-event time distribution and the conditional probability density function and so on are unclear, such as the time series of the ETAS model.

6 Bayesian updating for the time series of the ETAS model

In this section, Bayesian updating is applied to the time series of the ETAS model (Tanaka & Umeno, 2021). In this case, due to the correlations among events, it is difficult to derive the inverse probability density function and its approximation function analytically. Therefore, we compute the approximation function (Equation (36)) and its kernel part (Equation (37)) using the numerical Bayesian updating method. The maximum peak time of the kernel part is used as the estimate for the occurrence time of the next large event, and the effectiveness of forecasting based on that estimate is evaluated statistically.

6.1 Time series generation and Bayesian updating methods

We apply the numerical Bayesian updating method in §5.1 to the time series generated by Equation (1) with the parameter values $b = 1$, $\alpha = 0.8$, $\theta = 0.2$ ($p = 1.2$), $c = 0.01$, $M_0 = 3$, $\lambda_0 = 0.0007$, and $K = 0.0125$. The magnitude thresholds are

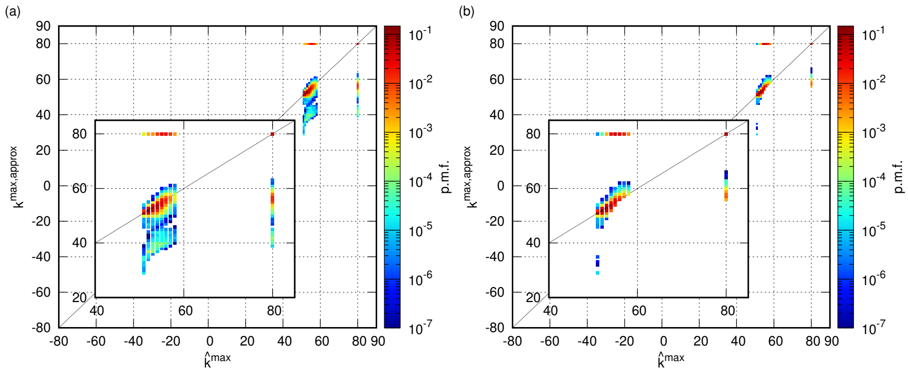

Figure 8. Joint probability mass function for $(\hat{k}^{\max}, k^{\max,\mathrm{approx}})$. Numerical search of the maximum peak is conducted for (a) $\tau_M > \max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ and (b) $\tau_M > T$. The horizontal line at $k^{\max,\mathrm{approx}} = 80$ and the vertical line at $\hat{k}^{\max} = 80$ correspond to the cases when the peak is not detected by the peak search.

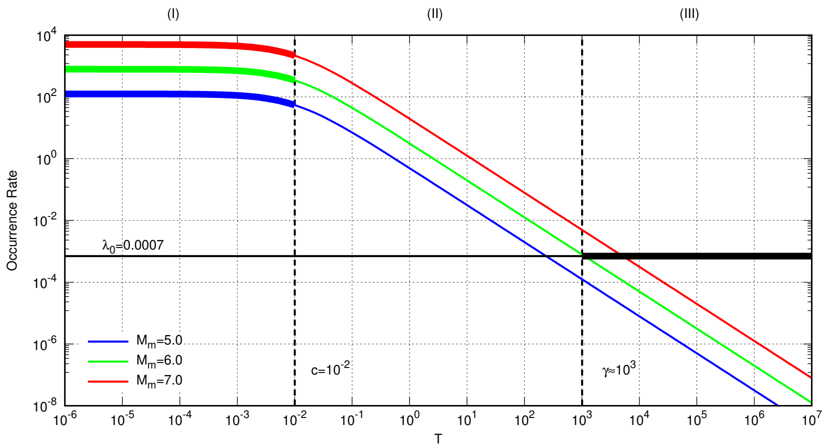

Figure 9. Omori–Utsu law for the parameter values in the text with a different mainshock magnitude M_m. The number of aftershocks per unit day against the elapsed time (T) from the mainshock obeys $\lambda(T) = K10^{\alpha(M_m - M_0)}/(T+c)^{\theta+1}$. The background rate ($\lambda_0 = 0.0007$) is also shown.

$(M, m) = (5.0, 3.0)$. Although the entire time series is stationary because the branching ratio ($n_{\mathrm{br}} \approx 0.785$) is less than 1, it is locally non-stationary obeying the Omori–Utsu law after a large event, as shown in Figure 9. The activity can be categorized into three regimes with respect to the elapsed time (T) from the mainshock, as summarized in Table 1.

We prepare 1100 time series, with each containing 10^5 events. First, random numbers generated from five different seed values are used to generate 240 time series for each seed. Among them, those contain events with magnitude above 10 are excluded. This is because the aftershock sequence excited by such an unrealistic large event do not fit within a single time series, and then, the non-stationarity affects the statistics of the sample data. We use 1100 of the remaining time series. $\mathcal{N} = 1000$ are used as the sample data to obtain statistics with $\Delta\tau_M = 0.1$; the interpolation and extrapolation procedures are conducted with $l_e = 5$ in the same way as explained in §5.1 (Figures S8–S12). Bayesian updating (Equations (43) and (44)) is applied to the remaining 100 time se-

Table 1. *Three Regimes in the time series of the ETAS model*

Category	Regime	Property
(I)	$T \lesssim c (= 0.01)$	Stationary, high occurrence rate
(II)	$c \lesssim T \lesssim \gamma (\approx 10^3)$	Non-stationary, relaxation process
(III)	$\gamma \lesssim T$	Stationary, low occurrence rate ($\lesssim \lambda_0$)

ries. The maximum range of k is $-120 \leq k \leq 70$, and the n-th update from the occurrence time of the event above M is conduced in the range $\max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\} < \tau_M$. The numerical update is conducted when the lower interval is above 0 (for the occurrence times recorded to 20 decimal places); otherwise, the update is skipped.

The following normalizations are performed in the calculations of the Bayesian updating. The result of the calculation in Equation (43) can be very large. In order to compute the approximation functions together with Equation (44), it is necessary to use the function value of p_{Mm}^{kernel} as it is, though it can exceed p_{sup}. Therefore, to avoid such enlargement, we normalize the result of Equation (43) for each update by subtracting the following numerical integration from Equation (43).

$$\ln \left(\sum_{\substack{k;\tau_{M,k}>T \\ p_{\mathrm{sup}}>\ln p_{Mm}^{\mathrm{kernel}}>p_{\mathrm{inf}}}} p_{Mm}^{\mathrm{kernel}}(\tau_{M,k}|\tau_m^{(1)}, \cdots, \tau_m^{(n)})(\ln 10)\tau_{M,k}\Delta\tau_M \right). \tag{47}$$

Further, it is necessary to subtract Equation (47) from the correction term in Equation (44) at the same time (thereby the entire approximation function is multiplied by a constant). Thus, for each update of Equations (43) and (44), the numerical integration (47) is computed and subtracted from both.

6.2 Comparison of the approximation function and its kernel part

It is difficult to obtain $P_i(\tau_m|\tau_M)$ for the ETAS model, and therefore, we examine the contribution from the correction term to the approximation function as follows. Instead of $P_i(\tau_m|\tau_M)$, we calculate the probability density functions $P^{\mathrm{L}}(\tau_m|\tau_M)$ and $P^{\mathrm{R}}(\tau_m|\tau_M)$ (Figures S10 and S11). According to the Omori–Utsu law, we consider that these two are the end-members of $P_i(\tau_m|\tau_M)$. Then, the approximation functions are calculated numerically by replacing all $P_i(\tau_m|\tau_M)$'s in Equation (44) by either $P^{\mathrm{L}}(\tau_m|\tau_M)$ or $P^{\mathrm{R}}(\tau_m|\tau_M)$. We denote the maximum peak times of these approximation functions by $\tau_M^{\mathrm{max,L}}$ and $\tau_M^{\mathrm{max,R}}$, and their corresponding k's in Equation (41) by $k^{\mathrm{max,L}}$ and $k^{\mathrm{max,R}}$, respectively. Similarly, they are denoted by τ_M^{max} and k^{max} for the kernel part, hereafter. The numerical search of $k^{\mathrm{max,L}}$, $k^{\mathrm{max,R}}$, and k^{max} is conducted in the same way as indicated in §5.2 in the range $\tau_M > \max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$.

Figure 10 shows the joint p.m.f. of $(k^{\mathrm{max}}, k^{\mathrm{max,L}})$ and $(k^{\mathrm{max}}, k^{\mathrm{max,R}})$, which is calculated in the same way as indicated in §5.2. The maximum peak time of the kernel part is not significantly affected by the correction term, and then, it can be used to infer that of the inverse probability density function. However, its confidence interval or average cannot be used because the correction term is not taken into account. In the following, we use the maximum peak time of the kernel part (τ_M^{max}) as the estimator of the occurrence time of the next large event with a magnitude above M, and we discuss the effectiveness of the forecasting based on the estimates.

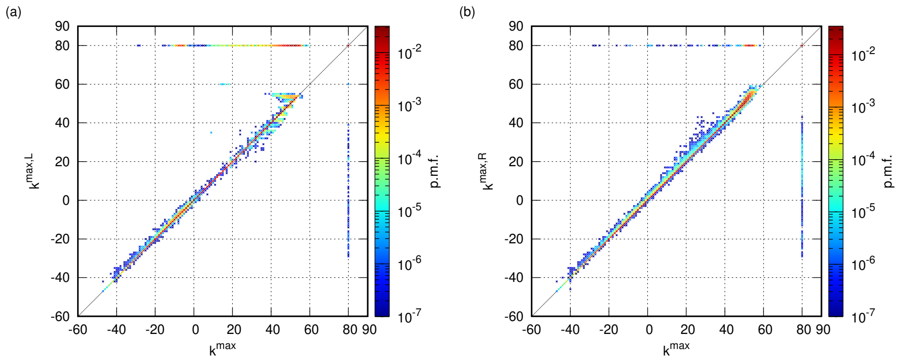

Figure 10. Joint probability mass functions for (a) $\left(k^{\max}, k^{\max,\mathrm{L}}\right)$ and (b) $\left(k^{\max}, k^{\max,\mathrm{R}}\right)$. The horizontal lines at $k^{\max,\mathrm{L}} = 80$ and $k^{\max,\mathrm{R}} = 80$ and the vertical line $k^{\max} = 80$ are the cases where the peak is not detected.

6.3 Estimation of the occurrence time of the next large event and effectiveness of forecasting

We denote the estimate at the n-th update by $\tau_M^{\max,\mathrm{n}} (= 10^{(k^{\max,\mathrm{n}}+0.5)\Delta\tau_M})$, and the actual elapsed time of the next large event from the previous one by τ_M^*. We evaluate the accuracy of the estimation at the n-th update using the relative error (δ_n), which is given as

$$\delta_n := \frac{\tau_M^* - \tau_M^{\max,\mathrm{n}}}{\tau_M^*}, \tag{48}$$

Equation (48) considers that the error $|\tau_M^* - \tau_M^{\max,\mathrm{n}}|$ gets larger as τ_M^* becomes longer. The relative error makes it possible to evaluate the accuracy in a manner that is comparable regardless of τ_M^*.

The accuracy at the n-th update is judged by whether δ_n is within the threshold (δ_{th})

$$-\delta_{\mathrm{th}} \le \delta_n \le \delta_{\mathrm{th}}. \tag{49}$$

When Equation (49) is satisfied, the estimation at the n-th update is judged to be plausible for the given threshold value δ_{th} in the present paper. This is equivalent for the actual occurrence time to be within the range

$$\frac{\tau_M^{\max,\mathrm{n}}}{(1+\delta_{\mathrm{th}})} \le \tau_M^* \le \frac{\tau_M^{\max,\mathrm{n}}}{(1-\delta_{\mathrm{th}})}. \tag{50}$$

Based on the above accuracy at each update, we further evaluate whether a series of estimations yields effective forecasting. Here, effective forecasting implies that $\tau_M^{\max}$ takes a nearly constant value around τ_M^* continuously from well before the occurrence time of the next large event. This can be quantitatively expressed as follows: Let $n_{\le\mathrm{th}}$ be the number of consecutive updates immediately before the next large event, in which Equation (49) is satisfied. Further, we denote the last update as the n_{fin}-th update. When the sequence of updates with a sufficiently long $n_{\le\mathrm{th}}$ exists in the range of $n \in (n_{\mathrm{fin}} - n_{\le\mathrm{th}}, n_{\mathrm{fin}}]$, we consider the forecasting to be effective. We judge the stability of $\tau_M^{\max,\mathrm{n}}$ by Equation (49), and therefore, δ_{th} should not be too large. In the present paper, we set $\delta_{\mathrm{th}} = 0.5$ and 0.25.

To observe the relationship between the effectiveness of forecasting and the stationarity of the time series, we examine the occurrence rate (R_n), variation of its log ($\Delta \log_{10} R_n$), and variation of log-estimate ($\Delta k_n^{\max}$) defined below.

$$R_n := 10/\left(t_{n+9} - t_n\right), \tag{51}$$

$$\Delta \log_{10} R_n := \log_{10} R_{n+10} - \log_{10} R_n, \tag{52}$$

$$\Delta k_n^{\max} := k^{\max,\mathrm{n}+10} - k^{\max,\mathrm{n}}, \tag{53}$$

where t_n represents the occurrence time of the n-th update.

Figures S11–S13 show examples of Bayesian updating for each regime in Table 1. Although these are only examples and not all updating proceeds in this way, these examples suggest that the stability of the estimate is related to the stationarity of the time series.

6.4 Statistical analysis of the effectiveness of forecasting

We show the results of the statistical analysis on the effectiveness of forecasting. Only the cases of $n_{\mathrm{fin}} \geq 30$ are used in the analysis to ensure that the temporal information of lower intervals is fully reflected in the estimate. Figure 11(a) shows the total number of upper intervals (N) obtained from the test data for each $\tau_M^* \in [10^{0.5l}, 10^{0.5(l+1)})$ with $l \in \mathbb{Z}$. Further, N_{30} represents the total number of upper intervals such that $n_{\mathrm{fin}} \geq 30$, which is shown with the ratio to N. The updates included in these N_{30} upper intervals are analyzed.

Figures 11(b–d) show the results of the statistical analysis with $\delta_{\mathrm{th}} = 0.5$. Figure 11(b) shows the probability (P_{fin}) of $n_{\leq \mathrm{th}} > 0$ (or $|\delta_{n_{\mathrm{fin}}}| \leq \delta_{\mathrm{th}}$) for each τ_M^*. The average of P_{fin} for the overall τ_M^* is about 0.52, and the P_{fin} for each τ_M^* is about the same, except for $\tau_M^* > \langle \tau_M \rangle$ in which P_{fin} takes a higher probability around 0.67. Of such $n_{\leq \mathrm{th}} > 0$ cases, the proportion ($P_{\geq 30}$) of those with relatively long $n_{\leq \mathrm{th}} \geq 30$ is also shown in Figure 11(b) (the probability distribution of $n_{\leq \mathrm{th}}$ is shown in Figure S16(a)). Thus the regions of high $P_{\geq 30}$ are overlapped with regimes (I) and (III), though the former is shifted toward larger τ_M^*. On the other hand, $P_{\geq 30}$ is lower in regime (II); it gradually decreases as τ_M^* gets larger. This is consistent with the average of $n_{\leq \mathrm{th}}$ ($\langle n_{\leq \mathrm{th}} \rangle$, this average is taken for $n_{\leq \mathrm{th}} > 0$), but also with the average of its proportion to n_{fin} ($\langle n_{\leq \mathrm{th}}/n_{\mathrm{fin}} \rangle$) as shown in Figure 11(c). This implies that, as the fraction of non-stationary times in $[0, \tau_M^*)$ increases in regime (II), the domination rate of $n_{\leq \mathrm{th}}$ in the total n_{fin}-updates decreases gradually. These properties are preserved for $\delta_{\mathrm{th}} = 0.25$ (Figure S17).

Figure 12 shows the joint probability density-mass functions of $\Delta \log_{10} R$ and $\Delta k^{\max}$ calculated numerically for each τ_M^*. The case $k^{\max} = 80$ is excluded from the population. If τ_M^* is in the regions of high $P_{\geq 30}$, the distribution is almost symmetrically concentrated near the origin. This implies that, when the time series is dominated by stationarity ($\Delta \log_{10} R \approx 0$), the estimated value is stable ($\Delta k^{\max} \approx 0$). On the other hand, if τ_M^* is in regime (II), the probability density function gradually has a region in the second quadrant as τ_M^* gets larger. This region indicates the existence of a non-stationary time series in which the estimate has an increasing trend ($\Delta k^{\max} > 0$).

These results present the following conclusions. First, the probability that the relative error is within the threshold at the last update ($|\delta_{\mathrm{n_{fin}}}| \leq \delta_{\mathrm{th}}$) is almost independent of the actual occurrence time (τ_M^*). This suggests that the length of the upper interval can be estimated by the inverse probability density function reflecting the temporal pattern of lower intervals, at the last update when the information of the lower intervals can be utilized fully. Second, the stationarity of the time series is related to the stability of the estimate; if the time series is non-stationary, it causes the estimate $\tau_M^{\max}$ to shift. Third, the domination rate of stationarity in the time series determines the effectiveness of forecasting. Immediately or long after the large event, the stationary time series is dominant. Therefore, based on the second point mentioned above, the estimate becomes stable, which leads to an effective forecasting with a relatively long $n_{\leq \mathrm{th}}$. However, these regions are not identical to regimes (I) and (III). This is attributed to lag until the ratio of the non-stationary region in the time series becomes dominant. On the other hand, in regime (II), the non-stationarity becomes gradually dominant, which leads to the shifting of $\tau_M^{\max}$ and shortening of $n_{\leq \mathrm{th}}$.

Finally, we discuss the effectiveness of forecasting in terms of the duration time ($\tau_{\leq \mathrm{th}}$) during the $n_{\leq \mathrm{th}}$ updates. Figure 11(d) shows the average of the duration time ($\langle \tau_{\leq \mathrm{th}} \rangle$) and the average of its ratio to the actual occurrence time ($\langle \tau_{\leq \mathrm{th}}/\tau_M^* \rangle$) for each τ_M^* (the probability distribution of $\tau_{\leq \mathrm{th}}$ is shown in Figure S16(b)). Unlike $\langle n_{\leq \mathrm{th}} \rangle$ in Figure 11(c), $\langle \tau_{\leq \mathrm{th}} \rangle$ increases linearly as τ_M^* gets larger, and it is sufficiently long in regime (III). On the other hand, $\tau_{\leq \mathrm{th}}$ is very short in regime (I); however, the ratio $\langle \tau_{\leq \mathrm{th}}/\tau_M^* \rangle$ is high (nearly 0.7). Therefore, from the perspective of the time interval, the forecasting is also considered to be effective immediately or long after the large event.

7 Discussion and Conclusions

The Bayes' theorem and Bayesian updating on the inter-event times at different magnitude thresholds in a marked point process are considered. The analytical results for the uncorrelated time series are used to apply Bayesian updating to the time series of the ETAS model for examining its utility toward forecasting a large event using the temporal pattern of the smaller events.

First, the Bayes' theorem is considered for the general marked point process. The Bayes' theorem provides the relationship between the conditional and inverse probability density functions for the lengths of one upper interval and one lower interval. The inverse probability density function is represented by the generalized forms of the probability density functions of the inter-event times and the conditional probability density function. This inverse probability density function is derived for the uncorrelated time series analytically, and the condition to have a peak is also found.

The Bayes' theorem is extended to Bayesian updating that yields the inverse probability density function between the lengths of multiple consecutive lower intervals and the upper interval that includes them. Although the inverse probability density function is different for the updating manner, we consider the updating without the restriction on the position of the lower intervals. For the uncorrelated time series, the inverse probability density function and its approximation function are derived, and the latter approximation is shown to be reasonable using the distances.

Bayesian updating is applied to the time series of the ETAS model. We numerically analyze the approximation function and its kernel part. We use the maximum point of the kernel part as the estimate of the occurrence time of the next large event because the maximum peaks of these two functions are shown to not be different drastically. The accuracy of the estimation at each update is evaluated by the relative error with the actual occurrence time of the next large event; the effectiveness of the forecasting throughout the series of updates is judged by the continuity of the plausible estimations prior to the large event.

Statistical analysis indicates that the accuracy of the estimation at the last update does not drastically depend on the occurrence time of the next large event. This suggests that the inverse probability density function can estimate the occurrence time of the next large event in response to the temporal pattern of minor events. However, the continuity of plausible estimation depends on the occurrence time of the next large event. This is because the dominance rate of the non-stationary time series in which the estimate becomes unstable varies with the elapsed time from the previous large event obeying the Omori–Utsu law. The stationarity is dominant either immediately after or long

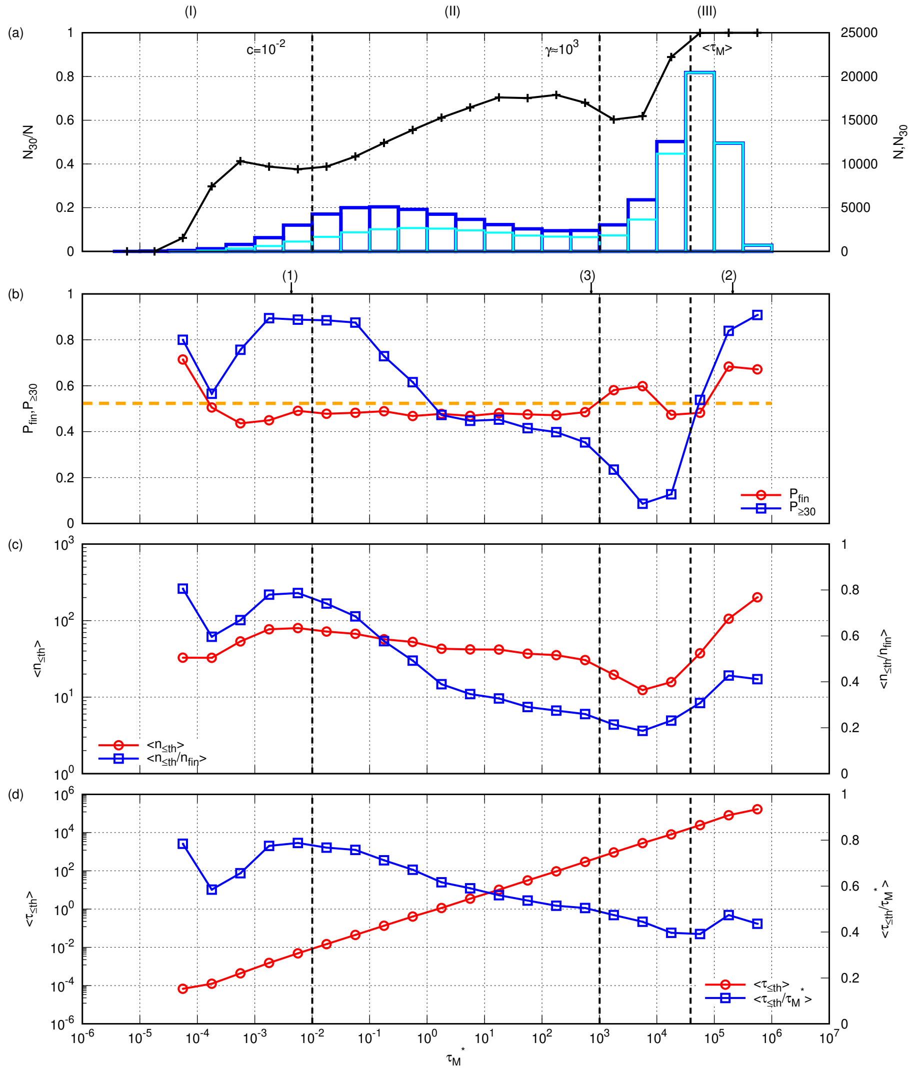

Figure 11. (a) (Blue) Number of upper intervals N, (Cyan) number of upper intervals that include at least 30 updates N_{30}, and (Black) their ratio N_{30}/N, for each τ_M^* in the test data. (b–d) Statistical results with $\delta_{\rm th} = 0.5$ for each τ_M^*. (1)–(3) indicates the τ_M^*'s of the examples in Figures S13–S15. (b) (Red) Probability $P_{\rm fin}$ that $n_{\leq{\rm th}} > 0$ holds (or the probability that Equation (49) is satisfied at the last ($n_{\rm fin}$-th) update), and (Orange dotted line) the average of $P_{\rm fin} \approx 0.52$ for the overall τ_M^*. (Blue) Proportion $P_{\geq 30}$ of such cases among them where $n_{\leq{\rm th}} \geq 30$. (c) (Red) Average of $n_{\leq{\rm th}}$, $\langle n_{\leq{\rm th}}\rangle$, and (Blue) the average of its proportion to the total number of updates, $\langle n_{\leq{\rm th}}/n_{\rm fin}\rangle$. (d) (Red) Average of the duration time $\tau_{\leq{\rm th}}$, $\langle \tau_{\leq{\rm th}}\rangle$ and (Blue) the average of its proportion to the actual occurrence time, $\langle \tau_{\leq{\rm th}}/\tau_M^*\rangle$.

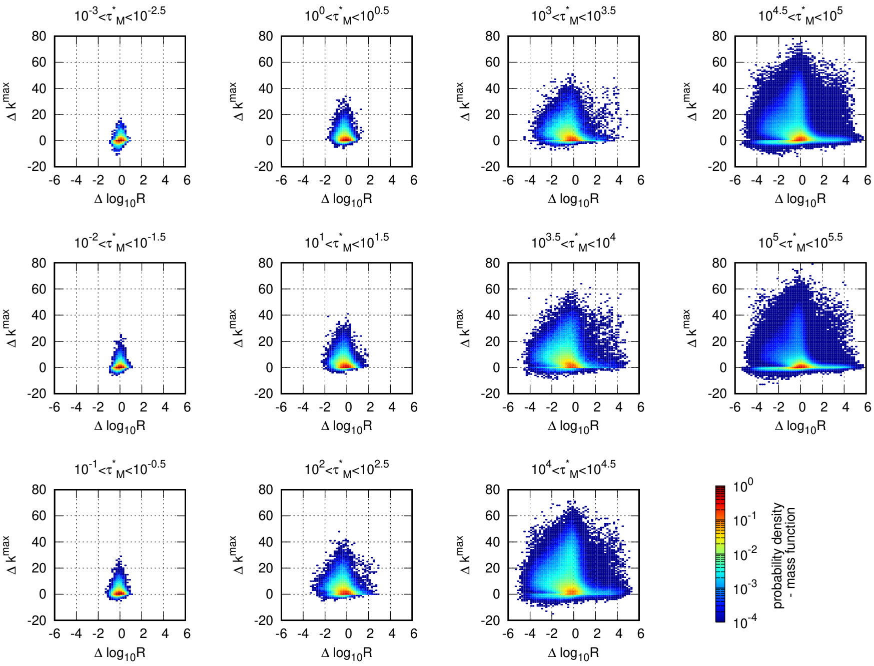

Figure 12. Joint probability density-mass functions of $\Delta \log_{10} R$ and $\Delta k^{\max}$ for each τ^*_M.

after the previous major event. Therefore, the forecasting by the Bayesian updating method can be effective for secondary disaster prevention in the former case, and for long-term risk assessment in the latter case.

The approximation function derived for the uncorrelated time series is applied in the Bayesian updating for the time series of the ETAS model. This allows us to perform the update in the convenient form of the product of the conditional probabilities. However, this implicitly assumes that there is no correlation between events and lower intervals; such an assumption can be reasonable for the stationary part of the time series, although it is not reasonable for non-stationary part. This probably is one of the reasons why forecasting is ineffective in the non-stationary regime.

Further, only the kernel part of the approximation function is used when estimating the occurrence time of the next large event for the time series of the ETAS model. The correction term needs to be investigated in detail to use the entire approximation function to perform point estimation by average, interval estimation, or probabilistic risk assessment using hazard rate. Comparison with the probabilistic evaluation using the inter-event time distribution becomes possible after clarifying the correction term.

Although the statistical property of Bayesian updating is examined for only one set of ETAS parameters, it is considered to be different for activities generated by other parameter values. For example, for the time series with the high background rate (λ_0) that corresponds to taking up a large spatial area, forecasting is considered to be less effective because in such time series, different mainshock-aftershocks sequences overlap (Touati et al., 2009) and the correlations between the upper and lower intervals are weakened. Further, if the background rate is low, forecasting is considered to be improved because a single mainshock-aftershocks sequence is exposed (Touati et al., 2009) and the correlation is easily reflected in the conditional probability. Forecasting is also considered to be improved for the time series with a large branching ratio ($n_{\rm br}$). If the branching ratio is large, the number of aftershocks generated by an event increases (Helmstetter

& Sornette, 2003), which increases the number of updates in the Bayesian updating; this is advantageous for forecasting.

In this study, only one lower threshold magnitude (m) is set for a given upper threshold (M). One approach for performing Bayesian updating using more temporal information of the lower intervals is to set multiple lower thresholds ($m_1 < m_2 < \cdots (< M)$). When setting such lower thresholds, it is better to refer to the condition of $\Delta m > \log_{10} 3/b$ so that the inverse probability density function has a peak. This condition indicates that there is a trade-off between the b-value and Δm, and then, the range of lower thresholds that can be set varies with the b-value. However, the condition $\Delta m > \log_{10} 3/b$ is for the uncorrelated time series; finding the corresponding condition for the time series of the ETAS model is a future work. Considering such an extension is important for applying the Bayesian updating method to the real seismic catalogs in which the number of earthquakes is limited.

To utilize as much temporal information on small events as possible using the method described above is important for applying the Bayesian updating method to real seismic catalogs with a limited number of data. Another idea to apply the Bayesian updating method to seismic catalogs while compensating for the shortage of data is to use the ETAS model in combination. The ETAS model can be used to generate a sufficient amount of synthetic data with the parameter set determined for the past seismic activity. From such synthetic data, statistical amounts necessary in the numerical Bayesian updating method is obtained precisely. Moreover, it is necessary to develop further ingenuity by studying the properties of the conditional and inverse probability density functions through the analysis of seismic catalogs. With these auxiliaries, the application of the Bayesian updating method to real seismic activity is expected to proceed while solving the limitation of seismic data.

Appendix A Derivation of the conditional and inverse probability density functions for the uncorrelated time series

First, we derive the conditional probability density function (Equation (15)) by substituting Equations (12)–(14) into Equation (11). The denominator of Equation (11) is

$$\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M) = A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle} + 1,$$

and the numerator is

$$\begin{aligned}
&\sum_{i=1}^{\infty} i\rho_{mM}(\tau_m|i,\tau_M)\Psi_{mM}(i|\tau_M) = e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m)\\
&+e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\frac{A_{\Delta m}}{\langle\tau_M\rangle}\sum_{i=0}^{\infty}(i+2)\frac{\left\{A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m}{\tau_M}\right)\right\}^i}{i!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m}{\tau_M}\right)}\theta(\tau_M-\tau_m)\\
&= e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m) + e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left\{A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m}{\tau_M}\right)+2\right\}\theta(\tau_M-\tau_m).
\end{aligned}$$

Equation (15) is obtained by rearranging the above equations.

We confirm that Equation (2) with this conditional probability in its kernel has the exponential distribution (Equation (10)) as the solution. By dividing both sides of Equation (2) by N_m and rewriting it using $N_M/N_m = \langle\tau_m\rangle/\langle\tau_M\rangle$ as well as Equation (15)

$$\begin{aligned}
p_m(\tau_m) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\int_{\tau_m}^{\infty}\Big[&e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m)\\
&+\frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\left\{A_{\Delta m}\frac{\tau_M-\tau_m}{\langle\tau_M\rangle}+2\right\}\theta(\tau_M-\tau_m)\Big]p_M(\tau_M),
\end{aligned}\qquad\text{(A1)}$$

where the following general relation is used.

$$\frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} = \sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M).$$

We show that the r.h.s. of Equation (A1) is equivalent to the l.h.s., $p_m(\tau_m) = e^{-\frac{\tau_m}{\langle\tau_m\rangle}}/\langle\tau_m\rangle$. Substitute $p_M(\tau_M) = e^{-\frac{\tau_M}{\langle\tau_M\rangle}}/\langle\tau_M\rangle$ into the r.h.s. of Equation (A1) and note that $A_{\Delta m} + 1 = \langle\tau_M\rangle/\langle\tau_m\rangle$; the integral involving the delta function (R_1) is

$$R_1 = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} e^{-\frac{\tau_m}{\langle\tau_m\rangle}}, \tag{A2}$$

and the integral involving the step function (R_2) is

$$\begin{aligned} R_2 &= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^3} A_{\Delta m} e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}} \int_{\tau_m}^{\infty} \left\{ A_{\Delta m}\frac{\tau_M - \tau_m}{\langle\tau_M\rangle} + 2 \right\} e^{-\frac{\tau_M}{\langle\tau_M\rangle}} d\tau_M \\ &= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} A_{\Delta m}(A_{\Delta m}+2) e^{-\frac{\tau_m}{\langle\tau_m\rangle}}. \end{aligned} \tag{A3}$$

Therefore, the r.h.s. of Equation (A1) is shown to be equivalent to the l.h.s. of Equation (A1) as follows:

$$\begin{aligned} R_1 + R_2 &= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2} \left(1 + A_{\Delta m}\right)^2 e^{-\frac{\tau_m}{\langle\tau_m\rangle}} \\ &= \frac{1}{\langle\tau_m\rangle} e^{-\frac{\tau_m}{\langle\tau_m\rangle}}. \end{aligned} \tag{A4}$$

Second, we derive the inverse probability density function (Equation (16)). From Equation (15), the generalized probability density functions for the uncorrelated time series are derived as

$$\begin{aligned} z_m(\tau_m) &= \frac{\tau_m}{\langle\tau_m\rangle^2} e^{-\frac{\tau_m}{\langle\tau_m\rangle}}, \\ z_M(\tau_M) &= \frac{\tau_M}{\langle\tau_M\rangle^2} e^{-\frac{\tau_M}{\langle\tau_M\rangle}}, \\ z_{mM}(\tau_m|\tau_M) &= \frac{\tau_m}{\tau_M} e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}} \left[\delta(\tau_M - \tau_m) + \frac{A_{\Delta m}}{\langle\tau_M\rangle} \left\{ A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1 - \frac{\tau_m}{\tau_M}\right) + 2 \right\} \theta(\tau_M - \tau_m) \right]. \end{aligned}$$

Equation (16) is obtained by substituting the above equations in Equation (8).

Derivative of Equation (16) by τ_M is

$$\frac{\partial}{\partial\tau_M} p_{Mm}(\tau_M(>\tau_m)|\tau_m) = -\frac{\langle\tau_m\rangle^2}{\langle\tau_M\rangle^5} A_{\Delta m}^2 e^{\frac{\tau_m - \tau_M}{\langle\tau_M\rangle}} \left[\tau_M - \left\{ \tau_m + \langle\tau_M\rangle \left(1 - \frac{2}{A_{\Delta m}}\right) \right\} \right].$$

Therefore, the inverse probability density function has a peak at

$$\tau_M^{\max} = \tau_m + \langle\tau_M\rangle \left(1 - \frac{2}{A_{\Delta m}}\right),$$

under the condition of $\tau_M^{\max} > \tau_m$, which is equivalent to

$$\Delta m > \frac{\log_{10} 3}{b}.$$

Appendix B Derivation of Equation (24) from Equation (23)

The summation part in the r.h.s. of Equation (23) is

$$
\begin{aligned}
&\sum_{i=n}^{\infty}(i-n+1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M)\prod_{j=2}^{n}\rho_{mM}\left(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right)\\
&=\Psi_{mM}(n|\tau_M)\rho_{mM}(\tau_m^{(1)}|n,\tau_M)\prod_{j=2}^{n}\rho_{mM}\left(\tau_m^{(j)}|n-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right)\\
&+\sum_{i=n+1}^{\infty}(i-n+1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M)\prod_{j=2}^{n}\rho_{mM}\left(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right).
\end{aligned}
$$

The first term on the r.h.s. of the above equation is transformed by substituting Equations (12)–(14) as

$$
\begin{aligned}
&\frac{\left[A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right]^{n-1}}{(n-1)!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\frac{(n-1)}{\tau_M}\left(\frac{\tau_M-\tau_m^{(1)}}{\tau_M}\right)^{n-2}\frac{(n-2)}{\tau_M-\tau_m^{(1)}}\left(\frac{\tau_M-\tau_m^{(1)}-\tau_m^{(2)}}{\tau_M-\tau_m^{(1)}}\right)^{n-3}\\
&\cdots\frac{\delta(\tau_M-\sum_{i=1}^{n}\tau_m^{(i)})}{\tau_M-\sum_{i=1}^{n-2}\tau_m^{(i)}}=\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{n-1}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta\left(\tau_M-\sum_{i=1}^{n}\tau_m^{(i)}\right).
\end{aligned}
\tag{B1}
$$

The second term except the step function is also transformed as

$$
\begin{aligned}
&\sum_{i=n+1}^{\infty}(i-n+1)\frac{\left[A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right]^{i-1}}{(i-1)!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\frac{(i-1)}{\tau_M}\left(\frac{\tau_M-\tau_m^{(1)}}{\tau_M}\right)^{i-2}\\
&\times\prod_{j=2}^{n}\frac{(i-j)}{\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}}\left(\frac{\tau_M-\sum_{k=1}^{j}\tau_m^{(k)}}{\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}}\right)^{i-j-1}\\
&=\sum_{i=n+1}^{\infty}\frac{(i-n+1)}{(i-n-1)!}\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{i-1}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\left(\tau_M-\sum_{k=1}^{n}\tau_m^{(k)}\right)^{i-n-1}\\
&=\sum_{i=0}^{\infty}\frac{i+2}{i!}\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{i+n}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\left(\tau_M-\sum_{k=1}^{n}\tau_m^{(k)}\right)^{i}\\
&=\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{n}e^{-A_{\Delta m}\frac{\sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}}\sum_{i=0}^{\infty}\frac{i+2}{i!}\left\{\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left(\tau_M-\sum_{k=1}^{n}\tau_m^{(k)}\right)\right\}^{i}e^{-A_{\Delta m}\frac{\tau_M-\sum_{k=1}^{n}\tau_m^{(k)}}{\langle\tau_M\rangle}}\\
&=\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{n}e^{-A_{\Delta m}\frac{\sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}}\left(A_{\Delta m}\frac{\tau_M-\sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_M\rangle}+2\right).
\end{aligned}
\tag{B2}
$$

Finally, Equation (24) is obtained by substituting Equations (B1) and (B2) in Equation (23), with the denominator of the r.h.s. of Equation (23)

$$
\prod_{i=1}^{n}p_m(\tau_m^{(i)})=\frac{1}{\langle\tau_m\rangle^n}e^{-\frac{\sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_m\rangle}}.
$$

Appendix C Another Bayesian updating method

In this appendix, we consider another method of Bayesian updating from the one introduced in §4; this method considers the consecutive lower intervals in the order of the appearance from the last event with magnitude greater than M. We derive the inverse probability density function for this updating method in the uncorrelated time series.

Let $N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ be the total number of such upper intervals of length τ_M that include the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ start from the leftmost one in the upper interval. Further, we denote the inverse probability density function for this updating by $p^*_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$. We derive it by representing $N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ in two ways as follows:

First, we derive $N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ by counting the total number of the upper intervals of length τ_M that include the leftmost consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$. The position of the first interval in the sequence of the consecutive lower intervals is fixed at the leftmost one in an upper interval, and therefore, the number of the sequence $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ in the time series is

$$N_M \prod_{i=1}^{n} p_m(\tau_m^{(i)}) d\tau_m^n.$$

Among them, the number of sequences that belong to the same upper interval is

$$N_M \left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1} \prod_{i=1}^{n} p_m(\tau_m^{(i)}) d\tau_m^n.$$

Therefore, the first representation is obtained as

$$\begin{aligned} &N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\ &\quad = N_M \left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1} \left\{\prod_{i=1}^{n} p_m(\tau_m^{(i)})\right\} p^*_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) d\tau_M d\tau_m^n. \end{aligned}$$

This equation is rewritten using Equation (10) in the explicit form as

$$\begin{aligned} &N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\ &\quad = N_M \left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1} \frac{1}{\langle\tau_m\rangle^n} e^{-\frac{\sum_{i=1}^n \tau_m^{(i)}}{\langle\tau_m\rangle}} p^*_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) d\tau_M d\tau_m^n. \end{aligned} \tag{C1}$$

Second, we derive $N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ by counting the total number of consecutive lower intervals that start from the leftmost one in the upper intervals of length τ_M. There is only one way for the sequence of consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ to be involved in each of the $N_M p_M(\tau_M) d\tau_M$ upper intervals of length τ_M. The probability of the occurrence of that sequence in the upper interval is, when $i(\geq n)$-lower intervals are included in it

$$\rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i - j + 1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right) d\tau_m^n.$$

Therefore, the second representation is obtained as

$$\begin{aligned} &N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\ &= N_M p_M(\tau_M) \sum_{i=n}^{\infty} \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i - j + 1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right) d\tau_M d\tau_m^n. \end{aligned}$$

This equation is rewritten in the explicit form using Equations (12)–(14) in the same way as in Appendix B.

$$\begin{aligned} &N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) = N_M \frac{1}{\langle\tau_M\rangle} e^{-\frac{\tau_M}{\langle\tau_M\rangle}} d\tau_M d\tau_m^n \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{n-1} \\ &\times \left\{ e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}} \delta\left(\tau_M - \sum_{i=1}^{n} \tau_m^{(i)}\right) + \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right) e^{-A_{\Delta m}\frac{\sum_{i=1}^n \tau_m^{(i)}}{\langle\tau_M\rangle}} \theta\left(\tau_M - \sum_{i=1}^{n} \tau_m^{(i)}\right) \right\}. \end{aligned} \tag{C2}$$

Finally, $p^*_{Mm}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)})$ is derived from Equations (C1) and (C2) as

$$\begin{aligned} &p^*_{Mm}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)}) \\ &= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\left\{\frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-\frac{\tau_M-\sum_{i=1}^n\tau_m^{(i)}}{\langle\tau_M\rangle}}\theta\left(\tau_M-\sum_{i=1}^n\tau_m^{(i)}\right)+e^{-\frac{\tau_M-\sum_{i=1}^n\tau_m^{(i)}}{\langle\tau_m\rangle}}\delta\left(\tau_M-\sum_{i=1}^n\tau_m^{(i)}\right)\right\}. \end{aligned} \tag{C3}$$

This is different from Equation (24), which reflects the difference whether the position of lower intervals is specified.

Appendix D Derivation of Equation (29)

First, we substitute Equations (12)–(14) into Equation (28)

$$\begin{aligned} P^{\mathrm{R}}(\tau_m|\tau_M) &= P^{\mathrm{L}}(\tau_m|\tau_M) \\ &= e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m) \\ &+\sum_{i=2}^{\infty}\frac{(i-1)}{\tau_M}\left(1-\frac{\tau_m}{\tau_M}\right)^{i-2}\frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right)^{i-1}}{(i-1)!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\theta(\tau_M-\tau_m). \end{aligned}$$

In the above equation, the summation part of the term including the step function can be transformed as

$$\begin{aligned} &\sum_{i=2}^{\infty}\frac{(i-1)}{\tau_M}\left(1-\frac{\tau_m}{\tau_M}\right)^{i-2}\frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right)^{i-1}}{(i-1)!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}} \\ &= \frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\sum_{i=0}^{\infty}\frac{\left\{A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m}{\tau_M}\right)\right\}^i}{i!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m}{\tau_M}\right)} \\ &= \frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}. \end{aligned}$$

Finally, Equation (29) is obtained by rearranging the above equations.

Appendix E Derivation of Equation (34)

In this appendix, $P(\tau_m^{(1)},\cdots,\tau_m^{(l)}|\tau_M)$ is derived for the uncorrelated time series. First, we divide the summation in Equation (33) into two parts:

$$\begin{aligned} P(\tau_m^{(1)},\cdots,\tau_m^{(l)}|\tau_M) &= \sum_{i=l}^{\infty}\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M)\prod_{j=2}^{l}\rho_{mM}\left(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right) \\ &= \Psi_{mM}(l|\tau_M)\rho_{mM}(\tau_m^{(1)}|l,\tau_M)\prod_{j=2}^{l}\rho_{mM}\left(\tau_m^{(j)}|l-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(j)}\right) \\ &+\sum_{i=l+1}^{\infty}\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M)\prod_{j=2}^{l}\rho_{mM}\left(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right) \end{aligned}$$

This equation is further rewritten by substituting Equations (12)–(14) in the same way as in Appendix B. The second term on the r.h.s. except for the step function is

$$
\begin{aligned}
&\sum_{i=l+1}^{\infty} \frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right)^{i-1}}{(i-1)!} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}} \frac{i-1}{\tau_M}\left(\frac{\tau_M-\tau_m^{(1)}}{\tau_M}\right)^{i-2} \\
&\times \prod_{j=2}^{l} \frac{(i-j)}{\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}}\left(\frac{\tau_M-\sum_{k=1}^{j}\tau_m^{(k)}}{\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}}\right)^{i-j-1} \\
&= \sum_{i=l+1}^{\infty} \frac{1}{(i-l-1)!}\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{i-1} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}} \left(\tau_M-\sum_{k=1}^{l}\tau_m^{(k)}\right)^{i-l-1} \\
&= \sum_{i=0}^{\infty} \frac{1}{i!}\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{i+l} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}} \left(\tau_M-\sum_{k=1}^{l}\tau_m^{(k)}\right)^{i} \\
&= \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{l} e^{-A_{\Delta m}\frac{\sum_{i=1}^{l}\tau_m^{(i)}}{\langle\tau_M\rangle}} \sum_{i=0}^{\infty} \frac{\left\{\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left(\tau_M-\sum_{k=1}^{l}\tau_m^{(k)}\right)\right\}^{i}}{i!} e^{-\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left(\tau_M-\sum_{k=1}^{l}\tau_m^{(k)}\right)} \\
&= \prod_{i=1}^{l} P_i(\tau_m^{(i)}|\tau_M),
\end{aligned}
$$

where $P_i(\tau_m^{(i)}|\tau_M) := \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right) e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}$.

Therefore

$$
P(\tau_m^{(1)},\cdots,\tau_m^{(l)}|\tau_M) = \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{l-1} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta\left(\tau_M-\sum_{i=1}^{l}\tau_m^{(i)}\right)+\prod_{i=1}^{l}P_i(\tau_m^{(i)}|\tau_M)\theta\left(\tau_M-\sum_{i=1}^{l}\tau_m^{(i)}\right).
$$

Finally, because $\tau_M \gneqq \sum_{i=1}^{l}\tau_m^{(i)}$ holds for $l<n$ by the condition of Equation (26)

$$
P(\tau_m^{(1)},\cdots,\tau_m^{(l)}|\tau_M) = \prod_{i=1}^{l}P_i(\tau_m^{(i)}|\tau_M).
$$

Appendix F Derivation of Equation (39)

In this appendix, the approximation function of the inverse probability density function for the uncorrelated time series (Equation (39)) is derived. By substituting Equations (10), (15), and (34) into Equation (36)

$$
\begin{aligned}
&p_{Mm}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)}) \\
&= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}+1\right)}{\left(A_{\Delta m}\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}}\left[\prod_{i=1}^{n}\frac{e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left\{A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m^{(i)}}{\tau_M}\right)+2\right\}}{\frac{1}{\langle\tau_m\rangle}e^{-\frac{\tau_m^{(i)}}{\langle\tau_m\rangle}}\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}+1\right)}\right]\frac{1}{\langle\tau_M\rangle}e^{-\frac{\tau_M}{\langle\tau_M\rangle}} \\
&-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\frac{(n-1)}{\left(A_{\Delta m}\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}}\left[\prod_{i=1}^{n}\frac{\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}}{\frac{1}{\langle\tau_m\rangle}e^{-\frac{\tau_m^{(i)}}{\langle\tau_m\rangle}}}\right]\frac{1}{\langle\tau_M\rangle}e^{-\frac{\tau_M}{\langle\tau_M\rangle}},
\end{aligned}
\tag{F1}
$$

where the following relation is used.

$$
\begin{aligned}
\frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} &= \sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M) \\
&= A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}+1.
\end{aligned}
$$

The two products ($[\cdots]$'s) in Equation (F1) are respectively transformed as

$$\prod_{i=1}^{n} \frac{\left(\frac{A_{\Delta m}}{\langle \tau_M \rangle}\right) e^{-A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle}}}{\frac{1}{\langle \tau_m \rangle} e^{-\frac{\tau_m^{(i)}}{\langle \tau_m \rangle}}} = \left(A_{\Delta m} \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^n \prod_{i=1}^{n} e^{-A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle} + \frac{\tau_m^{(i)}}{\langle \tau_m \rangle}}$$
$$= \left(A_{\Delta m} \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^n \prod_{i=1}^{n} e^{\frac{\tau_m^{(i)}}{\langle \tau_M \rangle}}. \tag{F2}$$

$$\prod_{i=1}^{n} \frac{e^{-A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle}} \frac{A_{\Delta m}}{\langle \tau_M \rangle} \left\{A_{\Delta m} \frac{\tau_M}{\langle \tau_M \rangle} \left(1 - \frac{\tau_m^{(i)}}{\tau_M}\right) + 2\right\}}{\frac{1}{\langle \tau_m \rangle} e^{-\frac{\tau_m^{(i)}}{\langle \tau_m \rangle}} \left(A_{\Delta m} \frac{\tau_M}{\langle \tau_M \rangle} + 1\right)}$$
$$= \left(A_{\Delta m} \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^n \left(\prod_{i=1}^{n} e^{-A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle} + \frac{\tau_m^{(i)}}{\langle \tau_m \rangle}}\right) \left\{\prod_{i=1}^{n} \frac{A_{\Delta m} \frac{\tau_M}{\langle \tau_M \rangle} + 1 - \left(A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle} - 1\right)}{A_{\Delta m} \frac{\tau_M}{\langle \tau_M \rangle} + 1}\right\}$$
$$= \left(A_{\Delta m} \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^n \left(\prod_{i=1}^{n} e^{\frac{\tau_m^{(i)}}{\langle \tau_M \rangle}}\right) \left\{\prod_{i=1}^{n} \left(1 - \frac{\tau_m^{(i)} - \frac{\langle \tau_M \rangle}{A_{\Delta m}}}{\tau_M + \frac{\langle \tau_M \rangle}{A_{\Delta m}}}\right)\right\}. \tag{F3}$$

Finally, Equation (39) is derived by substituting Equations (F2) and (F3) into Equation (F1).

Appendix G Relation between the inverse probability density function and its approximation function in the uncorrelated time series

In this appendix, we discuss the relation between the inverse probability density function (Equation (23)) and Equation (36), i.e., the approximations made on Equation (23) that correspond to the assumptions made in §4.2 to derive Equation (36).

The summation in Equation (23) can be decomposed into

$$\sum_{i=n}^{\infty} (i - n + 1) \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i - j + 1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right)$$
$$= -(n-1) \sum_{i=n}^{\infty} \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i - j + 1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right)$$
$$+ \sum_{i=n}^{\infty} i \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i - j + 1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right). \tag{G1}$$

The first term on the r.h.s. of Equation (G1) is equivalent to $-(n-1) \prod_{i=1}^{n} P_i$ (Appendix E), and then, this term formally coincides with the correction term in Equation (36). Therefore, the second term corresponds to the kernel part (Equation (37)).

n-consecutive lower intervals must be included in only one upper interval. Under this condition, three constraints are imposed on the lower intervals, which appear on the l.h.s. of Equation (G1) as follows: (1) The number of lower intervals included in the upper interval must be larger than or equal to n. Then, the summation is taken in the range of $i \geq n$. (2) The way to choose the n-consecutive intervals from the i-lower intervals in an upper interval is only $(i - n + 1)$. If the first lower interval (or the leftmost one in the sequence of the consecutive lower intervals) is in either remaining $(n-1)$ ways, the sequence overflows from the upper interval. (3) The probability of the length of the j-th interval in the consecutive lower intervals depends on the way other (k-th, $1 \leq k <$

j) lower intervals appear, i.e., it is dependent on the remained time $\tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}$ and number of pieces of lower intervals $(i-j+1)$, $\rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right)$.

These constraints are relaxed in the derivation in §4.2. In the view in §4.2, the upper intervals of length τ_M are collected and the new time series is generated as shown in Figure 5. For this new time series, the only constraint imposed on the lower intervals is that they are included in the upper interval of length τ_M; each interval is assumed to occur independently. Therefore, the three constraints are changed in the following manner: (1) The new time series is generated by gathering all upper intervals of length τ_M, regardless of the number of lower intervals included in it. In addition, the restriction on the range of the summation ($i \geq n$) does not make much sense because the consecutive lower intervals are not assumed to be within only one upper interval, i.e., it is expanded to $i \geq 1$. (2) The number of ways to choose the n-consecutive intervals from i-lower intervals is unchanged; this exceeds the above mentioned upper limit ($i - n + 1$) although such cases are subtracted by the first term on the r.h.s. of Equation (G1), i.e., the correction term in Equation (36). (3) The constraints imposed on the condition in ρ_{mM} are removed; because the probability of the length of j-th interval is only not affected by other lower intervals, the temporal part of ρ_{mM} is replaced by τ_M (Equation (G2)). In addition, the constraint on the number of division can be eliminated by taking the average (Equation (G3)).

$$\rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right) \approx \rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M\right) \tag{G2}$$

$$\approx \frac{\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(j)}|i, \tau_M)}{\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M)} \tag{G3}$$

$$= p_{mM}(\tau_m^{(j)}|\tau_M).$$

Thus, ρ_{mM}'s are simply replaced by the conditional probability density functions.

In this way, the approximate view in §4.2 implies the following replacement in the exact inverse probability density function.

$$\begin{aligned}
&\sum_{i=n}^{\infty} i\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{n} \tau_m^{(k)}\right) \\
&\approx \sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M) \prod_{j=1}^{n} p_{mM}(\tau_m^{(j)}|\tau_M) \\
&= \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \prod_{j=1}^{n} p_{mM}(\tau_m^{(j)}|\tau_M).
\end{aligned}$$

Appendix H Distance between the inverse probability density function and the inter-event time distribution

In this Appendix, we derive the distance between the inverse probability density function (Equation (24)) and the inter-event time distribution ($p_M(\tau_M)$)

$$D(p_{Mm}||p_M) := \int_T^{\infty} |p_\theta(\tau_M, T) - p_M(\tau_M)|^2 \, d\tau_M, \tag{H1}$$

where

$$\begin{aligned}
p_\theta(\tau_M, T) &= \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle^2}\left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right) e^{-\frac{\tau_M - T}{\langle \tau_M \rangle}} \left(A_{\Delta m}\frac{\tau_M - T}{\langle \tau_M \rangle} + 2\right), \\
p_M(\tau_M) &= \frac{1}{\langle \tau_M \rangle} e^{-\frac{\tau_M}{\langle \tau_M \rangle}}.
\end{aligned}$$

By substituting these functions in Equation (H1), the distance is derived as

$$D(p_{Mm}||p_M) = \frac{\langle\tau_M\rangle C_1^2}{4} + \frac{C_1 C_2(T)}{2} + \frac{C_2(T)^2}{2\langle\tau_M\rangle}, \tag{H2}$$

where

$$C_1 = \frac{1}{\langle\tau_M\rangle}\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^2,$$
$$C_2(T) = 2\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) - e^{-\frac{T}{\langle\tau_M\rangle}}.$$

Appendix I On the cause of the separation of $\langle D'(p_{Mm}^{\mathrm{approx}}||p_M)\rangle$ and $\langle D''(p_{Mm}^{\mathrm{approx}}||p_M)\rangle$ at large T

In this appendix, we examine the cause of the separation between $\langle D'(p_{Mm}^{\mathrm{approx}}||p_M)\rangle$ and $\langle D''(p_{Mm}^{\mathrm{approx}}||p_M)\rangle$ at long-elapsed time T. Let us compare Figure 7 to Figures S6(a) and (c) for $\mathcal{N} = 10^5$. The separation is suppressed compared to that shown in Figure 7, which indicates that the fluctuations in the spline functions of $\boldsymbol{P}_1$ caused by a relatively small number of samples in the calculation of $\boldsymbol{P}_1$ are suppressed by increasing the sample data. This leads to the reduction of errors in the calculations (44), and to the improvement of the calculation of distance in Equation (46).

In addition, we tested numerical updating with $\mathcal{N} = 10^5$ by excluding some larger columns of the matrix $\boldsymbol{P}_1$, i.e., by using the following matrix $\boldsymbol{P}_1'$ with an integer l_c

$$\boldsymbol{P}_1' = \left[[P_{1,jk}]_{j=j_{\min}^{(k)},\cdots,j_{\max}^{(k)}}\right]_{k=k_{\min},\cdots,k_{\max}-l_c}. \tag{I1}$$

For this $\boldsymbol{P}_1'$, the interpolation and extrapolation procedures are conducted in the same way as in §5.1, and the numerical updating is executed.

Figures S6(b) and (d) show the results of the distance for such updating with (b) $\Delta\tau_M = 0.1$ and $l_c = 5$, and (d) $\Delta\tau_M = 0.025$ and $l_c = 20$. Compared to the results obtained using $\boldsymbol{P}_1$ in Figures S6(a) and (c), the separation is suppressed further. Combined with the results for the kernel part, these results suggest the following; the number of samples to calculate $\boldsymbol{P}_1$ is so small compared to that of the conditional probability (the number of sample is only one for an upper interval for $\boldsymbol{P}_1$ whereas all the lower intervals included in an upper interval are used as a sample to calculate the conditional probability), in particular for a large k, that its fluctuation becomes too large to compute the correction term precisely.

Acknowledgments

The idea of generalized probability density functions and the simplified relation (7) are by Y. Aizawa. We would like to thank Editage (www.editage.com) for English language editing. This work was supported by JST SPRING, Grant Number JPMJSP2110.

References

Aizawa, Y., Hasumi, T., & Tsugawa, S. (2013). Seismic statistics. *Nonlinear Phenomena in Complex Systems*, *16*(2), 116–130.

Aizawa, Y., & Tsugawa, S. (2014). Aftershock cascade of the 3.11 earthquake (2011) in fukushima-miyagi area. In D. Matrasulov & H. E. Stanley (Eds.), *Nonlinear phenomena in complex systems: From nano to macro scale* (pp. 21–33). Dordrecht: Springer Netherlands.

Bak, P., Christensen, K., Danon, L., & Scanlon, T. (2002). Unified scaling law for earthquakes. *Physical Review Letters*, *88*(17), 178501.

Bottiglieri, M., de Arcangelis, L., Godano, C., & Lippiello, E. (2010). Multiple-time scaling and universal behavior of the earthquake interevent time distribution. *Physical Review Letters*, *104*(15), 158501.

Corral, A. (2004). Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. *Physical Review Letters*, *92*(10), 108501.

Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in california. *Bulletin of the Seismological Society of America*, *34*(4), 185–188.

Helmstetter, A., & Sornette, D. (2002). Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research: Solid Earth*, *107*(B10), ESE 10-1-ESE 10-21. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001JB001580` doi: https://doi.org/10.1029/2001JB001580

Helmstetter, A., & Sornette, D. (2003). Importance of direct and indirect triggered seismicity in the etas model of seismicity. *Geophysical Research Letters*, *30*(11). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003GL017670` doi: https://doi.org/10.1029/2003GL017670

Lippiello, E., Corral, Á., Bottiglieri, M., Godano, C., & Arcangelis, L. (2012). Scaling behavior of the earthquake intertime distribution: Influence of large shockss and time scales in the omori law. *Physical Review E*, *86*(6), 066119.

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

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*(2), 379–402.

Omori, F. (1894). On aftershocks of earthquakes. *The journal of the College of Science, Imperial University of Tokyo, Japan*, *7*, 111–200.

Saichev, A., & Sornette, D. (2006). "universal" distribution of interearthquake times explained. *Phys. Rev. Lett.*, *97*, 078501. Retrieved from `https://link.aps.org/doi/10.1103/PhysRevLett.97.078501` doi: 10.1103/PhysRevLett.97.078501

Scholz, C. H. (2002). *The mechanics of earthquakes and faulting.* Cambridge university press.

Tanaka, H., & Aizawa, Y. (2017). Detailed analysis of the interoccurrence time statistics in seismic activity. *Journal of the Physical Society of Japan*, *86*(2), 024004.

Tanaka, H., & Umeno, K. (2021). Bayesian updating for time-intervals of different magnitude thresholds in marked point process and its application to time-series of ETAS model. *ESS Open Archive*. Retrieved from `https://doi.org/10.1002%2Fessoar.10509635.1` doi: 10.1002/essoar.10509635.1

Touati, S., Naylor, M., & Main, I. G. (2009). Origin and nonuniversality of the earthquake interevent time distribution. *Physical Review Letters*, *102*(16), 168501.

Utsu, T. (1961). A statistical study on the occurrence of aftershocks. *Geophysical Magazine*, *30*, 521–605.

Webb, D. J. (1974). The statistics of relative abundance and diversity. *Journal of Theoretical Biology*, *43*(2), 277–291.

Supporting Information for "Bayesian updating on time intervals at different magnitude thresholds in a marked point process and its application to synthetic seismic activity"

Hiroki Tanaka [1] and Ken Umeno [1]

[1]Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto-shi, JAPAN

Contents of this file

1. Figures S1 to S17

Introduction The following figures are provided in this supporting information. Figure S1–S7 present information regarding the numerically generated uncorrelated time series. Figure S1 shows the probability density functions of the length of the inter-event times ($p_m(\tau_m)$ and $p_M(\tau_M)$) and their interpolation and extrapolation functions. Figure S2 shows the conditional probability density function of the length of the lower intervals included in the upper interval of length τ_M ($p_{mM}(\tau_m|\tau_M)$) and its interpolation and extrapolation functions. Figure S3 shows the conditional probability density function of the length of the left-most lower interval included in the upper interval of length τ_M ($P_1(\tau_m|\tau_M)$) and its interpolation and extrapolation functions. Figure S4 shows the average number of lower intervals included in the upper interval of length τ_M ($\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M}$) and its interpolation and extrapolation functions. Figure S5 shows an example of Bayesian updating. Figure S6 shows the average distances between the distribution functions and the (part of) approximation functions including the results of numerical Bayesian updating using statistical amounts calculated using the $\mathcal{N} = 10^5$ time series. Figure S7 shows the joint probability mass function of the logarithm of the positions of the maximum peak of the inverse probability density function and its approximation function calculated by numerical Bayesian updating with $\mathcal{N} = 10^5$. Figures S8–S17 are about the numerically generated time series of the ETAS model. Figure S8 shows $p_m(\tau_m)$ and $p_M(\tau_M)$, and their interpolation and extrapolation functions. Figure S9 shows $p_{mM}(\tau_m|\tau_M)$ and its interpolation and extrapolation functions. Figure S10 and S11 show the conditional probability density functions of the left-most and right-most lower intervals included in the upper interval of length τ_M ($p^{\mathrm{L}}(\tau_m|\tau_M)$ and $p^{\mathrm{R}}(\tau_m|\tau_M)$, respectively) and their interpolation and extrapolation functions. Figure S12 shows $\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M}$ and its interpolation and extrapolation functions. Figure S13–S15 show examples of Bayesian updating for $\delta_{\mathrm{th}} = 0.5$. Figure S16 shows the probability density functions of $n_{\leq \mathrm{th}}$ and $\tau_{\leq \mathrm{th}}$ for $\delta_{\mathrm{th}} = 0.5$. Figure S17 shows the statistical results of the effectiveness of forecasting for $\delta_{\mathrm{th}} = 0.25$.

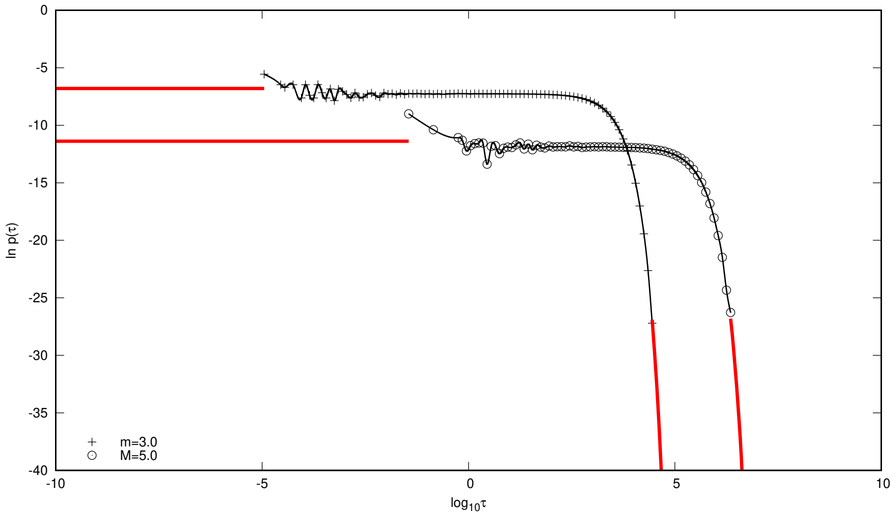

Figure S1. Probability density functions of the length of inter-event times at magnitude thresholds $m = 3.0$ and $M = 5.0$ calculated for the numerically generated uncorrelated time series, and their interpolation and extrapolation functions. Symbols ($+$ and $\odot$) show the probability density functions obtained numerically. The intervals between the symbols are interpolated by cubic spline functions represented by the black curves. Further, outside of the range covered by the symbols are extrapolated by the fitting functions at the edge represented by the red lines; constant function ($\ln p(\tau) \equiv C$) on the small side and the exponential function ($\ln p(\tau) = A\tau + B$) on the large side. The parameter values (A, B, C) are determined by the least squares method using 10 points at each end.

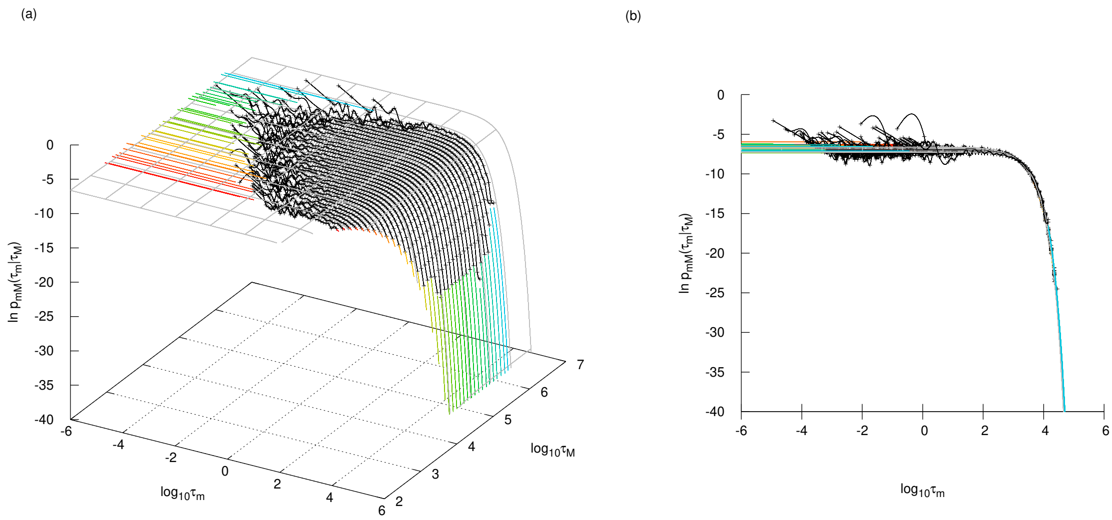

Figure S2. Conditional probability density function of the length of the lower interval (τ_m) included in the upper interval of length τ_M for the uncorrelated time series, and its interpolation and extrapolation functions in (a) an oblique view and (b) a horizontal view parallel to the $\log_{10}\tau_M$-axis. Gray curved surface is the part of the step function of Equation (15). Symbols (+) represent the probability density function obtained numerically. For each τ_M, only data points with 30 or more points in the range of $\tau_M > \tau_m$ are displayed. The intervals between the symbols in the $\log_{10}\tau_m$-axis direction are interpolated by cubic spline functions represented by the black curves. Outsides of the range covered by the symbols in the $\log_{10}\tau_m$-axis direction are extrapolated by the fitting functions at the edge represented by the colored lines (the color varies with $\log_{10}\tau_M$); constant function ($\ln p(\tau) \equiv C$) on the small τ_m side and the exponential function ($\ln p(\tau) = A\tau + B$) on the large τ_m side for each τ_M. The parameter values (A, B, C) are determined by the least squares method using 10 points at each end for each τ_M.

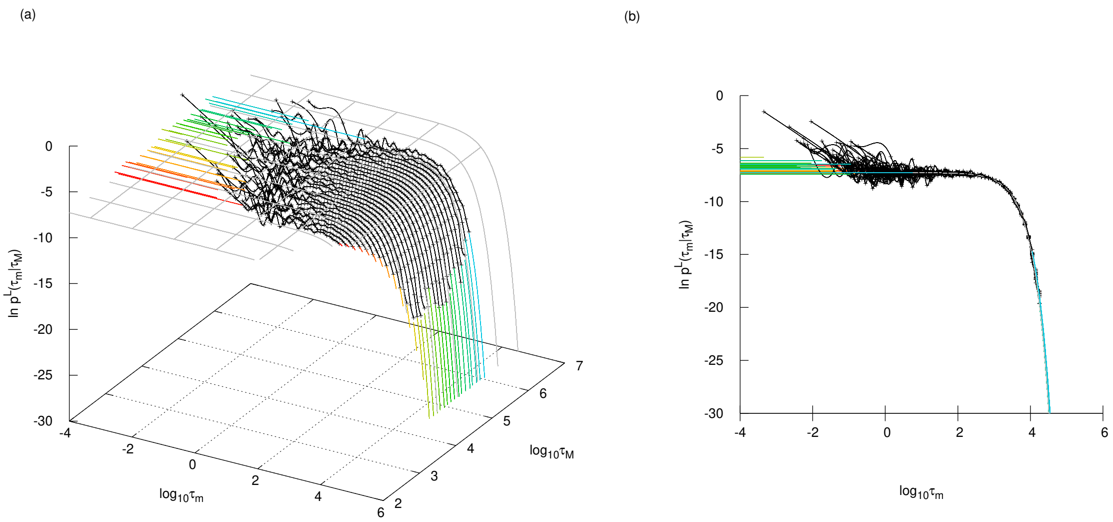

Figure S3. Conditional probability density function of the length of the left-most lower interval included in the upper interval of length τ_M for the uncorrelated time series, and its interpolation and extrapolation functions in (a) an oblique view and (b) a horizontal view parallel to the $\log_{10}\tau_M$-axis. Gray curved surface shows the part of the step function of Equation (29). The description of the figure is the same as in Figure S2.

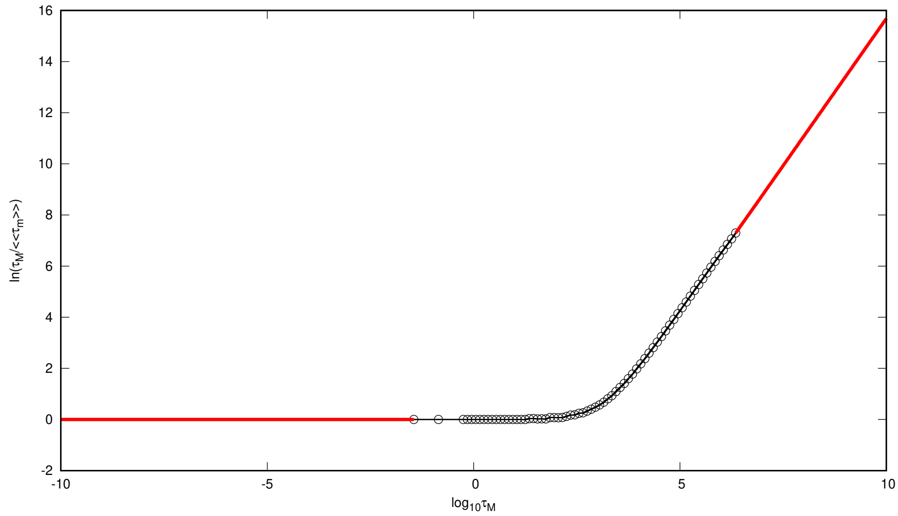

Figure S4. Average number of lower intervals included in the upper interval of length τ_M ($\tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M}$) for the uncorrelated time series, and its interpolation and extrapolation functions. Symbols ($\odot$) indicate the results obtained numerically. The intervals between the symbols are interpolated by cubic spline functions represented by the black curves. Outsides of the range covered by the symbols are replaced or extrapolated by following functions represented by red lines; $\ln \tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M} \equiv 0$ on the small side, whereas the fitting function $\tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M} = A\tau_M + B$ on the large side with the parameter values determined by the least squares method using 10 points at the end.

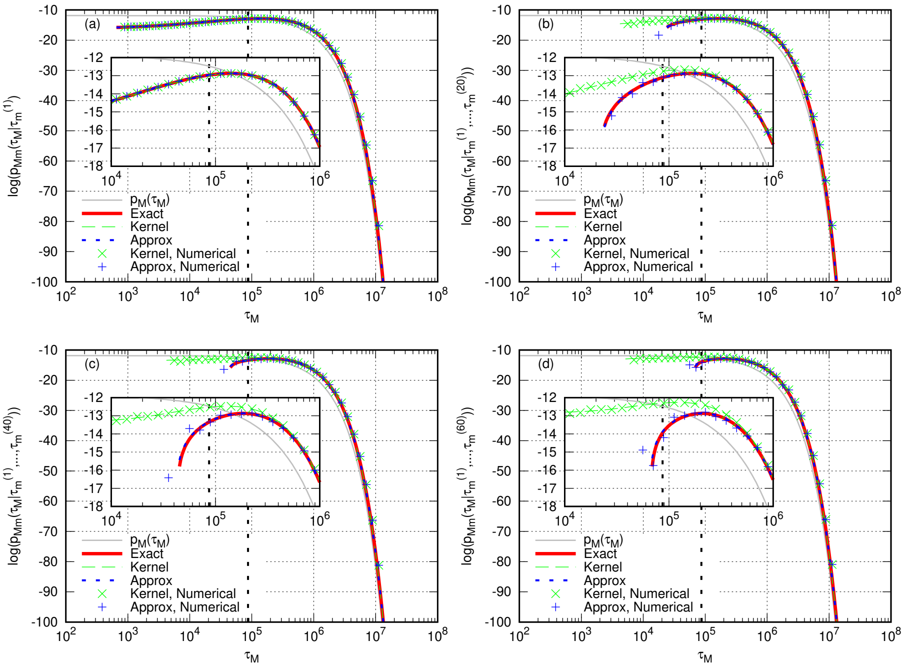

Figure S5. Examples of the Bayesian updating on a numerically generated uncorrelated time series. In this example, the total number of updates between the events with a magnitude above M is 73. The horizontal axis represents the elapsed time from the previous event with a magnitude above M, and the vertical axis is the logarithmic scale of the inverse probability density function and its approximation function for n = (a) 1, (b) 20, (c) 40, and (d) 60, respectively. Gray curve is $p_M(\tau_M)$ and black vertical dotted line indicates the actual elapsed time of the next large event with a magnitude above M. At each update, (Exact) the inverse probability density function (Equation (24)), (Approx) the approximation function (Equation (39)), and (Kernel) its kernel part (Equation (40)) are shown. Further, the results with the numerical updating method (Equations (43) and (44)) are shown for the approximation function and its kernel part; (Kernel, Numerical) represents the result of Equation (43), and (Approx, Numerical) is for Equations (43) and (44).

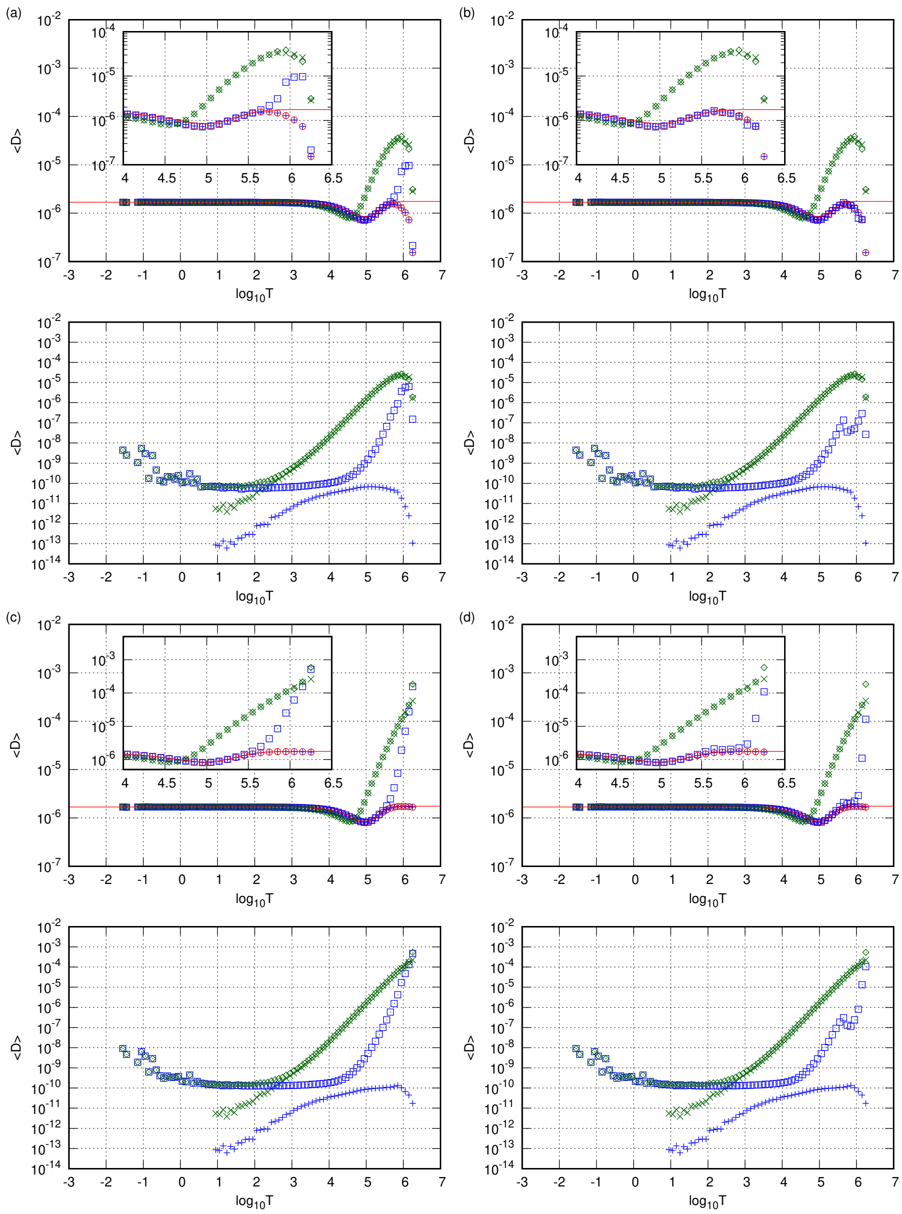

Figure S6. Average distances for elapsed time T from the last event greater than M. Each symbol represents the same distance as in Figure 7. The results of the numerical updating of $\langle D'' \rangle$ are calculated using statistical amounts in Equation (42) taken from the $\mathcal{N} = 10^5$ time series with (a) $\Delta\tau_M = 0.1$ and $l_c = 0$, (b) $\Delta\tau_M = 0.1$ and $l_c = 5$, (c) $\Delta\tau_M = 0.025$ and $l_c = 0$, and (d) $\Delta\tau_M = 0.025$ and $l_c = 20$.

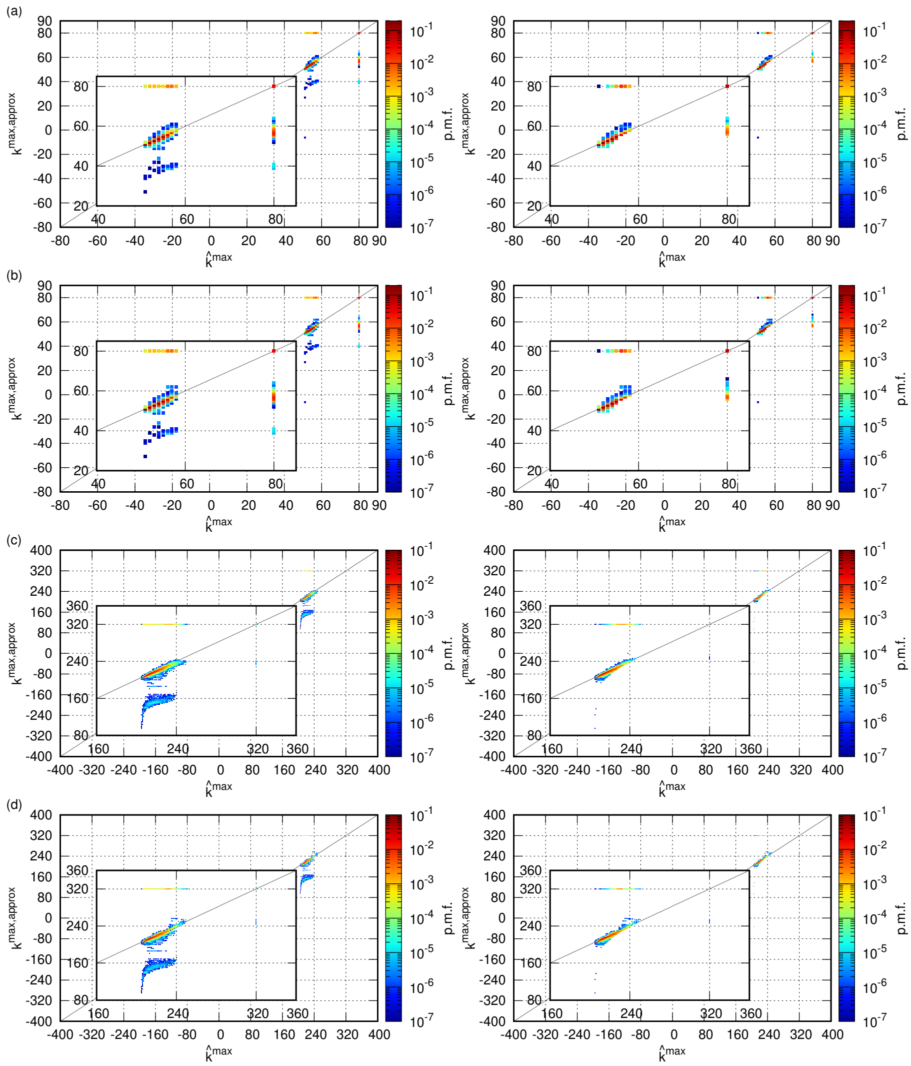

Figure S7. Joint probability mass function of $(\hat{k}^{\max}, k^{\max,\text{approx}})$. $k^{\max,\text{approx}}$ is obtained using the numerical updating method with statistical amounts in Equation (42) taken from the $\mathcal{N} = 10^5$ time series with (a) $\Delta\tau_M = 0.1$ and $l_c = 0$, (b) $\Delta\tau_M = 0.1$ and $l_c = 5$, (c) $\Delta\tau_M = 0.025$ and $l_c = 0$, and (d) $\Delta\tau_M = 0.025$ and $l_c = 20$. For (a) and (b), the horizontal lines at $k^{\max,\text{approx}} = 80$ and the vertical line at $\hat{k}^{\max} = 80$ correspond to cases when no peak is detected by the peak search. Further, for (c) and (d), the lines at $k^{\max,\text{approx}} = 320$ and at $\hat{k}^{\max} = 320$ correspond to the no peak cases. The left panels are results when the peak search is conducted in the range of $\tau_M > \max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ and the right panels in the range of $\tau_M > T$.

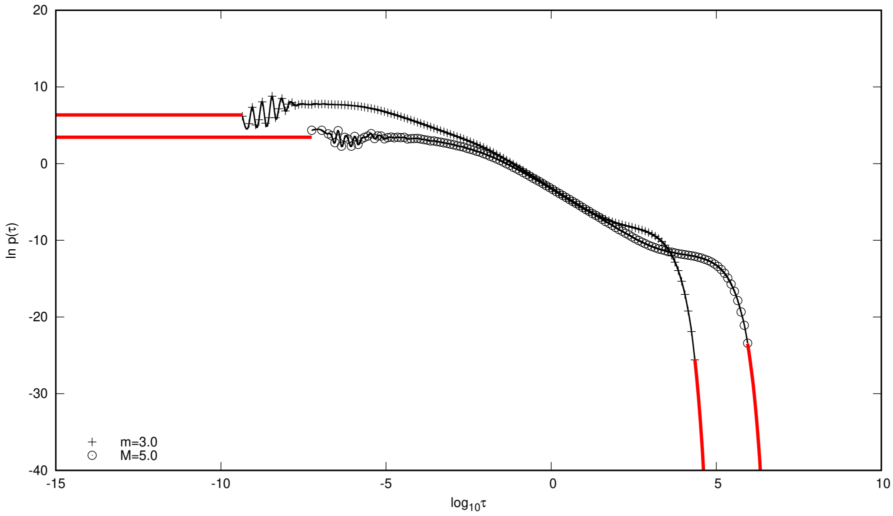

Figure S8. Probability density functions of the length of inter-event times at magnitude thresholds $m = 3.0$ and $M = 5.0$ calculated for the numerically generated time series of the ETAS model, and their interpolation and extrapolation functions. The description of the figure is the same as in Figure S1.

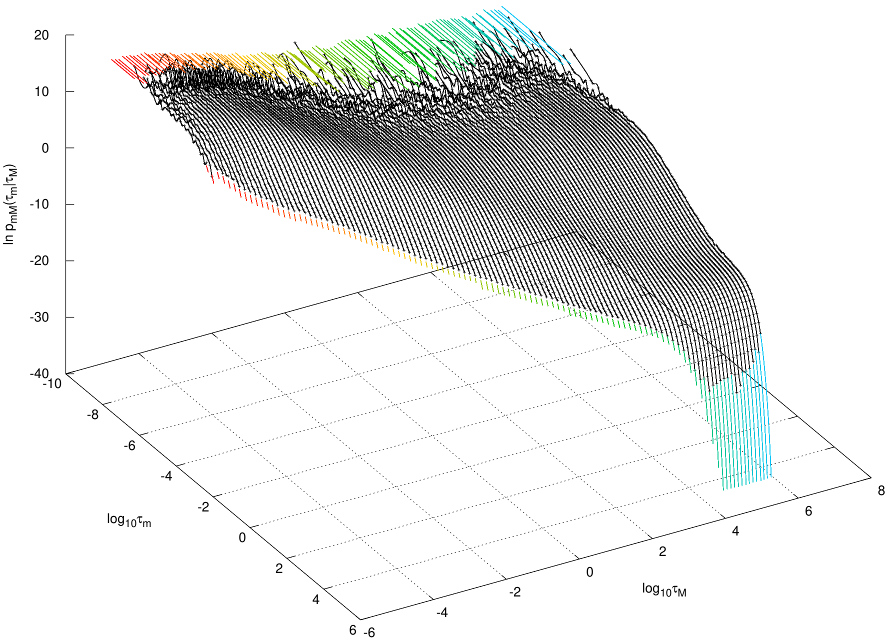

Figure S9. Conditional probability density function of the length of the lower interval (τ_m) included in the upper interval of length τ_M for the numerically generated time series of the ETAS model, and its interpolation and extrapolation functions. The description of the figure is the same as in Figure S2.

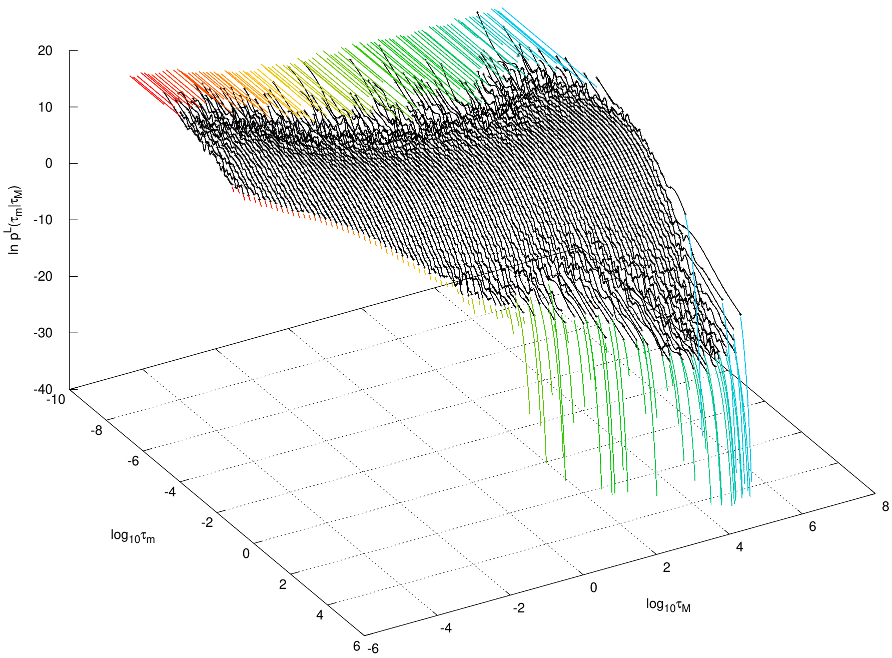

Figure S10. Conditional probability density function of the length of the left-most lower interval included in the upper interval of length τ_M for the numerically generated time series of the ETAS model, and its interpolation and extrapolation functions. The description of the figure is the same as in Figure S3.

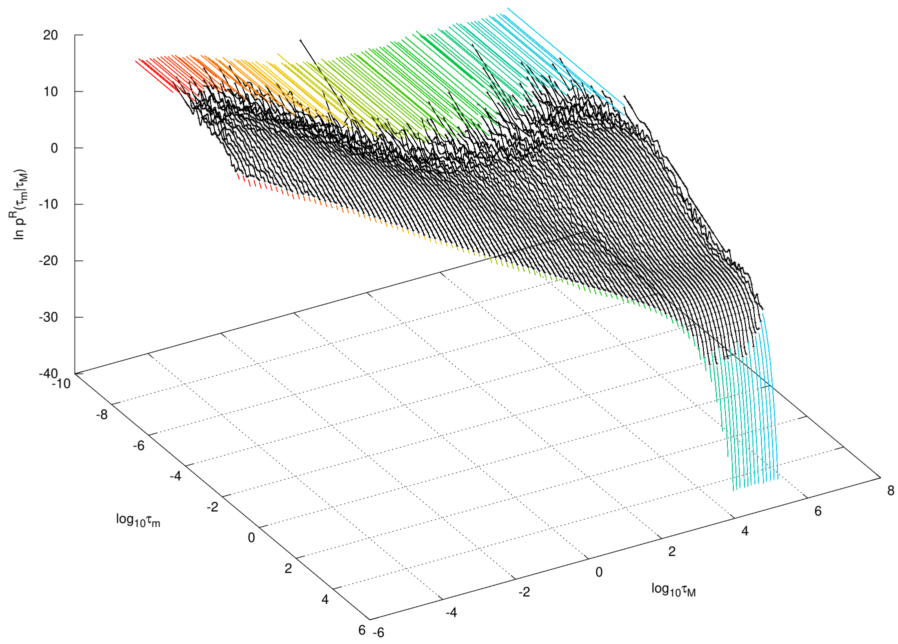

Figure S11. Conditional probability density function of the length of the right-most lower interval included in the upper interval of length τ_M for the numerically generated time series of the ETAS model, and its interpolation and extrapolation functions. The description of the figure is the same as in Figure S3.

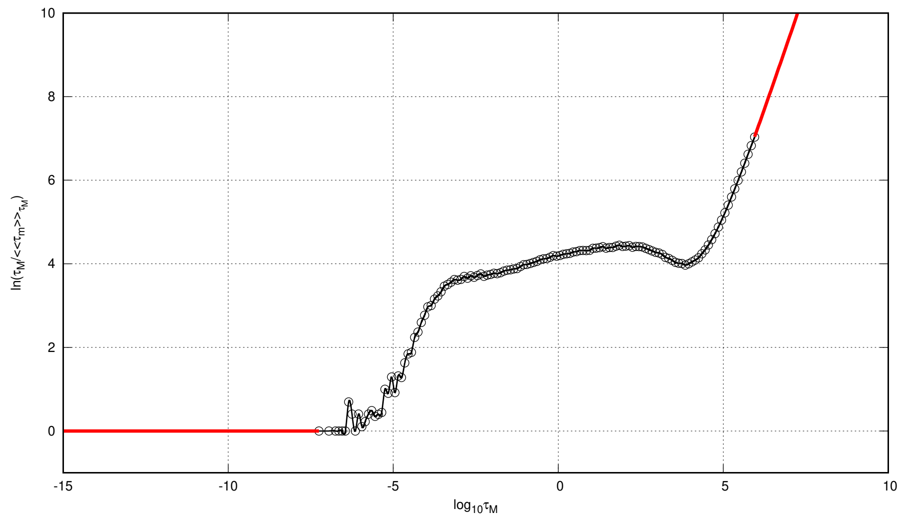

Figure S12. Average number of lower intervals included in the upper interval of length τ_M ($\tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M}$) for the numerically generated time series of the ETAS model. The description of the figure is the same as in Figure S4.

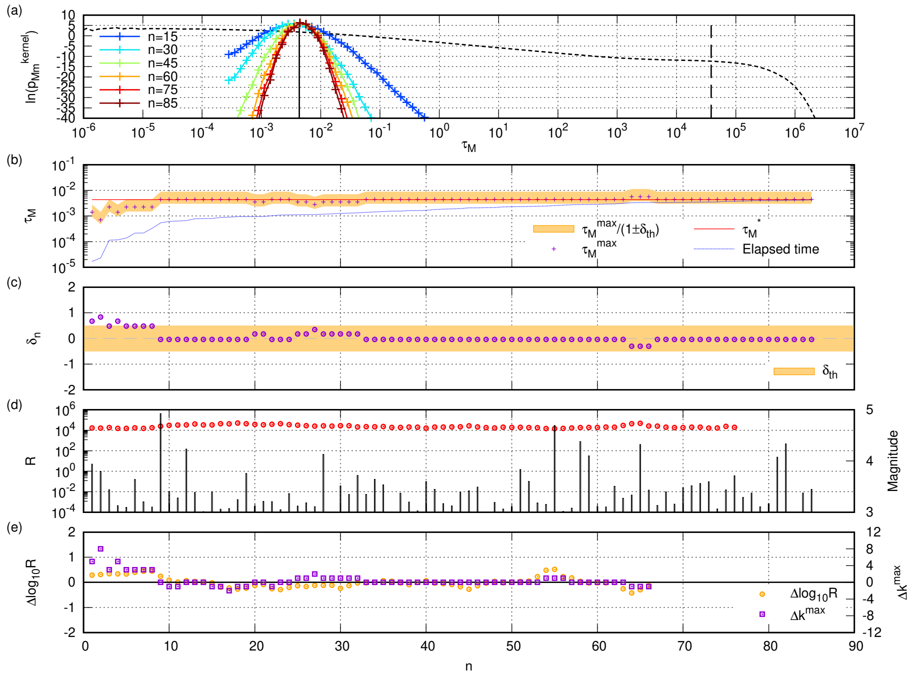

Figure S13. First example of Bayesian updating. This is the case where τ_M^* is in regime (I). (a) Around the peak of the kernel part for each update. The dotted curve indicates $p_M(\tau_M)$. (b) Evolutions of the estimate ($\tau_M^{\mathrm{max,n}}$) and the tolerance of error $\left[\frac{\tau_M^{\mathrm{max,n}}}{1+\delta_{\mathrm{th}}}, \frac{\tau_M^{\mathrm{max,n}}}{1-\delta_{\mathrm{th}}}\right]$ in Equation (50) with $\delta_{\mathrm{th}} = 0.5$. The elapsed time from the last large event is indicated by the dotted line. (c) Evolution of the relative error (δ_n). The orange band indicates the tolerance range $\pm\delta_{\mathrm{th}}$. (d) Evolution of the occurrence rate (R_n defined by Equation (51)). The magnitude of the event at each update is indicated by black bars. (e) Evolutions of the variation of the log-occurrence rate ($\Delta \log_{10} R_n$ defined by Equation (52)) and the variation of the log-estimate ($\Delta k_n^{\mathrm{max}}$ defined by Equation (53)). In this example, the occurrence rate is high, and it stays almost constant as shown in (d). The kernel part has a peak as shown in (a). Its maximum peak time ($\tau_M^{\mathrm{max,n}}$) continues to be nearly constant around τ_M^* from well before the large event as shown in (b); this is confirmed in (c), which indicates that $|\delta_n| \leq \delta_{\mathrm{th}}$ is satisfied consecutively for $n \in (n_{\mathrm{fin}} - n_{\leq \mathrm{th}}, n_{\mathrm{fin}}]$ with a long $n_{\leq \mathrm{th}}$, and in (e), that shows that $\Delta k_n^{\mathrm{max}}$ fluctuate around 0. Then, in this example, τ_M^* is judged to be effectively forecasted for the setting of $\delta_{\mathrm{th}} = 0.5$.

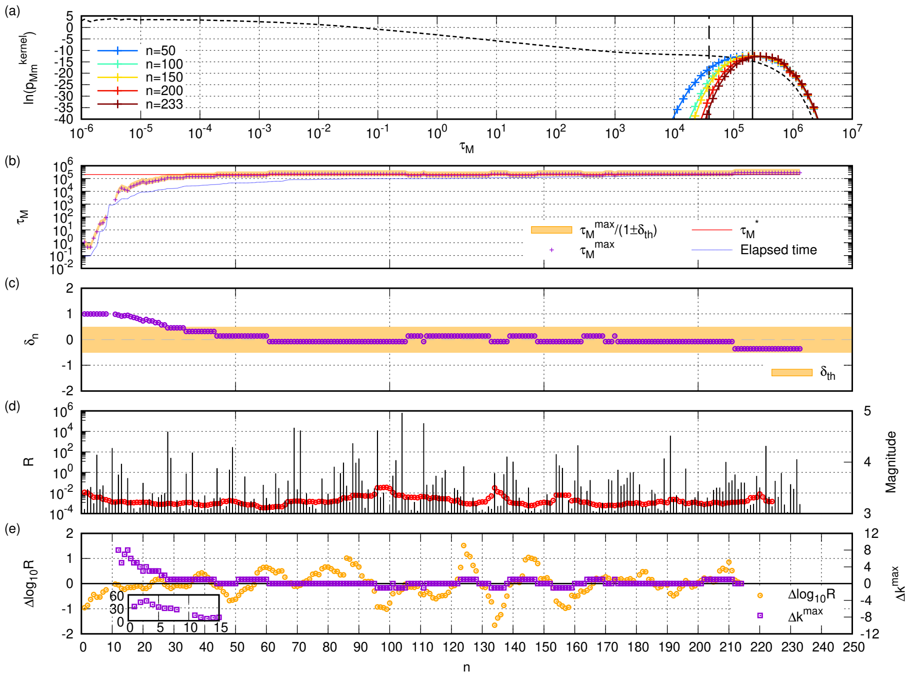

Figure S14. Second example of Bayesian updating. This is the case that τ_M^* is in regime (Ⅲ). In some cases, the kernel part does not have the maximum peak, and the estimate is not determined, which causes some jumps in the time series. The inset in (e) shows $\Delta k_n^{\max}$ for small number of updates, indicating a rapid variation of $k^{\max,n}$ at small n. Other descriptions of the figure is the same as in Figure S13. In this example, the occurrence rate is low and keeps almost constant around $\lambda_0 = 0.0007$, as shown in (d). (b) and (c) show that the maximum peak time of the kernel part ($\tau_M^{\max,n}$) transitions to around τ_M^*, and it consecutively satisfies $|\delta_n| \leq \delta_{\rm th} = 0.5$ from long to immediately before the next large event. Further, this is also confirmed by $\Delta k_n^{\max} \approx 0$ in (e). Therefore, in this example, the forecasting is judged to be conducted effectively.

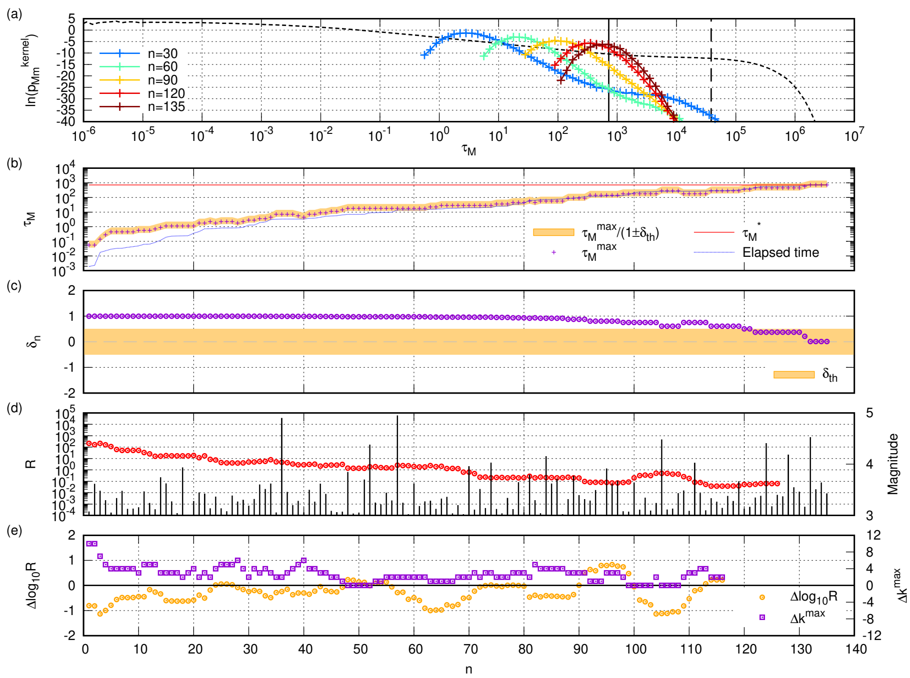

Figure S15. Third example of Bayesian updating. This is the case that τ_M^* is in regime (II). The descriptions of the figure are the same as in Figure S13. In this example, unlike in Figures S13 and S14, the time series is dominated by the non-stationary activity as shown in (d) and (e). Although $|\delta_n| \le \delta_{\rm th} = 0.5$ is satisfied only immediately before the large shock, $\tau_M^{\rm max,n}$ continues shifting and $|\delta_n| \le \delta_{\rm th}$ does not hold as shown in (b), (c), and (e). Therefore, the forecasting is not effective in this case.

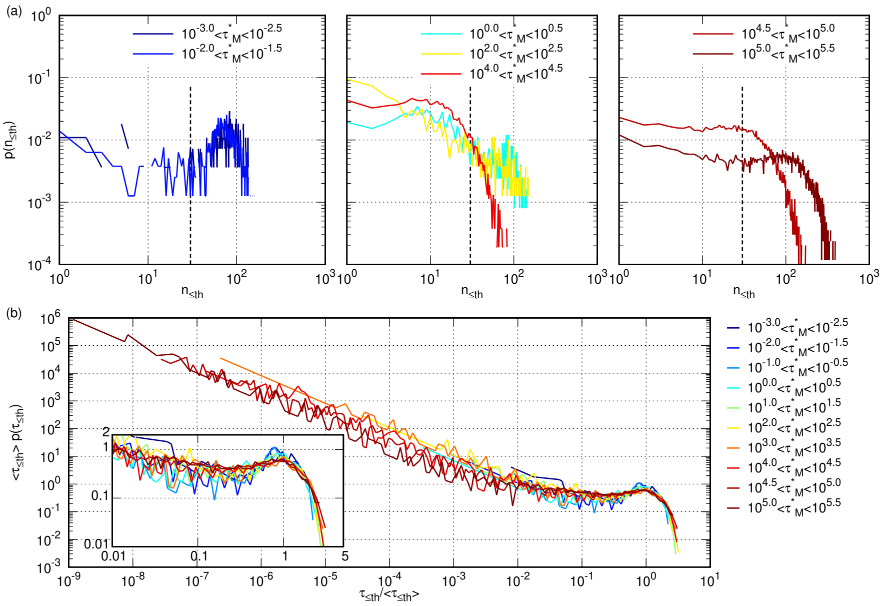

Figure S16. (a) Probability distributions ($p(n_{\leq \mathrm{th}})$) of the number of consecutive updates ($n_{\leq \mathrm{th}}$) with $|\delta_n| \leq \delta_{\mathrm{th}} = 0.5$ for each τ^*_M. Only the cases with $n_{\leq \mathrm{th}} > 0$ are included in the population. The vertical dotted line indicates $n_{\leq \mathrm{th}} = 30$. (b) Probability density function ($p(\tau_{\leq \mathrm{th}})$) of the duration time ($\tau_{\leq \mathrm{th}}$) during n_{th}-updates with $\delta_{\mathrm{th}} = 0.5$ for each τ^*_M. The distributions rescaled by the averages ($\langle \tau_{\leq \mathrm{th}} \rangle := \int_0^\infty \tau_{\leq \mathrm{th}} p(\tau_{\leq \mathrm{th}}) d\tau_{\leq \mathrm{th}}$) are shown.

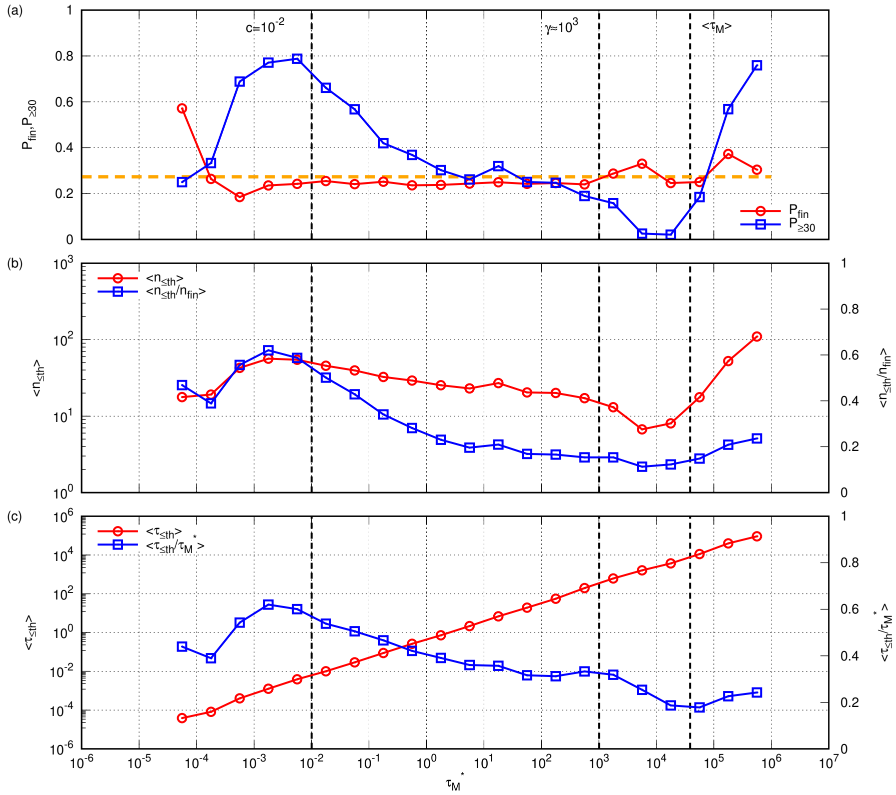

Figure S17. Statistical results for each τ_M^* with $\delta_{\rm th} = 0.25$. (a) $P_{\rm fin}$, $P_{\geq 30}$, and the average of $P_{\rm fin}(\approx 0.27)$ for the overall τ_M^*; (b) $\langle n_{\leq \rm th}\rangle$ and $\langle n_{\leq \rm th}/n_{\rm fin}\rangle$; and (c) $\langle \tau_{\leq \rm th}\rangle$ and $\langle \tau_{\leq \rm th}/\tau_M^*\rangle$ are shown.