Journal of the Physical Society of Japan **93**, 024001 (2024)
https://doi.org/10.7566/JPSJ.93.024001

Bayesian Updating on Time Intervals at Different Magnitude Thresholds in a Marked Point Process and Its Application to Synthetic Seismic Activity

Hiroki Tanaka* and Ken Umeno†

Graduate School of Informatics, Kyoto University, Sakyo, Kyoto 606-8501, Japan

(Received January 17, 2023; accepted November 9, 2023; published online December 25, 2023)

We present a Bayesian updating method on the inter-event time intervals at different magnitude thresholds in a marked point process toward probabilistic forecasting of an upcoming large event using temporal information on smaller events. Bayes' theorem 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. We derive the inverse probability density function and the condition under which it has a peak. Further, we derive its approximation function that consists of the kernel part that includes the product of conditional probabilities and a correction term. Bayesian updating is applied numerically to the time series of a seismic activity model. We estimate the time of an upcoming large event using the maximum point of the kernel part and evaluate its accuracy by the relative error with the actual occurrence time. We further evaluate the forecasting effectiveness by the continuity of updates with acceptable accuracy before the large event. Statistical results indicate that forecasting is relatively effective immediately or long after the last major event in which stationarity dominates in time series.

1. Introduction

The probabilistic forecasting of the timing of future major earthquakes is important for seismic risk assessment. Such forecasting is based on earthquakes' magnitude and time information that can be illustrated by a marked point process schematically shown in Fig. 1. A basic approach involves using the hazard rate, which is calculated using the inter-event time distribution[1,2] extracted from the marked point process. The inter-event time distribution is defined as a probability density function of a length 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 has been studied in seismic activity not only for risk assessment[2] but also for advancing the understanding of statistical properties in seismic activity from several perspectives. Such studies encompass the unification of seismic laws[3–10] that include the Gutenberg–Richter (GR)[11] and Omori–Utsu[12,13] laws, and the scaling universality in inter-event time distributions[5,14] that were critically examined[15–20] and further extended[21–23] with the help of the Epidemic Type Aftershock Sequence (ETAS) model.[24–29]

The ETAS model is capable of generating an inhomogeneous marked Poisson process like seismic activity:[24,30–33] the event time t is stochastically determined by the occurrence rate [$\lambda(t)$] that is represented by the combination of the Omori–Utsu and Utsu–Seki[34,35] laws[24–29] as,[24,30]

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

where $\{t_j, M_j\}$ ($j \in \mathbb{N}$) are the time and magnitude of events before time t; while the magnitude is generated randomly and independently obeying the GR law, $P(M)$ (probability density of magnitude at M) $\propto 10^{-bM}$.[24] M_0 represents the minimum magnitude and $(\lambda_0, K, \alpha, c, \theta, b)$ the parameters that characterize the activity; in particular, λ_0 represents the constant rate for background seismicity.[24,30] The combination of the remaining parameters yields the branching ratio $n_{\mathrm{br}} = \frac{K}{\theta c^\theta}\frac{b}{b-\alpha}$ (when $\theta > 0$)[30] that determines the stationarity of the time series as well as the ratio of aftershocks generated by mainshocks.[36]

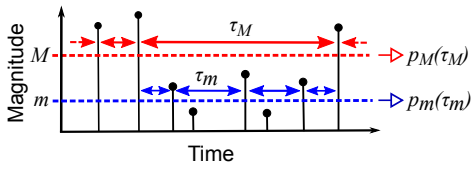

Fig. 1. (Color online) Schematic of a marked point process to represent the time and magnitude for each event in seismic activity (after Refs. 38 and 39).

The ETAS model provides a standard seismicity for detecting anomalous activity.[24,25] This model has been extended to spatio-temporal versions[26,28] and further hierarchical space-time version.[27,29] The conditional intensity function provides the risk of an event at a given time and space based on the history of seismic activity,[29,37] 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.[38]

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 a lower interval length (τ_m) provided that it is inside the upper interval of length τ_M (Fig. 1).[38] Inter-event time distributions at magnitude thresholds m and M are

connected by an integral equation, the kernel of which includes this conditional probability density function.[38] This integral equation is given as

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

where N_m and N_M represent the total number of intervals at magnitude thresholds m and M, respectively, and $\langle\langle \tau_m \rangle\rangle_{\tau_M}$ represents the average of the conditional probability density function, $\langle\langle \tau_m \rangle\rangle_{\tau_M} := \int_0^{\infty} \tau_m p_{mM}(\tau_m|\tau_M)\, d\tau_m$.[38]

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 Bayes' theorem and Bayesian updating on the intervals at different magnitude thresholds in this paper; further, we report the results of the numerical analysis related to the inverse probability density function.[39]

In Sect. 2, we derive Bayes' theorem for intervals at different magnitude thresholds in the marked point process. In Sect. 3, the inverse probability density function is derived for the uncorrelated time series that corresponds to the background seismicity of the ETAS model. In Sect. 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 Sect. 5. In Sect. 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, Sect. 7 presents additional discussions and conclusions. Variables and constants that emerge in this paper are summarized in Table I.

2. Bayes' Theorem for Inter-event Times at Different Magnitude Thresholds

We consider 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 $p_{Mm}(\tau_M|\tau_m)$. Here $p_{Mm}(\tau_M|\tau_m)$ represents the probability density function of the upper interval length under the condition that it includes a lower interval of length τ_m.

Let $N_{mM}(\tau_M, \tau_m)$ be 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 $[\tau_m, \tau_m + d\tau_m)$ (Fig. 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:

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 [Fig. 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 $p_{Mm}(\tau_M|\tau_m)\, d\tau_M$. Therefore

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

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 [Fig. 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

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

From Eqs. (3) and (4)

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

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

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

Equation (6) can be considered as 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

$$z_m(\tau_m) := \frac{\tau_m}{\langle \tau_m \rangle} p_m(\tau_m),$$

$$z_M(\tau_M) := \frac{\tau_M}{\langle \tau_M \rangle} p_M(\tau_M), \quad (7)$$

$$z_{mM}(\tau_m|\tau_M) := \frac{\tau_m}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M).$$

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

$$z_m(\tau_m) = \int_0^{\infty} z_{mM}(\tau_m|\tau_M) z_M(\tau_M)\, d\tau_M, \quad (8)$$

$$p_{Mm}(\tau_M|\tau_m) = \frac{z_{mM}(\tau_m|\tau_M) z_M(\tau_M)}{z_m(\tau_m)}. \quad (9)$$

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

Table I. Variables and constants used in this paper.

Symbols	Meaning
m, M	Lower (m) and upper (M) magnitude thresholds for an marked point process.
Δm	$:= M - m$.
τ_m, τ_M	Inter-event time interval for the point process at magnitude threshold m, M.
$p_m(\tau_m), p_M(\tau_M)$	Probability density functions of τ_m, τ_M (inter-event time distributions).
t_j	Time of j-th event in ETAS time series or of j-th Bayesian update.
M_j	Magnitude of j-th event in ETAS time series.
$\lambda(t)$	Event occurrence rate at time t.
λ_0	Constant occurrence rate of background seismicity.
K, α, c, θ	Parameters in the ETAS model to determine history-dependence.
M_0	The minimum magnitude in ETAS time series.
$P(M)$	Probability density of magnitude being M.
b	Parameter named the b-value that characterizes the Gutenberg–Richter law.
$p_{mM}(\tau_m\|\tau_M)$	Conditional probability density function of a lower interval length given the upper interval length τ_M.
N_m, N_M	Number of intervals at magnitude threshold m, M.
$\langle\langle \tau_m \rangle\rangle_{\tau_M}$	Average of the conditional probability density function $[p_{mM}(\tau_m\|\tau_M)]$.
$p_{Mm}(\tau_M\|\tau_m)$	Inverse probability density function of an upper interval length given the length τ_m of a lower interval in it.
$N_{mM}(\tau_m\|\tau_M)$	Number of upper and lower interval pairs of lengths τ_M and τ_m.
$\langle\tau_m\rangle, \langle\tau_M\rangle$	Average of $p_m(\tau_m), p_M(\tau_M)$.
$z_m(\tau_m), z_M(\tau_M)$	Generalized inter-event time distributions at magnitude threshold m, M.
$z_{mM}(\tau_m\|\tau_M)$	Generalized conditional probability density function.
$\rho_{mM}(\tau_m\|i, \tau_M)$	Conditional probability density function of a lower interval length given the upper interval with length τ_M including i lower intervals.
$\Psi_{mM}(i\|\tau_M)$	Probability mass function of the number of lower intervals in the upper interval of length τ_M.
$A_{\Delta m}$	$:= \langle\tau_M\rangle/\langle\tau_m\rangle - 1$ $(= 10^{b\Delta m} - 1)$.
$\delta(\cdot)$	Dirac's delta function.
$\theta(\cdot)$	Unit step function.
$\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\}$	n consecutive lower intervals of lengths $\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\}$.
$N_{mM}(\tau_M, \tau_m^{(1)}, \ldots, \tau_m^{(n)})$	Number of combinations of an upper interval of length τ_M and $\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\}$.
$p_{Mm}(\tau_M\|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$	Inverse probability density function of an upper interval length given $\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\}$ in it.
$\mathcal{T}$	Sum of the lengths of consecutive lower intervals.
$N'_{mM}(\tau_m^{(1)}, \ldots, \tau_m^{(n)}\|\tau_M)$	Number of $\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\}$ in the new time series in Fig. 5(b).
$P^{\mathrm{L}}(\tau_m\|\tau_M)$ $[P^{\mathrm{R}}(\tau_m\|\tau_M)]$	Probability density function of the left (right) most lower interval length in an upper interval of length τ_M.
$P(\tau_m^{(1)}, \ldots, \tau_m^{(l)}\|\tau_M)$	Probability density function of the left (right) most lower intervals lengths in an upper interval of length τ_M.
$P_i(\tau_m^{(i)}\|\tau_M)$	Probability density function of the i-th interval length in consecutive lower intervals in an upper interval of length τ_M.
$p_{Mm}(\tau_M\|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$	Inverse probability density function of an upper interval length given $\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\}$.
$p_{Mm}^{\mathrm{approx}}(\tau_M\|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$	Approximation function of $p_{Mm}(\tau_M\|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$.
$p_{Mm}^{\mathrm{kernel}}(\tau_M\|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$	Kernel part of the approximation function $p_{Mm}^{\mathrm{approx}}(\tau_M\|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$.
$p_{Mm}^{\mathrm{correct}}(\tau_M\|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$	Correction term in the approximation function $p_{Mm}^{\mathrm{approx}}(\tau_M\|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$.
$\tau_{m,j}, \tau_{M,k}$	Discretized intervals defined by Eq. (42).
$j_{\min}, j_{\max}$ $(j_{\min}^{(k)}, j_{\max}^{(k)})$	Minimum and maximum of j in Eq. (42) (with explicit dependence on k).
$k_{\min}, k_{\max}$	Minimum and maximum of k in Eq. (42).
l_e	Parameter for the calculation regarding extrapolation of the range of the bivariate distributions.
l_c	Parameter for testing the effect of the edge in the $\boldsymbol{P}_1$.
$\Delta\tau_m, \Delta\tau_M$	Increments in discretized intervals.
$\boldsymbol{p}_m, \boldsymbol{p}_M$	Numerically calculated inter-event time distributions defined by Eq. (43).
$\boldsymbol{p}_{mM}, \boldsymbol{P}_1$	Numerically calculated conditional probability density function and $P_1(\tau_m\|\tau_M)$ defined by Eq. (43).
$\mathcal{N}$	Number of time series for sample data.
$p_{\sup}, p_{\inf}$	Upper and lower bounds imposed on numerical calculations.
$D(f\|\|g)$	Distance between two square-integrable functions f and g defined by Eq. (46).
$D'(\cdot\|\|\cdot)$	Distance calculated by Eq. (47) with Eqs. (40) and (41).
$D''(\cdot\|\|\cdot)$	Distance calculated by Eq. (47) with Eqs. (37) and (38).
T	Elapsed time from the event with magnitude above M.
$\hat{\tau}_M^{\max}$	Maximum peak time in Eq. (25) which is discretized in Eq. (42).
$\tau_M^{\max,\mathrm{approx}}$	Maximum peak time in Eq. (37).
$k^{\max}, k^{\max,\mathrm{approx}}$	The k corresponds to $\hat{\tau}_M^{\max}$ and $\tau_M^{\max,\mathrm{approx}}$ by Eq. (42).
$\tau_M^{\max}$ $(\tau_M^{\max,\mathrm{n}})$	Maximum peak time of the kernel part in Eq. (38) (at the n-th update).
$k^{\max}$ $(k^{\max,\mathrm{n}})$	The k corresponds to $\tau_M^{\max}$ $(\tau_M^{\max,\mathrm{n}})$ by Eq. (42) (at the n-th update).
$\tau_M^{\max,\mathrm{L}}$, $\tau_M^{\max,\mathrm{R}}$	Maximum peak time in Eq. (37) where P_i are all replaced by P^{L} or P^{R}.
$k^{\max,\mathrm{L}}, k^{\max,\mathrm{R}}$	The k corresponds to $\tau_M^{\max,\mathrm{L}}$ and $\tau_M^{\max,\mathrm{R}}$ by Eq. (42).
τ_M^*	The time interval from the previous to the next events with magnitudes $> M$.
δ_n	Relative error between $\tau_M^{\max,\mathrm{n}}$ and τ_M^* defined by Eq. (49).
δ_{th}	Threshold of relative error δ_n to judge the accuracy of the estimation.
n_{fin}	Total number of updates until the next event greater than M.
$n_{\leq}$	Consecutive number of updates that satisfy $\|\delta_n\| \leq \delta_{\mathrm{th}}$ and include n_{fin}-th update.
R_n,	Occurrence rate of events defined by Eq. (52).
$\Delta\log_{10} R_n, \Delta k_n^{\max}$	Variations of $\log_{10} R_n$ and $k^{\max,\mathrm{n}}$ defined by Eqs. (53) and (54).
P_{fin}	Probability of $\|\delta_{n_{\mathrm{fin}}}\| \leq \delta_{\mathrm{th}}$.
$P_{\geq 30}$	Probability of $n_{\leq} \geq 30$ among $n_{\leq} > 0$.
$\tau_{\leq \mathrm{th}}$	Time duration corresponds to the $n_{\leq}$ consecutive updates.

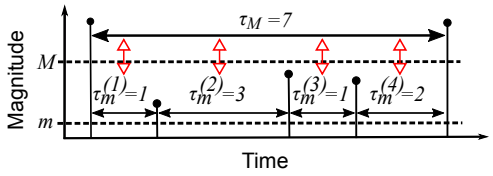

Fig. 2. (Color online) Schematic of the approach to count the number of pairs of upper and lower intervals whose lengths are τ_M and τ_m, respectively. Four pairs are shown in the figure, and $N_{mM}(7,1) = 2$, $N_{mM}(7,2) = 1$, and $N_{mM}(7,3) = 1$.

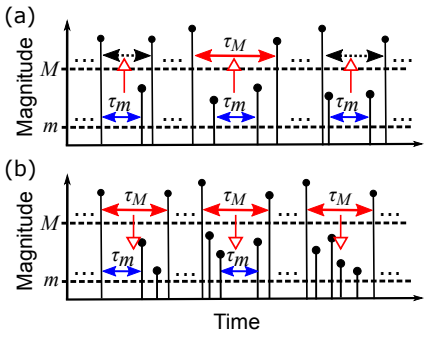

Fig. 3. (Color online) Schematic of the two approaches for calculating $N_{mM}(\tau_M, \tau_m)$. (a) The first approach involves counting the cumulative total number of the upper intervals of length τ_M that include the lower interval of length τ_m. (b) The second approach involves counting the number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

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 $\lambda(t) \equiv \lambda_0$. In this case, the magnitudes and inter-event times obey the following probability density functions independently.

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

$$p_m(\tau_m) = \frac{1}{\langle \tau_m \rangle} e^{-\frac{\tau_m}{\langle \tau_m \rangle}}. \tag{11}$$

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

$$p_{mM}(\tau_m|\tau_M) = \frac{\sum_{i=1}^{\infty} i \rho_{mM}(\tau_m|i,\tau_M)\Psi_{mM}(i|\tau_M)}{\sum_{i=1}^{\infty} i \Psi_{mM}(i|\tau_M)}, \tag{12}$$

where i ($\in \mathbb{N}$) 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 a lower interval length 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

$$\Psi_{mM}(i|\tau_M) = \frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right)^{i-1}}{(i-1)!} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}, \tag{13}$$

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 $i = 1$,

$$\rho_{mM}(\tau_m|1,\tau_M) = \delta(\tau_M - \tau_m), \tag{14}$$

where $\delta(\cdot)$ represents the Dirac's delta function. For $i \geq 2$,[40,41]

$$\rho_{mM}(\tau_m|i,\tau_M) = \frac{(i-1)}{\tau_M}\left(1 - \frac{\tau_m}{\tau_M}\right)^{i-2}\theta(\tau_M - \tau_m), \tag{15}$$

where $\theta(\cdot)$ represents the unit step function that takes 1 for positive argument and 0 for negative or 0 argument.

From Eqs. (13)–(15), $p_{mM}(\tau_m|\tau_M)$ is derived as (Appendix A)

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

This conditional probability composed of Eqs. (13)–(15) certainly has exponential distributions as the solution of Eq. (2) (Appendix A).

Second, we derive $p_{Mm}(\tau_M|\tau_m)$. From Eqs. (11) and (16), $p_{Mm}(\tau_M|\tau_m)$ is obtained as (Appendix A)

$$p_{Mm}(\tau_M|\tau_m) = \frac{e^{-\frac{\tau_M-\tau_m}{\langle\tau_m\rangle}}\delta(\tau_M-\tau_m) + \frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-\frac{\tau_M-\tau_m}{\langle\tau_M\rangle}}\left(A_{\Delta m}\frac{\tau_M-\tau_m}{\langle\tau_M\rangle}+2\right)\theta(\tau_M-\tau_m)}{(A_{\Delta m}+1)^2}. \tag{17}$$

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

$$\tau_M^{\max} = \tau_m + \langle\tau_M\rangle\left(1-\frac{2}{A_{\Delta m}}\right), \tag{18}$$

when the next condition is satisfied (Appendix A).

$$\Delta m > \frac{\log_{10}3}{b}. \tag{19}$$

4. Bayesian Updating for Uncorrelated Time Series

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. We derive the inverse probability density function $p_{Mm}(\tau_M|\tau_m^{(1)},\ldots,\tau_m^{(n)})$, as well as its approximation function, for the upper interval under the condition that it includes the consecutive lower intervals of lengths $\{\tau_m^{(1)},\ldots,\tau_m^{(n)}\}$.

4.1 Inverse probability density function

As in Sect. 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 $\{\tau_m^{(1)},\ldots,\tau_m^{(n)}\}$ included in it denoted by $N_{mM}(\tau_M,\tau_m^{(1)},\ldots,\tau_m^{(n)})$ in two ways.

First, we derive $N_{mM}(\tau_M,\tau_m^{(1)},\ldots,\tau_m^{(n)})$ by counting the cumulative total number of the upper intervals of length τ_M that include the consecutive lower intervals of lengths $\{\tau_m^{(1)},\ldots,\tau_m^{(n)}\}$ [Fig. 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 Fig. 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

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

Therefore

$$N_{mM}(\tau_M,\tau_m^{(1)},\tau_m^{(2)}) = N_m\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)p_m(\tau_m^{(1)})p_m(\tau_m^{(2)}) \times p_{Mm}(\tau_M|\tau_m^{(1)},\tau_m^{(2)})\,d\tau_m^2\,d\tau_M. \tag{20}$$

Equation (20) is generalized for n (≥ 2) consecutive lower intervals.

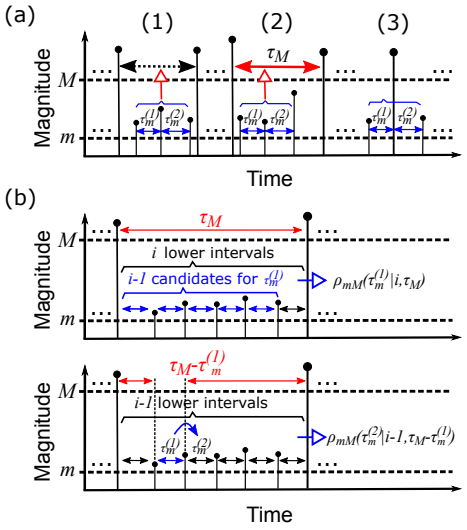

Fig. 4. (Color online) Schematic of the two approaches to calculate $N_{mM}(\tau_M,\tau_m^{(1)},\tau_m^{(2)})$. (a) The first approach involves counting the cumulative total number of the upper intervals of length τ_M that include the consecutive lower intervals of lengths $\{\tau_m^{(1)},\tau_m^{(2)}\}$.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

$$N_{mM}(\tau_M,\tau_m^{(1)},\ldots,\tau_m^{(n)}) = N_m\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}\left(\prod_{i=1}^{n}p_m(\tau_m^{(i)})\right) \times p_{Mm}(\tau_M|\tau_m^{(1)},\ldots,\tau_m^{(n)})\,d\tau_m^n\,d\tau_M. \tag{21}$$

Second, we derive $N_{mM}(\tau_M,\tau_m^{(1)},\ldots,\tau_m^{(n)})$ by counting the total number of the consecutive lower intervals of lengths $\{\tau_m^{(1)},\ldots,\tau_m^{(n)}\}$ included in the upper interval of length τ_M [Fig. 4(b)]. To this end, we start with the case $n=2$ again [Fig. 4(b)]. When the upper interval of length τ_M includes i (≥ 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 (≥ 2)

$$N_{mM}(\tau_M,\tau_m^{(1)},\tau_m^{(2)}) = N_M p_M(\tau_M)\,d\tau_M\sum_{i=2}^{\infty}(i-1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M) \times \rho_{mM}(\tau_m^{(2)}|i-1,\tau_M-\tau_m^{(1)})\,d\tau_m^2. \tag{22}$$

Equation (22) is generalized for the case n (≥ 2) lower intervals as

$$N_{mM}(\tau_M, \tau_m^{(1)}, \ldots, \tau_m^{(n)}) = N_M p_M(\tau_M)\, d\tau_M \times \sum_{i=n}^{\infty} (i-n+1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1}\tau_m^{(k)}\right) d\tau_m^n. \tag{23}$$

From Eqs. (21) and (23), $p_{Mm}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$ is derived as

$$p_{Mm}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{1}{\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}} \frac{p_M(\tau_M)}{\prod_{i=1}^{n} p_m(\tau_m^{(i)})} \times \sum_{i=n}^{\infty} (i-n+1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1}\tau_m^{(k)}\right). \tag{24}$$

Furthermore, the explicit form of the inverse probability density function is derived by substituting Eqs. (11) and (13)–(15) into Eq. (24) as (Appendix B)

$$p_{Mm}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)}) = \left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^2 \left\{ e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_m\rangle}} \delta\left(\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}\right) + \frac{A_{\Delta m}}{\langle\tau_M\rangle} e^{-\frac{\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}}{\langle\tau_m\rangle}} \left[\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left(\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}\right) + 2\right] \theta\left(\tau_M - \sum_{i=1}^{n}\tau_m^{(i)}\right) \right\}. \tag{25}$$

Equation (25) includes the case $n = 1$ [Eq. (17)]. In addition, Eq. (25) is identical to Eq. (17) 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 Eqs. (18) and (19) holds for $p_{Mm}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$; it has a peak at

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

under the condition

$$\Delta m > \frac{\log_{10} 3}{b}. \tag{26}$$

In the above-mentioned Bayesian updating, the position of the consecutive lower intervals in an upper interval is not restricted. However, the 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 Eq. (25) (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.[42] 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 examining 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 (24) 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 subsection, we derive its approximation function with a convenient form applicable to the time series with correlations between events.

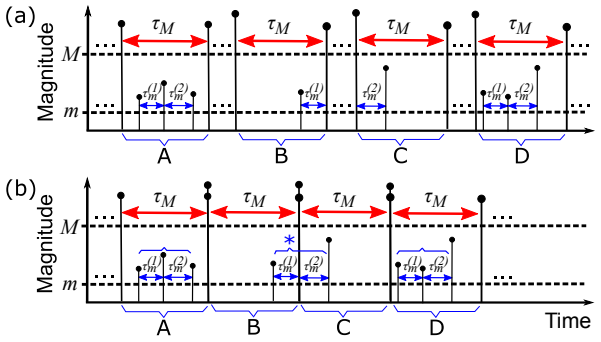

Fig. 5. (Color online) Schematic of another approach to count the total number of consecutive lower intervals of lengths $\tau_m^{(1)}$ and $\tau_m^{(2)}$ included in the upper interval of length τ_M. (a) First, pick up all upper intervals of length τ_M from the time series. (b) Second, generate new 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 $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$.

To this end, we use the approximate derivation of $N_{mM}(\tau_M, \tau_m^{(1)}, \ldots, \tau_m^{(n)})$ described below instead of the second approach for deriving Eq. (23). In the following, the upper and the lower consecutive intervals are assumed to satisfy

$$\tau_M \geq \sum_{i=1}^{n}\tau_m^{(i)}. \tag{27}$$

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 Fig. 5(a), and we use them to generate a new time series by connecting them in the order of appearance as in Fig. 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

$$N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M) \approx N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m^{(1)}|\tau_M) p_{mM}(\tau_m^{(2)}|\tau_M)\, d\tau_m^2\, d\tau_M. \tag{28}$$

$N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$ 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 upper interval [the case indicated by $*$ in Fig. 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

$$P^{\mathrm{R}}(\tau_m|\tau_M) = P^{\mathrm{L}}(\tau_m|\tau_M) = \sum_{i=1}^{\infty} \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m|i, \tau_M). \tag{29}$$

Here, the probability density for the rightmost lower interval is denoted by $P^{\mathrm{R}}(\tau_m|\tau_M)$, and the leftmost by $P^{\mathrm{L}}(\tau_m|\tau_M)$.

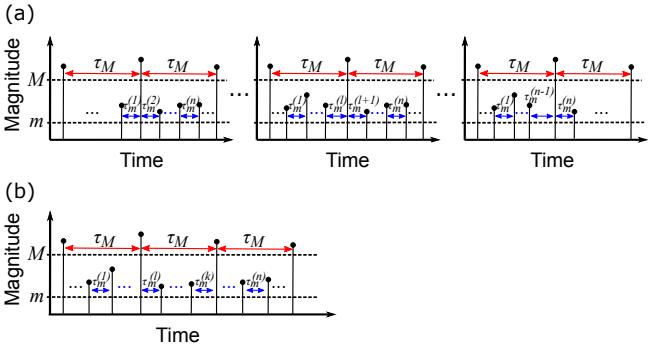

Fig. 6. (Color online) Schematic of the patterns of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\}$ excluded from $N'_{mM}(\tau_m^{(1)}, \ldots, \tau_m^{(n)}|\tau_M)$. (a) There are $(n-1)$ ways to divide the sequence of lower intervals by the event with a magnitude 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 (27).

Equation (29) can be explicitly written using Eqs. (13)–(15) as (Appendix D)

$$P^{\mathrm{R}}(\tau_m|\tau_M) = P^{\mathrm{L}}(\tau_m|\tau_M) = e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}} \delta(\tau_M - \tau_m) + \frac{A_{\Delta m}}{\langle\tau_M\rangle} e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}} \theta(\tau_M - \tau_m). \tag{30}$$

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 Eq. (31) is precisely $N_M p_M(\tau_M)\, d\tau_M - 1$] as

$$N_M p_M(\tau_M) P^{\mathrm{R}}(\tau_m^{(1)}|\tau_M) P^{\mathrm{L}}(\tau_m^{(2)}|\tau_M)\, d\tau_m^2\, d\tau_M. \tag{31}$$

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

$$N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)}) \approx N_M p_M(\tau_M) \left(\frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m^{(1)}|\tau_M) p_{mM}(\tau_m^{(2)}|\tau_M) - P^{\mathrm{R}}(\tau_m^{(1)}|\tau_M) P^{\mathrm{L}}(\tau_m^{(2)}|\tau_M) \right) d\tau_m^2\, d\tau_M. \tag{32}$$

Next, we consider the case n (≥ 3). Equation (28) is generalized as

$$N'_{mM}(\tau_m^{(1)}, \ldots, \tau_m^{(n)}|\tau_M) \approx N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \left(\prod_{i=1}^{n} p_{mM}(\tau_m^{(i)}|\tau_M) \right) d\tau_m^n\, d\tau_M. \tag{33}$$

From this $N'_{mM}(\tau_m^{(1)}, \ldots, \tau_m^{(n)}|\tau_M)$, the cases wherein the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\}$ are not included in the same upper interval need to be excluded. Considering the condition of Eq. (27), a sequence of consecutive lower intervals is divided by only one boundary event with a magnitude above M (Fig. 6). Let $P(\tau_m^{(1)}, \ldots, \tau_m^{(l)}|\tau_M)$ be the probability that the rightmost or leftmost lower intervals of the upper interval of length τ_M is $\{\tau_m^{(1)}, \ldots, \tau_m^{(l)}\}$ ($l \geq 2$). Then, as the position of the rightmost or leftmost lower intervals is confirmed among the i ($\geq l$) lower intervals, $P(\tau_m^{(1)}, \ldots, \tau_m^{(l)}|\tau_M)$ is

$$P(\tau_m^{(1)}, \ldots, \tau_m^{(l)}|\tau_M) = \sum_{i=l}^{\infty} \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{l} \rho_{mM}\left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)} \right). \tag{34}$$

By substituting Eqs. (13)–(15) into Eq. (34) (Appendix E)

$$P(\tau_m^{(1)}, \ldots, \tau_m^{(l)}|\tau_M) = \prod_{i=1}^{l} P_i(\tau_m^{(i)}|\tau_M), \text{ where } P_i(\tau_m^{(i)}|\tau_M) = \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle} \right) e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}. \tag{35}$$

There are $(n-1)$ possible choices for the boundary position of the consecutive lower intervals [Fig. 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

$$N_{mM}(\tau_M, \tau_m^{(1)}, \ldots, \tau_m^{(n)}) \approx N_M p_M(\tau_M) \left[\frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \prod_{i=1}^{n} p_{mM}(\tau_m^{(i)}|\tau_M) - (n-1) \prod_{i=1}^{n} P_i(\tau_m^{(i)}|\tau_M) \right] d\tau_m^n \, d\tau_M. \tag{36}$$

Therefore, from Eqs. (21) and (36), the approximation function $[p_{Mm}^{\text{approx}}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)})]$ of the inverse probability density function is derived as

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

Equation (37) is composed of two parts: the first term on the right hand side (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 $[p_{Mm}^{\text{kernel}}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)})]$ hereafter.

$$p_{Mm}^{\text{kernel}}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)}) = \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle} \frac{1}{\left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1}} \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \left(\prod_{i=1}^{n} \frac{p_{mM}(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})} \right) p_M(\tau_M). \tag{38}$$

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 $p_{Mm}^{\text{correct}}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$ as

$$\text{correction term} = (n-1) p_{Mm}^{\text{correct}}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)}), \tag{39}$$

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

Equation (37) can be explicitly written as (Appendix F)

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

The kernel part is explicitly expressed as

$$p_{Mm}^{\text{kernel}}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)}) = \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle^2} \left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right) \left(A_{\Delta m} \frac{\tau_M}{\langle \tau_M \rangle} + 1 \right) e^{-\frac{\tau_M - \sum_{i=1}^{n} \tau_m^{(i)}}{\langle \tau_M \rangle}} \prod_{i=1}^{n} \left[1 - \left(\frac{\tau_m^{(i)} - \frac{\langle \tau_M \rangle}{A_{\Delta m}}}{\tau_M + \frac{\langle \tau_M \rangle}{A_{\Delta m}}} \right) \right]. \tag{41}$$

Note that functions (37)–(41) do not satisfy the normalization condition. Furthermore, in some cases, $p_{Mm}^{\text{approx}}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$ in Eqs. (37) and (40) 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 Eq. (25) and the (part of) approximation function [Eqs. (37)–(41)] for the numerically generated uncorrelated time series, and we compare their properties. We examine the numerical method of 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 Eq. (1). In fact, it was numerically generated as the renewal process in which magnitudes and time intervals were generated randomly obeying Eqs. (10) and (11), respectively. We set the parameter values to be $b = 1$ and $\lambda_0 = 0.0007$. M_0 was set to 3. It should be noted that the magnitudes were set to be generated in the range greater than $M_0 = 3$; however, as the outputs were only six decimal places, a small number of events with magnitude $\equiv 3$ existed. Such cases were excluded from the analysis by setting the following lower magnitude threshold at $m = M_0$. The magnitude thresholds were set to $(M, m) = (5, 3)$. The b-value condition of Eq. (26) is satisfied for these settings. The occurrence time of each event was recorded to 20 decimal places. For such time series, Bayesian updating was applied as explained below.

Bayesian updating was 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 $\{\tau_m^{(1)}, \tau_m^{(2)}, \ldots, \tau_m^{(n)}\}$ into Eqs. (25), (40), and (41). 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 was 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 was not used). Therefore, we considered only cases where at least one event was (or two lower intervals were) included in an upper interval.

In addition, we used the following numerical method based on Eqs. (37) and (38). First, we generated $\mathcal{N}$ time series each contains 10^5 events as sample data. From these sample data, numerically obtained the statistics required for the calculating Eqs. (37) and (38), 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 inside 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 calculated it separately. Moreover, we calculated only $P_1(\tau_m|\tau_M)$ and used it instead of $P_i(\tau_m|\tau_M)$ for $i \geq 2$.

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

$$\tau_{m,j} := 10^{(j+0.5)\Delta\tau_m},$$
$$\tau_{M,k} := 10^{(k+0.5)\Delta\tau_M}, \tag{42}$$

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

$$\begin{aligned} \boldsymbol{p}_m &= [p_{m,j}]_{j=j_{\min},\ldots,j_{\max}}, \boldsymbol{p}_M = [p_{M,k}]_{k=k_{\min},\ldots,k_{\max}}, \\ \boldsymbol{p}_{mM} &= [[p_{mM,jk}]_{j=j_{\min}^{(k)},\ldots,j_{\max}^{(k)}}]_{k=k_{\min},\ldots,k_{\max}}, \\ \boldsymbol{P}_1 &= [[P_{1,jk}]_{j=j_{\min}^{(k)},\ldots,j_{\max}^{(k)}}]_{k=k_{\min},\ldots,k_{\max}}. \end{aligned} \tag{43}$$

In Eq. (43), $j_{\min}$, $j_{\max}$, $k_{\min}$, 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 Eq. (43). 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 were fully used only for $p_{mM}(\tau_m|\tau_M)$ and $P_1(\tau_m|\tau_M)$, and only 10^3 of them were 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 performed 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 interpolated 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 extrapolated the fitting curve for the edge 10 points (Fig. S1[43]). The distributions were 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 performed the same interpolations and extrapolations for $\tau_{m,j}$ (Figs. S2 and S3[43]). Meanwhile, for $\tau_{M,k}$, the domain was extended using the average of the functions at $\{\tau_{M,k_{\min}}, \ldots, \tau_{M,k_{\min}+l_e-1}\}$ as the substitute for $\tau_{M,k}$ with $k < k_{\min}$, whereas using the functions at $\{\tau_{M,k_{\max}-l_e+1}, \ldots, \tau_{M,k_{\max}}\}$ 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 were conducted in the same way as $\boldsymbol{p}_M$, although the extrapolation functions were different (Fig. S4[43]).

Thus, the discrete variable $\tau_{m,j}$ became continuous as τ_m, and the distribution functions were defined for all τ_m larger than 0. This made it possible to return a value for any input of the length of a lower interval when performing Bayesian updating. Further, the distribution functions were defined for any k in Eq. (42). 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 yielded the maximum range of Bayesian updating, the updating at the n-th step was performed within the range $\max\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\} < \tau_M$. The properties of the inverse probability density function and the (part of) approximation function were examined within this range.

The kernel parts of the approximation functions were computed by calculating Eq. (38) in a step-by-step manner as

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

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

$$\begin{aligned} \ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)}) &= \ln\left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(1)}|\tau_{M,k}) - \ln p_m(\tau_m^{(1)}) + \ln p_{M,k}, \\ \ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)}, \tau_m^{(2)}) &= -\ln\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(2)}|\tau_{M,k}) - \ln p_m(\tau_m^{(2)}) + \ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)}), \\ \ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)}, \tau_m^{(2)}, \tau_m^{(3)}) &= -\ln\left(1 - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(3)}|\tau_{M,k}) - \ln p_m(\tau_m^{(3)}) + \ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)}, \tau_m^{(2)}), \\ &\vdots \end{aligned} \tag{45}$$

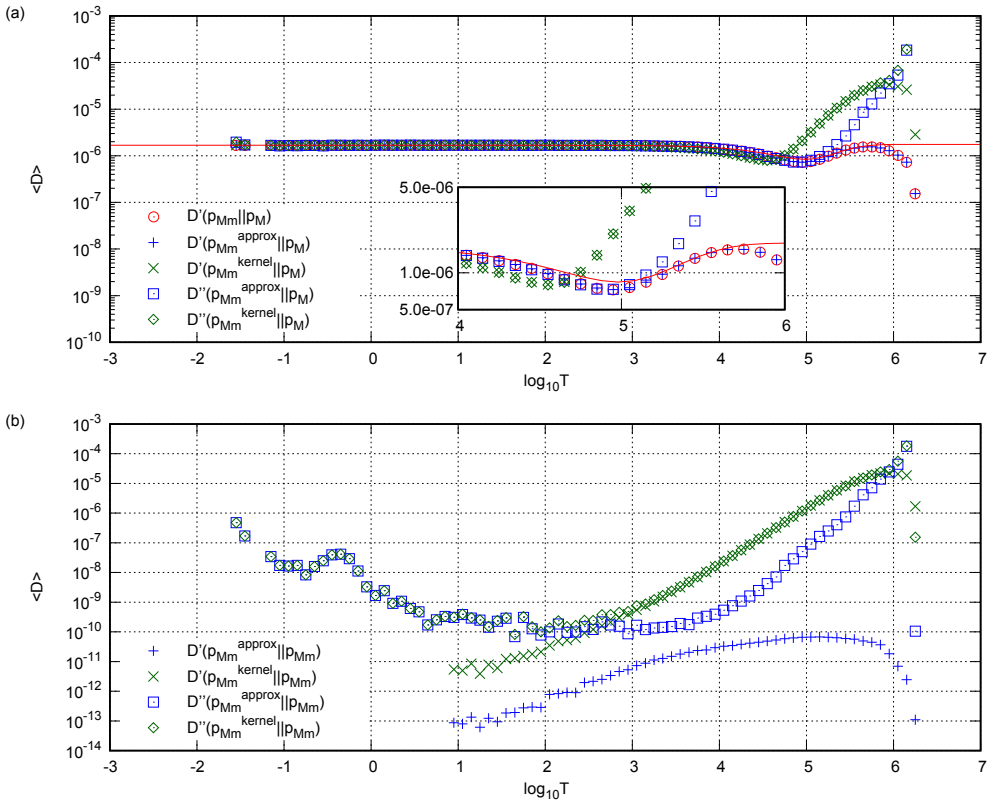

Fig. 7. (Color online) Average distances for each elapsed time (T) from the previous large event with a magnitude above M. (a) Distances between the inter-event time distribution and other function. $D(p_{Mm}\|p_M)$ [Eq. (H·2) in Appendix H] is shown by the red curve, and the symbols are numerical results for Eq. (47). (b) Distances between the inverse probability density function and other function numerically calculated by Eq. (47).

and then, we added $\ln(n-1)$ for each $\ln p^{\text{correct}}_{Mm}(\tau_M|\tau_m^{(1)},\ldots,\tau_m^{(n)})$.

The approximation functions were obtained by adding together the kernel part and the correction term calculated by these separate updates. The approximation functions were calculated only for such k's that $p_{\text{sup}} > \ln p^{\text{kernel}}_{Mm}$, $\ln p^{\text{correct}}_{Mm} > p_{\text{inf}}$. Here, p_{sup} $(= 600)$ and p_{inf} $(= -600)$ yielded the upper and lower limits of p^{kernel}_{Mm} and p^{correct}_{Mm} to ensure that these were within the range of the computer capacity. In addition, such k's for which the correction term was so large that Eq. (37) became negative were excluded.

Figure S5[43] shows an example of Bayesian updating for the uncorrelated time series. The inverse probability density function given by Eq. (25) has a characteristic peak that is not observed in $p_M(\tau_M)$. The correction term makes the kernel part obtained from Eq. (41) closer to the inverse probability density function. Moreover, the numerical calculations based on Eqs. (37) and (38) 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

$$D(f\|g) := \int_T^\infty |f(\tau_M) - g(\tau_M)|^2\, d\tau_M. \tag{46}$$

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 $f = p_{Mm}(\tau_M|\tau_m^{(1)},\ldots,\tau_m^{(n)})$ 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

$$D(f\|g) \simeq \sum_{\substack{k;\,\tau_{M,k}>T \\ p_{\text{sup}}>\ln f,\ln g>p_{\text{inf}}}} |f(\tau_{M,k}) - g(\tau_{M,k})|^2 (\ln 10)\tau_{M,k}\Delta\tau_M. \tag{47}$$

$D(f\|g)$ was calculated for each update throughout the 100 test data time series. If no k's satisfied $p_{\text{sup}} > \ln f, \ln g > p_{\text{inf}}$, it was not included in the following calculation. The average distance $\langle D(f\|g)\rangle$ was 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 $f = p_{Mm}(\tau_M|\tau_m^{(1)},\ldots,\tau_m^{(n)})$, $p^{\text{approx}}_{Mm}(\tau_M|\tau_m^{(1)},\ldots,\tau_m^{(n)})$, $p^{\text{kernel}}_{Mm}(\tau_M|\tau_m^{(1)},\ldots,\tau_m^{(n)})$, and $g = p_M(\tau_M)$. In addition to the analytical calculation in Eq. (46) for $D(p_{Mm}\|p_M)$, the results of the numerical integration of Eq. (47) are presented; the calculations using Eqs. (40) and (41) are indicated by $D'(\cdot\|\cdot)$. The results of the calculation using Eqs. (37) and (38) with the numerical method in Sect. 5.1 with $\mathcal{N} = 10^3$ and

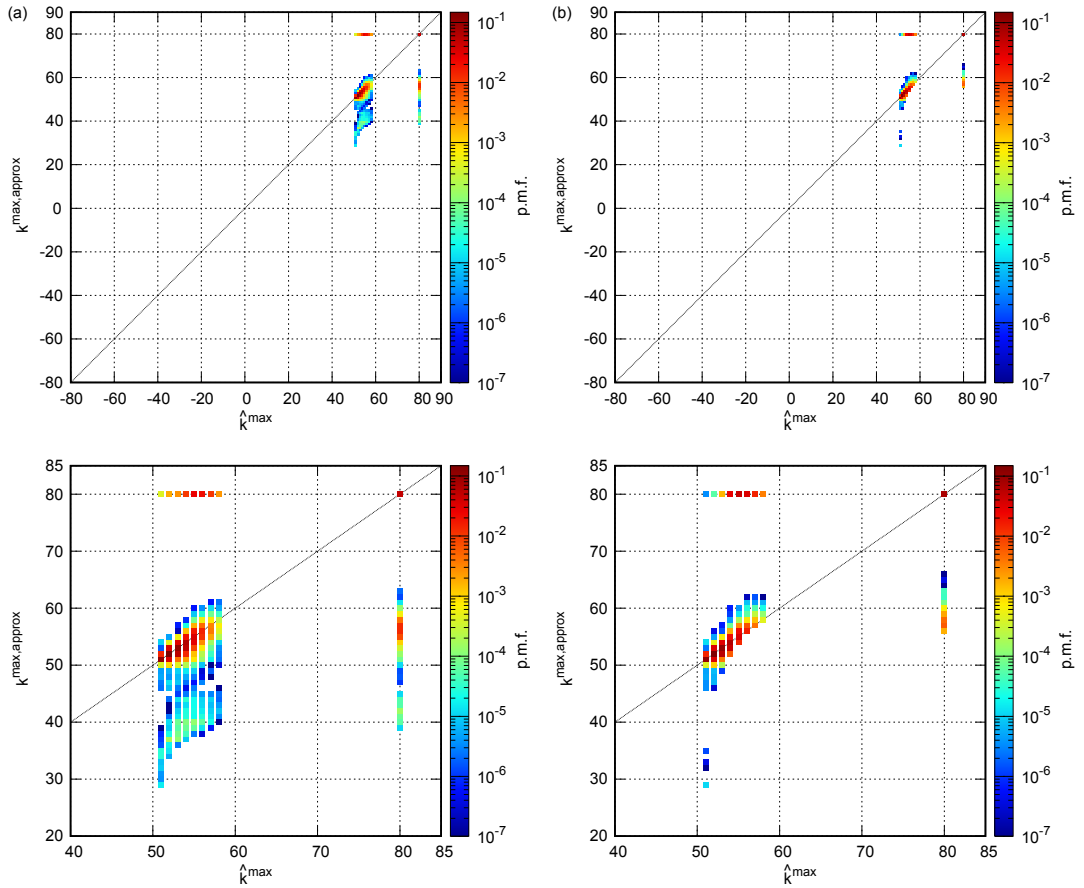

Fig. 8. (Color online) Joint probability mass function for $(\hat{k}^{\max}, k^{\max,\mathrm{approx}})$. Numerical search of the maximum peak is conducted for (a) $\tau_M > \max\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\}$ and (b) $\tau_M > T$. The horizontal line at $k^{\max,\mathrm{approx}} = 80$ and the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

$\Delta\tau_M = 0.1$ are presented by $D''(\cdot\|\cdot)$. The results for $\mathcal{N} = 10^5$ with $\Delta\tau_M = 0.1$ and 0.025 are shown in Fig. S6.[43)]

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 $p_{Mm}^{\mathrm{approx}}(\tau_M|\tau_m^{(1)}, \ldots, \tau_m^{(n)})$ 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 $\Delta\tau_M = 0.025$ [Figs. S6(c) and S6(d)[43)]], 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 Eq. (44) 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 Fig. 7(b), which shows the average distances between $f = p_{Mm}^{\mathrm{approx}}, p_{Mm}^{\mathrm{kernel}}$ and $g = p_{Mm}$ calculated by Eq. (47). This separation can be attributed to the calculation of the correction term in Eq. (45), 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 was 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 Eq. (25) and its approximation function in Eq. (37) with the numerical updating method, each denoted by $\hat{\tau}_M^{\max}$ and $\tau_M^{\max,\mathrm{approx}}$. Both functions are discretized as Eq. (42); the corresponding k in Eq. (42) is denoted by $\hat{k}^{\max}$ and $k^{\max,\mathrm{approx}}$, respectively.

$\hat{k}^{\max}$ and $k^{\max,\mathrm{approx}}$ were numerically searched for each update. These were determined as such k that the function took the maximum value within the range for which the above-mentioned numerical results were obtained, while excluding its edges. Thus, if $\hat{k}^{\max}$ or $k^{\max,\mathrm{approx}}$ was located at such edges, it was not considered the peak and was 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 were not obtained for any k (when the correction term exceeded the kernel part for all k), $k^{\max,\mathrm{approx}}$ was 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 Fig. S7.[43)] Here, the population is all the pairs of $(\hat{k}^{\max}, k^{\max,\mathrm{approx}})$ obtained for each update throughout the test data. In the following, we further discuss the area where the maximum peaks appeared. The maximum peak search was conducted in the two ranges; (a) $\tau_M > \max\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\}$, and (b) $\tau_M > T$. In the former case, the p.m.f. was bimodal; the higher peak existed around $\hat{k}^{\max} =$

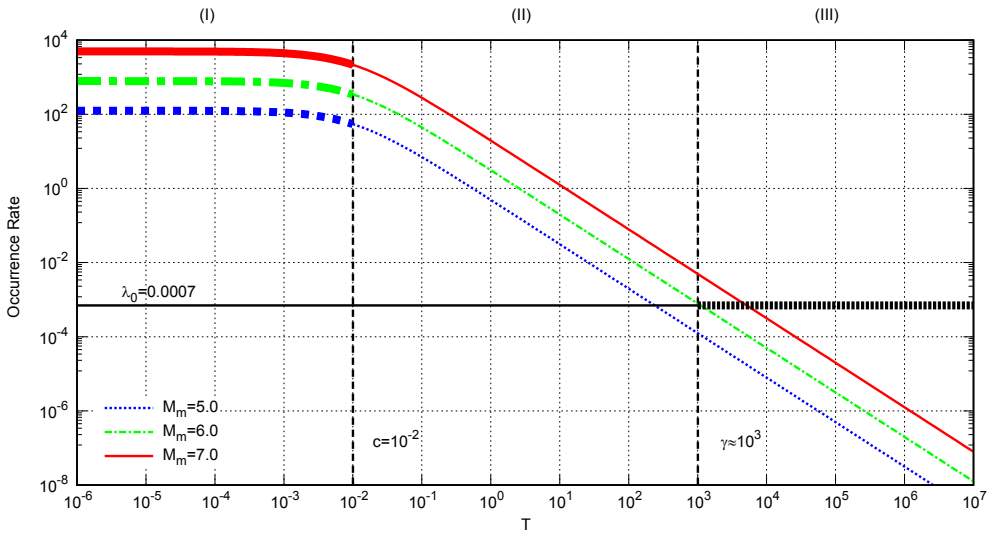

Fig. 9. (Color online) Omori–Utsu law for the parameter values in the text with a different mainshock magnitude M_m (after Ref. 39). The number of aftershocks per unit day against the elapsed time (T) from the mainshock obeys $\lambda(T) = K10^{\alpha(M_m - M_0)}/(T + c)^{\theta+1}$. The background rate ($\lambda_0 = 0.0007$) is also shown.

$k^{\max,\text{approx}}$, and the other lower peak around $\hat{k}^{\max} > k^{\max,\text{approx}}$. The second peak disappeared in the latter case, and the first peak is intrinsic, i.e., the positions of the maximum peak were close between the inverse probability density function and its approximation function. The situation was the same for other cases (Fig. S7[43]). These results indicate that it is the off-peak region of the approximation function that contributes to the separation of the average distances.

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. 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 [Eq. (37)] and its kernel part [Eq. (38)] using the numerical Bayesian updating method. The maximum peak time of the kernel part is used as the estimate for the time to occur the next large-magnitude event greater than the upper threshold, and the effectiveness of forecasting based on that estimate is evaluated statistically.

6.1 Time series generation and Bayesian updating methods

We applied the numerical Bayesian updating method in Sect. 5.1 to the time series generated by Eq. (1)[26,44] with the parameter values $b = 1$, $\alpha = 0.8$, $\theta = 0.2$, $c = 0.01$, $M_0 = 3$, $\lambda_0 = 0.0007$, and $K = 0.0125$; the parameter values were set in partially reference to the preceding numerical study[20] so that the branching ratio is less than 1.[30] The magnitude thresholds were $(M, m) = (5, 3)$. As in Sect. 5.1, we found some events (149 out of 1.1×10^8) had magnitude $\equiv 3$, those were excluded by the setting the lower threshold $m = 3$. 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 nearly the Omori–Utsu law after a large event, as shown in Fig. 9 [note that the local Omori–Utsu law in Eq. (1) is different from the global aftershock decay,[30] and thus, the actual decaying must be slightly different from shown in Fig. 9]. The activity can be categorized into three regimes with respect to the elapsed time (T) from the mainshock, as summarized in Table II.

Table II. 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 $(\lesssim \lambda_0)$

We prepared 1100 time series, with each containing 10^5 events. First, random numbers generated from five different seed values were used to generate 240 time series for each seed. Among them, those containing events with magnitude ≥ 10 were excluded. This is because the aftershock sequence excited by such an unrealistic large event does not fit within a single time series, and then, the non-stationarity affects the statistics of the sample data. We used 1100 of the remaining time series. $\mathcal{N} = 1000$ were used as the sample data to obtain statistics with $\Delta\tau_M = 0.1$; the interpolation and extrapolation procedures were conducted with $l_e = 5$ in the same way as explained in Sect. 5.1 (Figs. S8–S12[43]). Bayesian updating [Eqs. (44) and (45)] was applied to the remaining 100 time series. The maximum range of k was set to $-120 \leq k \leq 70$, and the n-th update from the occurrence time of the event above M was conduced in the range $\max\{\tau_m^{(1)}, \ldots, \tau_m^{(n)}\} < \tau_M$. The numerical update was conducted when the lower interval was above 0 (for the occurrence times recorded to 20 decimal places); otherwise, the update was skipped.

The following normalizations were performed in the calculations of Bayesian updating. The result of the

calculation in Eq. (44) can be very large. In order to compute the approximation functions together with Eq. (45), 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 normalized the result of Eq. (44) for each update by subtracting the following numerical integration from Eq. (44).

$$\ln\left(\sum_{\substack{k;\,\tau_{M,k}>T \\ p_{\mathrm{sup}}>\ln p_{Mm}^{\mathrm{kernel}}>p_{\mathrm{inf}}}} p_{Mm}^{\mathrm{kernel}}(\tau_{M,k}|\tau_m^{(1)},\ldots,\tau_m^{(n)})(\ln 10)\tau_{M,k}\Delta\tau_M\right). \tag{48}$$

Further, it is necessary to subtract Eq. (48) from the correction term in Eq. (45) at the same time (thereby the entire approximation function is multiplied by a constant). Thus, for each update of Eqs. (44) and (45), the numerical integration (48) was 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 examined the contribution from the correction term to the approximation function as follows. Instead of $P_i(\tau_m|\tau_M)$, we calculated the probability density functions $P^{\mathrm{L}}(\tau_m|\tau_M)$ and $P^{\mathrm{R}}(\tau_m|\tau_M)$ (Figs. S10 and S11[43]). 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 were calculated numerically by replacing all $P_i(\tau_m|\tau_M)$'s in Eq. (45) 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 Eq. (42) 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} was conducted in the same way as indicated in Sect. 5.2 in the range $\tau_M > \max\{\tau_m^{(1)},\ldots,\tau_m^{(n)}\}$.

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 was calculated in the same way as indicated in Sect. 5.2. From the results when the maximum peaks were detected, 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 [in some cases, the maximum peak was undetected in the approximation function; particularly those with $P^{\mathrm{L}}(\tau_m|\tau_M)$ showed high probability at $k^{\mathrm{max,L}} = 80$. It should be noted that the discussion here is based on the cases without when the maximum peak time was undetected]. 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 when the event above M will occur, and we discuss the effectiveness of the forecasting based on the estimates.

6.3 *Estimation of the next large event timing and effectiveness of forecasting*

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

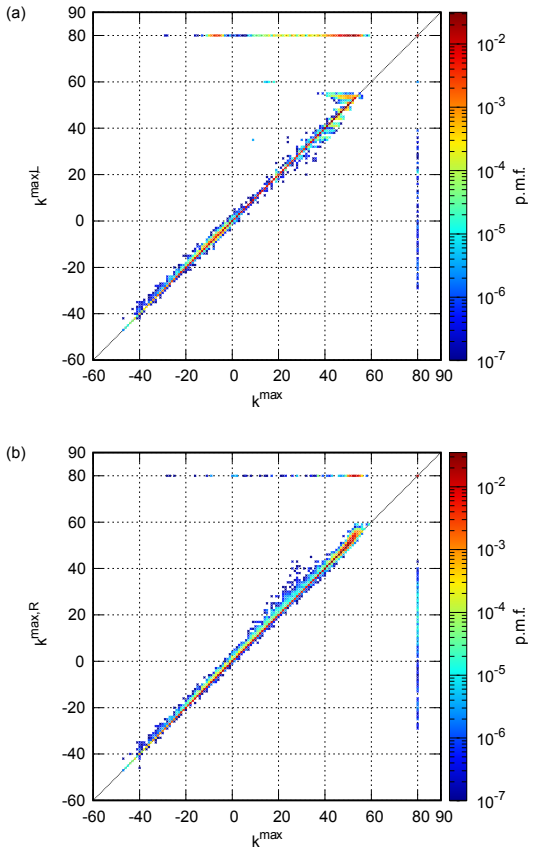

Fig. 10. (Color online) Joint probability mass functions for (a) (k^{max}, $k^{\mathrm{max,L}}$) and (b) ($k^{\mathrm{max}}, k^{\mathrm{max,R}}$). The horizontal lines at $k^{\mathrm{max,L}} = 80$ and $k^{\mathrm{max,R}} = 80$ and the vertical line $k^{\mathrm{max}} = 80$ are the cases where the peak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

$$\delta_n := \frac{\tau_M^* - \tau_M^{\mathrm{max,n}}}{\tau_M^*}. \tag{49}$$

Equation (49) considers that the error $|\tau_M^* - \tau_M^{\mathrm{max,n}}|$ 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 was judged by whether δ_n was within the threshold (δ_{th})

$$-\delta_{\mathrm{th}} \le \delta_n \le \delta_{\mathrm{th}}. \tag{50}$$

When Eq. (50) 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

$$\frac{\tau_M^{\mathrm{max,n}}}{(1+\delta_{\mathrm{th}})} \le \tau_M^* \le \frac{\tau_M^{\mathrm{max,n}}}{(1-\delta_{\mathrm{th}})}. \tag{51}$$

Based on the above accuracy at each update, we further evaluated whether a series of estimations yields effective forecasting. Here, effective forecasting implies that τ_M^{max} takes a nearly constant value around τ_M^* continuously from well before the elapsed time τ_M^*. 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 Eq. (50) 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 $n \in (n_{\mathrm{fin}} -$

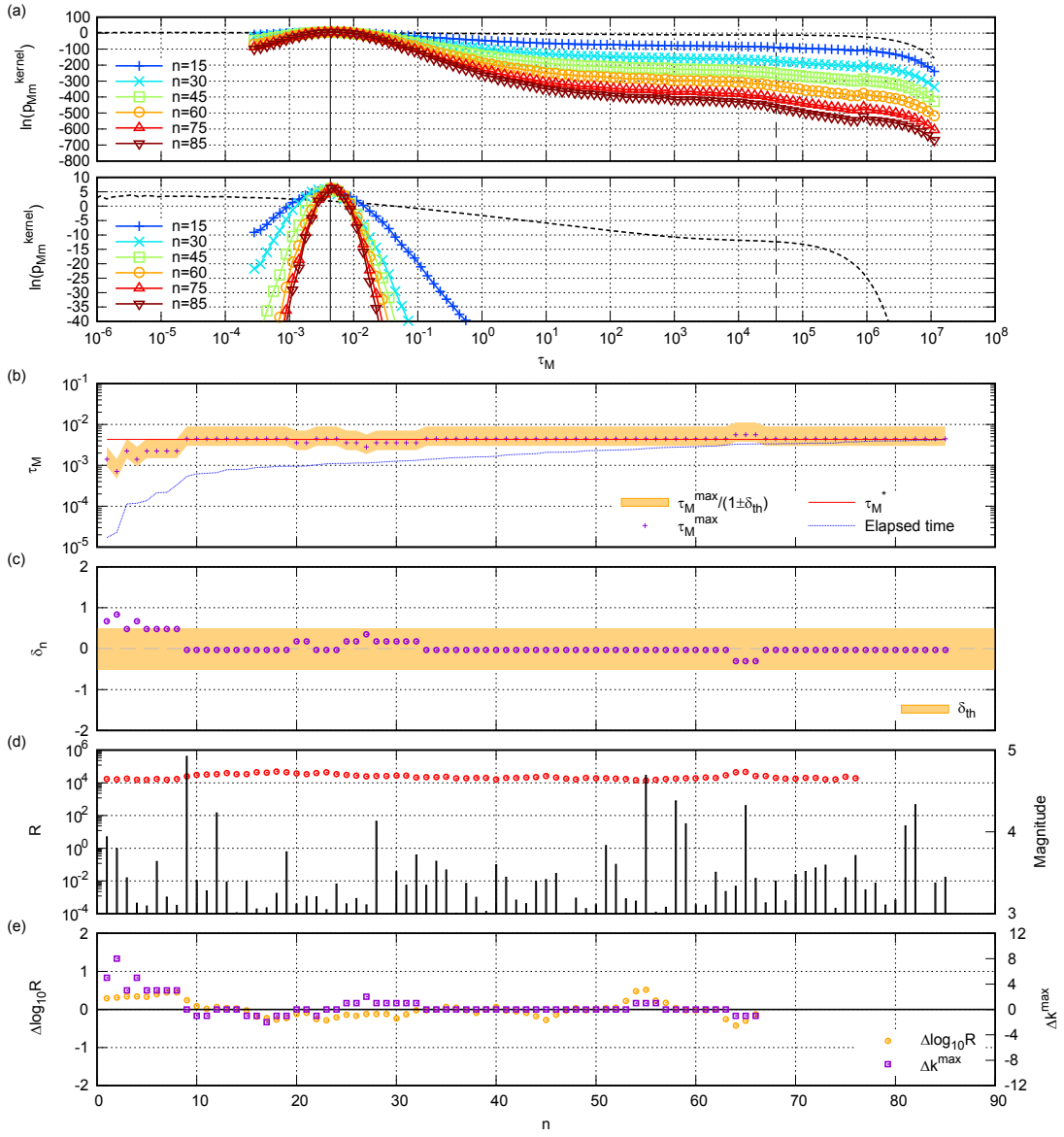

Fig. 11. (Color online) The first example of Bayesian updating and related amounts for the case where τ_M^* is in regime (I). (a) The kernel part for each update (n). The lower panel shows the enlarged view of the upper panel near the peak. The dotted curve indicates $p_M(\tau_M)$. The vertical solid line indicates τ_M^* and vertical dotted line the average $\langle\tau_M\rangle$. (b) Evolutions of the estimate ($\tau_M^{\max,n}$) and the tolerance of error $[\frac{\tau_M^{\max,n}}{1+\delta_{\rm th}}, \frac{\tau_M^{\max,n}}{1-\delta_{\rm th}}]$ in Eq. (51) with $\delta_{\rm th} = 0.5$. The elapsed time from the last larger-magnitude event than M is indicated by the blue dotted line. (c) Evolution of the relative error (δ_n). The orange band indicates the tolerance range $[-\delta_{\rm th}, \delta_{\rm th}]$. (d) Evolution of the occurrence rate [R_n defined by Eq. (52)]. 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 Eq. (53)] and the variation of the log-estimate [$\Delta k_n^{\max}$ defined by Eq. (54)].

$n_{\le\rm th}, n_{\rm fin}]$, we consider the forecasting to be effective. We judge the stability of $\tau_M^{\max,n}$ by Eq. (50), and therefore, $\delta_{\rm th}$ should not be too large. In the present paper, we set $\delta_{\rm th} = 0.5$ and 0.25.

To observe the relationship between the effectiveness of forecasting and the stationarity of the time series, we examined 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.

$$R_n := 10/(t_{n+9} - t_n), \tag{52}$$

$$\Delta \log_{10} R_n := \log_{10} R_{n+10} - \log_{10} R_n, \tag{53}$$

$$\Delta k_n^{\max} := k^{\max,n+10} - k^{\max,n}, \tag{54}$$

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

6.4 *Examples of Bayesian updating*

Figures 11–13 show examples of Bayesian updating and other related amounts for the cases where τ_M^* is included in each regime in Table II. $\delta_{\rm th}$ was set to 0.5.

Figure 11 shows the first example for $\tau_M^* \in$ (I). Figure 11(d) indicates that the occurrence rate is high, and it stays almost constant. The kernel part has a peak as shown in Fig. 11(a), and its maximum peak time ($\tau_M^{\max,n}$) continues to be nearly constant around τ_M^* from well before the large event as shown in Fig. 11(b); this is confirmed in Fig. 11(c), which indicates that $|\delta_n| \le \delta_{\rm th}$ is satisfied consecutively for $n \in (n_{\rm fin} - n_{\le\rm th}, n_{\rm fin}]$ with a long $n_{\le\rm th}$, and in Fig. 11(e), that shows that $\Delta k_n^{\max}$ fluctuates around 0. Therefore, in this example, τ_M^* is judged to be effectively forecasted.

Figure 12 is the second example for $\tau_M^* \in$ (III). In this example, the occurrence rate is low and keeps almost constant around $\lambda_0 = 0.0007$ as shown in Fig. 12(d). Figure 12(a) indicates that the kernel part has a peak and Figs. 12(b) and 12(c) show that the maximum peak time ($\tau_M^{\max,n}$) transitions to

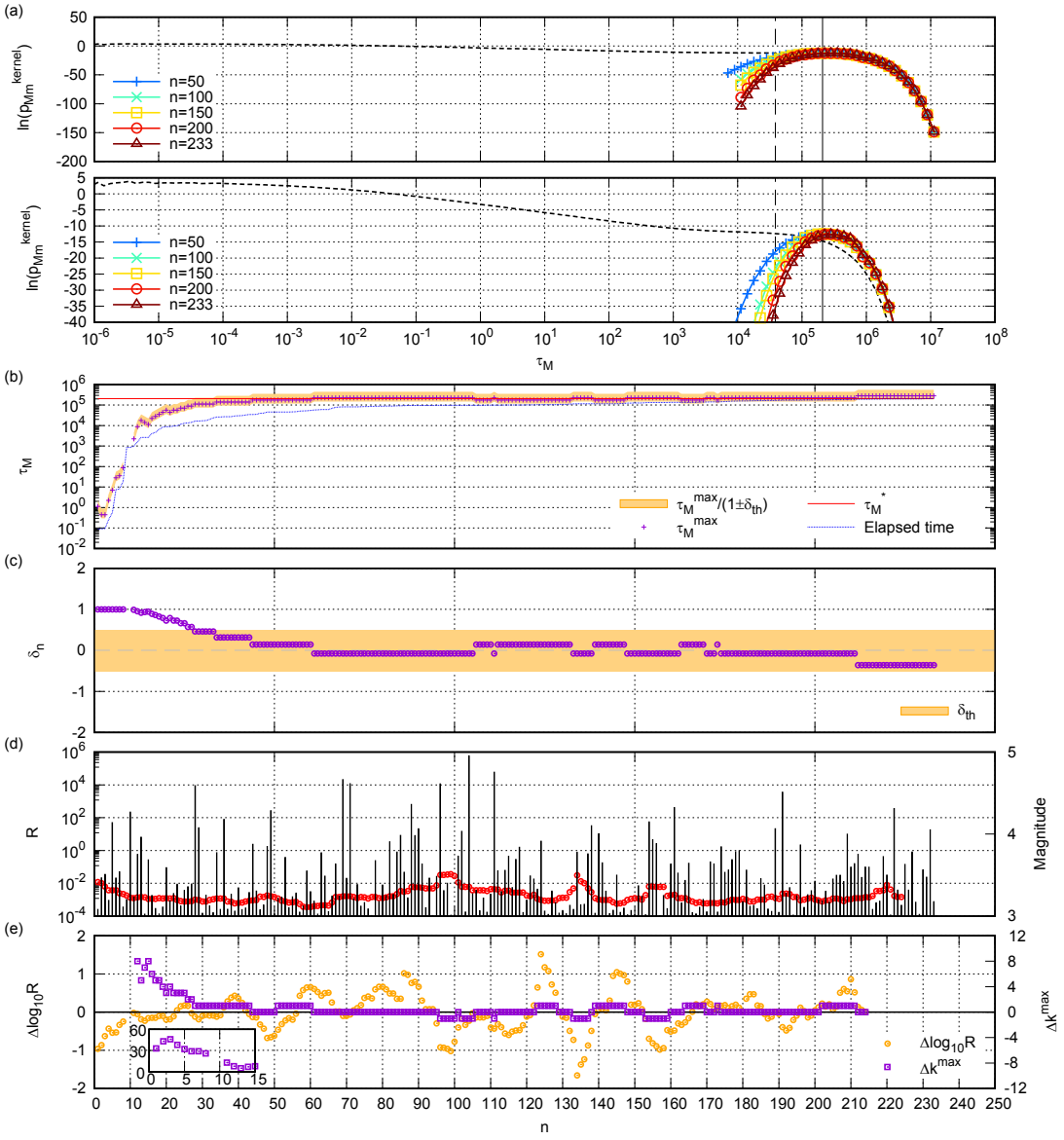

Fig. 12. (Color online) The second example for the case that τ_M^* is in regime (III). At some updates, the kernel part does not have the maximum peak and the estimate is not determined, which causes some jumps in the time series. The inset in (e) shows $\Delta k_n^{\max}$ at small update counts, indicating a rapid variation of $k^{\max,n}$ at small n. Other descriptions of the figure are the same as in Fig. 11.

around τ_M^*, and it consecutively satisfies $|\delta_n| \le \delta_{\rm th} = 0.5$ from long to immediately before the next large event. This is also confirmed by $\Delta k_n^{\max} \approx 0$ in Fig. 12(e). Thus in this case also the forecasting is judged to be effective.

Unlike these two examples, in the third example in Fig. 13 for $\tau_M^* \in$ (II), the time series is dominated by the non-stationary activity as shown in Figs. 13(d) and 13(e). Although $|\delta_n| \le \delta_{\rm th} = 0.5$ is satisfied only immediately before the large shock, $\tau_M^{\max,n}$ continues shifting and $|\delta_n| \le \delta_{\rm th}$ does not hold as shown in Figs. 13(b), 13(c), and 13(e). Thus, the forecasting is not effective in this case.

Although these are only examples and not all updating proceeded in these ways, the examples suggest that the stability of the estimate is related to the stationarity of the time series.

6.5 *Statistical analysis of the effectiveness of forecasting*

We show the results of the statistical analysis on the effectiveness of forecasting. Only the cases of $n_{\rm fin} \ge 30$ were used in the analysis to ensure that the temporal information of lower intervals was fully reflected in the estimate. Figure 14(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 $n_{\rm fin} \ge 30$, which is shown with the ratio to N. The updates included in these N_{30} upper intervals were analyzed.

Figures 14(b)–14(d) show the results of the statistical analysis with $\delta_{\rm th} = 0.5$. Figure 14(b) shows the probability ($P_{\rm fin}$) of $n_{\le {\rm th}} > 0$ (or $|\delta_{n_{\rm fin}}| \le \delta_{\rm th}$) for each τ_M^*. The average of $P_{\rm fin}$ for the overall τ_M^* is about 0.52, and the $P_{\rm fin}$ for each τ_M^* is about the same, except for $\tau_M^* > \langle \tau_M \rangle$ in which $P_{\rm fin}$ takes a higher probability around 0.67. Of such $n_{\le {\rm th}} > 0$ cases, the proportion ($P_{\ge 30}$) of those with relatively long $n_{\le {\rm th}} \ge 30$ is also shown in Fig. 14(b) [the probability distribution of $n_{\le {\rm th}}$ is shown in Fig. S13(a)[43]]. Thus the regions of high $P_{\ge 30}$ are overlapped with regimes (I) and (III), though the former is shifted toward larger τ_M^*. On the other hand, $P_{\ge 30}$ is lower in regime (II); it gradually decreases as τ_M^* gets larger. This is consistent with the average of $n_{\le {\rm th}}$ ($\langle n_{\le {\rm th}} \rangle$, this average

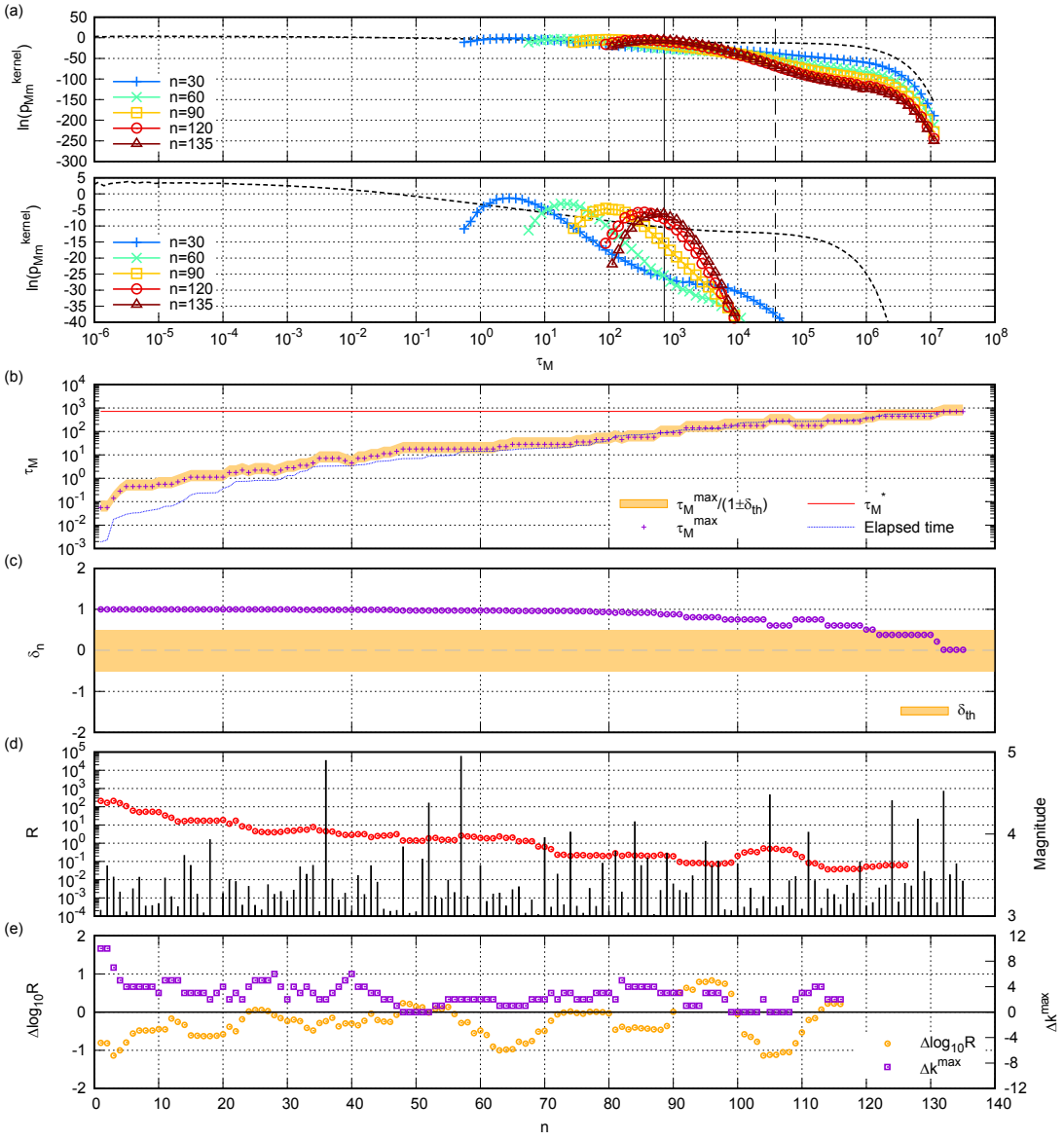

Fig. 13. (Color online) The third example for the case that τ_M^* is in regime (II). The descriptions of the figure are the same as in Fig. 11.

is taken for $n_{\le \mathrm{th}} > 0$), but also with the average of its proportion to n_{fin} ($\langle n_{\le \mathrm{th}}/n_{\mathrm{fin}} \rangle$) as shown in Fig. 14(c). This implies that, as the fraction of non-stationary times in $[0, \tau_M^*)$ increases in regime (II), the domination rate of $n_{\le \mathrm{th}}$ in the total n_{fin}-updates decreases gradually. These properties are preserved for $\delta_{\mathrm{th}} = 0.25$ (Fig. S14[43]).

Figure 15 shows the joint probability density-mass functions of $\Delta \log_{10} R$ and $\Delta k^{\max}$ calculated numerically for each τ_M^*. The case $k^{\max} = 80$ was excluded from the population. If τ_M^* is in the regions of high $P_{\ge 30}$, 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-mass 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 ($|\delta_{n_{\mathrm{fin}}}| \le \delta_{\mathrm{th}}$) 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 $n_{\le \mathrm{th}}$. 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 $n_{\le \mathrm{th}}$.

Finally, we discuss the effectiveness of forecasting in terms of the duration time ($\tau_{\le \mathrm{th}}$) during the $n_{\le \mathrm{th}}$ updates.

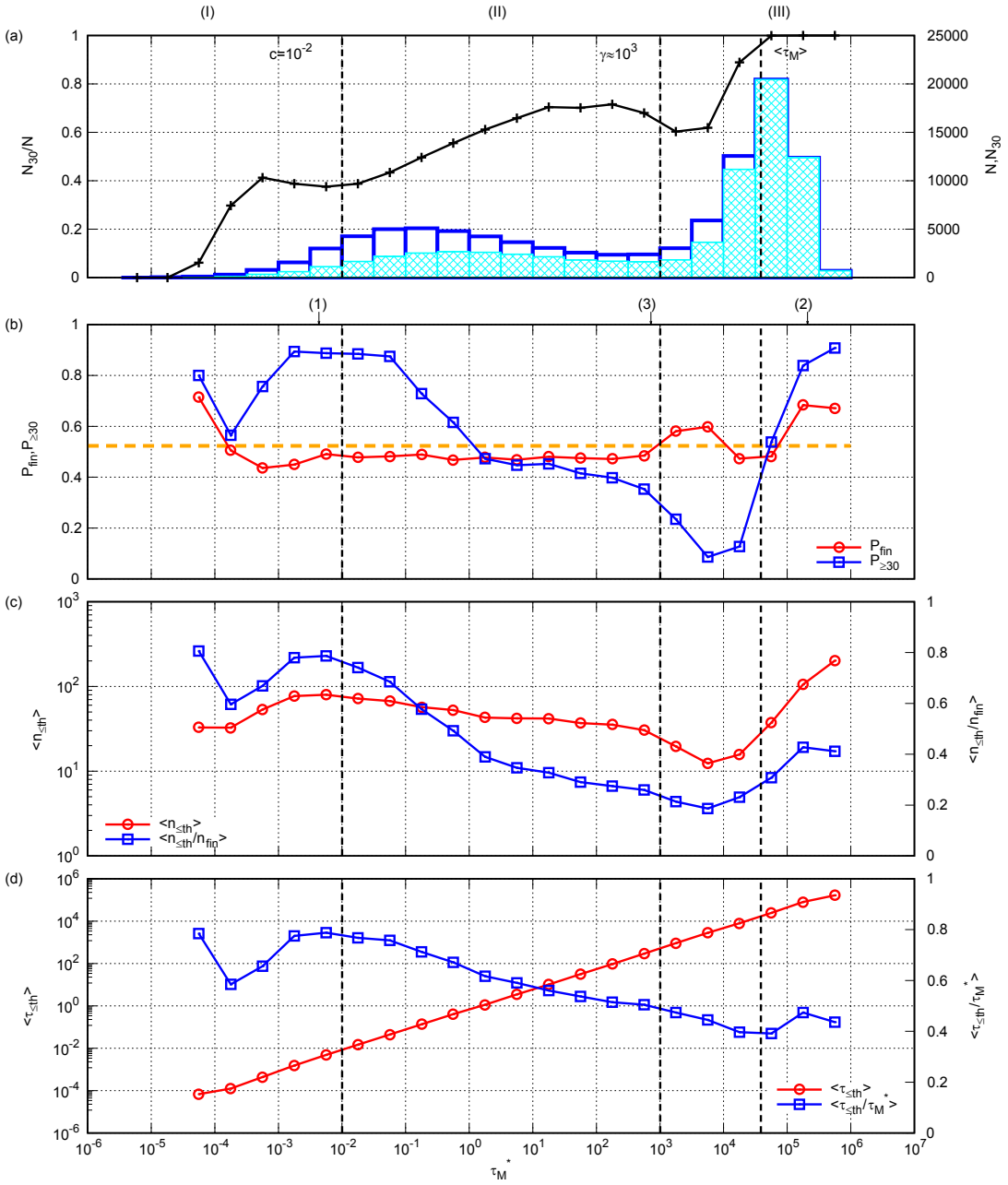

Fig. 14. (Color online) (a) (Blue bar) Number of upper intervals N, (Cyan shaded bar) number of upper intervals that include at least 30 updates N_{30}, and (Black points) their ratio N_{30}/N, for each τ_M^* in the test data. (b)–(d) Statistical results with $\delta_{\rm th} = 0.5$ for each τ_M^*. (1)–(3) indicates the τ_M^*'s of the examples in Figs. 11–13. (b) (Red circle) Probability $P_{\rm fin}$ that $n_{\le {\rm th}} > 0$ holds [or the probability that Eq. (50) 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 square) Proportion $P_{\ge 30}$ of such cases among them where $n_{\le {\rm th}} \ge 30$. (c) (Red circle) Average of $n_{\le {\rm th}}$, $\langle n_{\le {\rm th}} \rangle$, and (Blue square) the average of its proportion to the total number of updates, $\langle n_{\le {\rm th}}/n_{\rm fin} \rangle$. (d) (Red circle) Average of the duration time $\tau_{\le {\rm th}}$, $\langle \tau_{\le {\rm th}} \rangle$ and (Blue square) the average of its proportion to the actual occurrence time, $\langle \tau_{\le {\rm th}}/\tau_M^* \rangle$.

Figure 14(d) shows the average of the duration time ($\langle \tau_{\le {\rm th}} \rangle$) and the average of its ratio to the actual occurrence time ($\langle \tau_{\le {\rm th}}/\tau_M^* \rangle$) for each τ_M^* [the probability density of $\tau_{\le {\rm th}}$ is shown in Fig. S13(b)[43]]. Unlike $\langle n_{\le {\rm th}} \rangle$ in Fig. 14(c), $\langle \tau_{\le {\rm th}} \rangle$ increases linearly as τ_M^* gets larger, and it is sufficiently long in regime (III). On the other hand, $\tau_{\le {\rm th}}$ is very short in regime (I); however, the ratio $\langle \tau_{\le {\rm th}}/\tau_M^* \rangle$ is high (around 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

Bayes' theorem and Bayesian updating on the inter-event times at different magnitude thresholds in a marked point process were considered. The analytical results for the uncorrelated time series were 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, Bayes' theorem was considered for the general marked point process. 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 was represented by the generalized forms of the inter-event time distribution and the conditional probability density function.

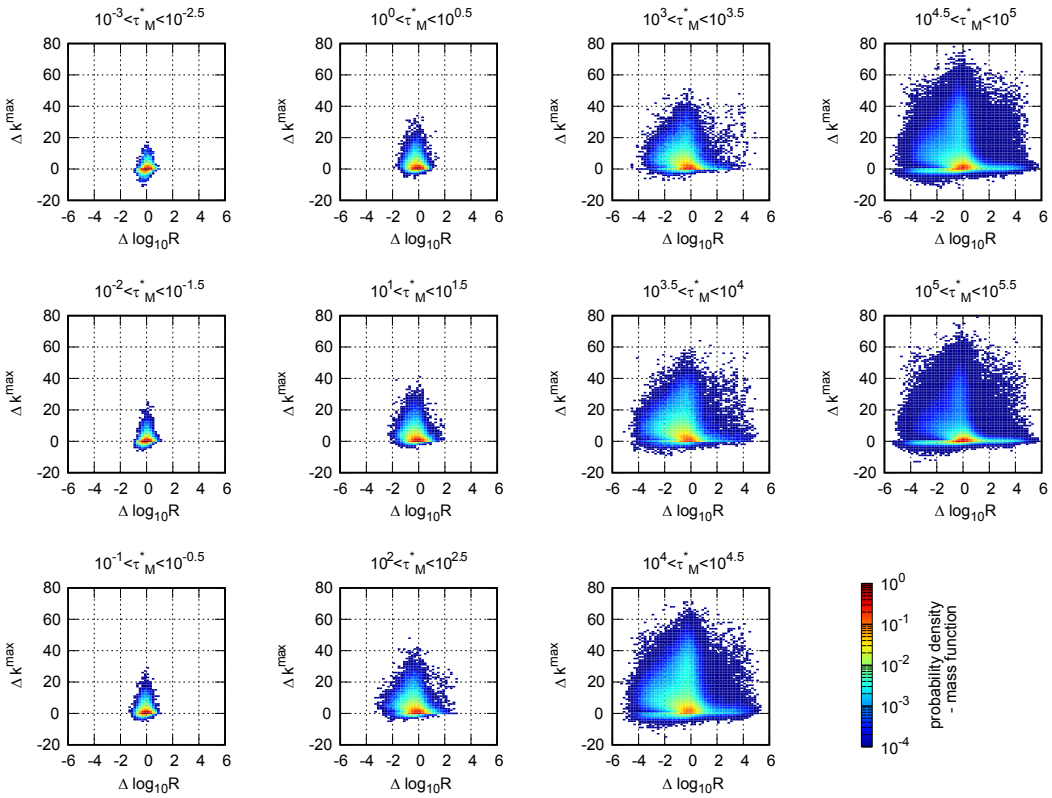

Fig. 15. (Color online) Joint probability density-mass functions of $\Delta \log_{10} R$ and $\Delta k^{\max}$ for each τ^*_M.

This inverse probability density function was derived for the uncorrelated time series analytically, and the condition to have a peak was also found.

Bayes' theorem was 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 considered 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 were derived, and the latter approximation was shown to be reasonable using the distances.

Bayesian updating was applied to the time series of the ETAS model. We numerically analyzed the approximation function and its kernel part. We used the maximum point of the kernel part as the estimate of when the larger-magnitude event than the upper threshold will occur because the maximum peaks of these two functions were shown to be not drastically different. The accuracy of the estimation at each update was evaluated by the relative error with the actual time the large event happened (τ^*_M); the effectiveness of the forecasting throughout the series of updates was judged by the continuity of the plausible estimations prior to the large event.

Statistical analysis indicated that the accuracy of the estimation at the last update was not drastically dependent on τ^*_M. This suggests that the inverse probability density function can estimate τ^*_M in response to the temporal pattern of minor events. However, the continuity of plausible estimation depended on τ^*_M. 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 was dominant either immediately after or long 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 was applied in 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 the non-stationary part. This probably is one of the reasons why forecasting was ineffective in the non-stationary regime.

In this study, we confirmed that the kernel part could be monomodal in the ETAS time series, which was not for the inter-event time distribution in our parameter setting, suggesting the advantage of the Bayesian approach for narrowing the possible range of the next large event timing down. Thus, we demonstrated qualitatively the superiority of the Bayesian approach to forecasting. However, this study did not compare their forecasting performance quantitatively. Probabilistic ways such as using the hazard function can make such a quantitative comparison, and thus it is necessary to further examine the inverse probability density or its approximation function in time series with correlations between events; a detailed analysis of the correction term

in the approximation function will be necessary. The kernel part enables us only the point estimate at the peak time, and therefore derivation of the entire approximation function is also significant for realizing point estimation by average, interval estimation, and probabilistic risk assessment in the Bayesian framework.

Although the statistical property of Bayesian updating was 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,[20] forecasting is considered to be less effective because in such time series, different mainshock–aftershocks sequences overlap[20] 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,[20] 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}); a larger branching ratio boosts aftershocks for a mainshock,[36] which increases the number of updates in Bayesian updating and thus is advantageous for forecasting.

In this study, only one lower threshold magnitude (m) was set for a given upper threshold (M). Although the lower threshold m can be set freely in between $[M_0, M)$, the theoretical result for the stationary marked Poisson process suggests that it is better to set m such that $\Delta m > \log_{10} 3/b$; under this condition, the inverse probability is monomodal other than the one by the delta function, and such a peak is convenient for the estimation of τ_M^*. 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. One approach for performing Bayesian updating using more temporal information of the lower intervals is to set multiple lower thresholds $[m_1 < m_2 < \cdots (< M)]$, all of them satisfy the condition $\Delta m > \log_{10} 3/b$. 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. It should be noted that 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.

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's capability to generate sufficient synthetic data with the parameter set determined for past seismic activity enables the preparation of precise statistical amounts necessary in the numerical Bayesian updating method. 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.

Another way to combine with the ETAS model is to replace the prior distribution of the Bayesian updating with the inter-event time distribution derived from the ETAS model's conditional intensity function; such replacement would incorporate the seismicity information before the last major earthquake in the Bayesian approach. From the viewpoint of probabilistic forecasting with the conditional intensity function, this replacement can be a way to improve forecasting by taking into account the correlation between a major earthquake and its preceding seismic pattern, which is typically not considered,[45] by combining the Bayesian approach. Note that, because this statement is based on the assumption that such a correlation between a seismic pattern and its following large shock that the ETAS model does not cover (for example, Refs. 46 and 47) can be managed by the conditional probability, this approach is not compatible with the method described in the previous paragraph to compensate for the shortage of data utilizing the ETAS model. Further examination of the Bayesian approach using actual seismic catalog data is necessary to clarify the credibility of the assumption and effectiveness of the approach.

Finally, extending the Bayesian approach to a spatiotemporal version is an important issue for more practical forecasting while incorporating spatial seismic features, which is discussed in Ref. 48.

Acknowledgments The idea and derivation of Eqs. (7) and (8) were first discussed by Y. Aizawa. This work was supported by JST SPRING, Grant Number JPMJSP2110.

Appendix A: Derivation of the Conditional and Inverse Probability Density Functions for the Uncorrelated Time Series

First, we derive the conditional probability density function [Eq. (16)] by substituting Eqs. (13)–(15) into Eq. (12). The denominator of Eq. (12) 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 (16) is obtained by rearranging the above equations.

We confirm that Eq. (2) with this conditional probability in its kernel has the exponential distribution [Eq. (11)] as the solution. By dividing both sides of Eq. (2) by N_m and rewriting it using $N_M/N_m = \langle\tau_m\rangle/\langle\tau_M\rangle$ as well as Eq. (16)

$$p_m(\tau_m) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\int_{\tau_m}^{\infty}\left[e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m) + \frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\left(A_{\Delta m}\frac{\tau_M-\tau_m}{\langle\tau_M\rangle}+2\right)\theta(\tau_M-\tau_m)\right]p_M(\tau_M), \tag{A·1}$$

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 Eq. (A·1) is equivalent to the left hand side (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 Eq. (A·1) and note that $A_{\Delta m}+1 = \langle\tau_M\rangle/\langle\tau_m\rangle$; the integral involving the delta function (R_1) is

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

and the integral involving the step function (R_2) is

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

Therefore, the r.h.s. of Eq. (A·1) is shown to be equivalent to the l.h.s. of Eq. (A·1) as follows:

$$\begin{aligned} R_1+R_2 &= \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2}(1+A_{\Delta m})^2e^{-\frac{\tau_m}{\langle\tau_m\rangle}} \\ &= \frac{1}{\langle\tau_m\rangle}e^{-\frac{\tau_m}{\langle\tau_m\rangle}}. \end{aligned} \tag{A·4}$$

Second, we derive the inverse probability density function [Eq. (17)]. From Eq. (16), the generalized probability density functions for the uncorrelated time series are derived as

$$z_m(\tau_m) = \frac{\tau_m}{\langle\tau_m\rangle^2}e^{-\frac{\tau_m}{\langle\tau_m\rangle}},$$

$$z_M(\tau_M) = \frac{\tau_M}{\langle\tau_M\rangle^2}e^{-\frac{\tau_M}{\langle\tau_M\rangle}},$$

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

Equation (17) is obtained by substituting the above equations in Eq. (9).

Derivative of Eq. (17) by τ_M is

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

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 Eq. (25) from Eq. (24)

The summation part in the r.h.s. of Eq. (24) is

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

The first term on the r.h.s. of the above equation is transformed by substituting Eqs. (13)–(15) as

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

The second term except the step function is also transformed as

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

Finally, Eq. (25) is obtained by substituting Eqs. (B·1) and (B·2) in Eq. (24), with the denominator of the r.h.s. of Eq. (24)

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

Appendix C: Another Bayesian Updating Method

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

First, we derive $N^*_{mM}(\tau_M,\tau_m^{(1)},\ldots,\tau_m^{(n)})$ by counting the total number of the upper intervals of length τ_M that include the leftmost consecutive lower intervals of lengths $\{\tau_m^{(1)},\ldots,\tau_m^{(n)}\}$. 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 $\{\tau_m^{(1)},\ldots,\tau_m^{(n)}\}$ 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

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

This equation is rewritten using Eq. (11) in the explicit form as

$$N^*_{mM}(\tau_M,\tau_m^{(1)},\ldots,\tau_m^{(n)}) = N_M\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}\frac{1}{\langle\tau_m\rangle^n}e^{-\frac{\sum_{i=1}^n\tau_m^{(i)}}{\langle\tau_m\rangle}}p^*_{Mm}(\tau_M|\tau_m^{(1)},\ldots,\tau_m^{(n)})\,d\tau_M\,d\tau_m^n. \tag{C·1}$$

Second, we derive $N^*_{mM}(\tau_M,\tau_m^{(1)},\ldots,\tau_m^{(n)})$ 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 $\{\tau_m^{(1)},\ldots,\tau_m^{(n)}\}$ to be involved in each of the $N_Mp_M(\tau_M)\,d\tau_M$ 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

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

This equation is rewritten in the explicit form using Eqs. (13)–(15) in the same way as in Appendix B.

$$\begin{aligned}N^*_{mM}(\tau_M,\tau_m^{(1)},\ldots,\tau_m^{(n)}) &= N_M\frac{1}{\langle\tau_M\rangle}e^{-\frac{\tau_M}{\langle\tau_M\rangle}}\,d\tau_M\,d\tau_m^n\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)^{n-1}\\ &\times\left[e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta\left(\tau_M-\sum_{i=1}^n\tau_m^{(i)}\right)+\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)e^{-A_{\Delta m}\frac{\sum_{i=1}^n\tau_m^{(i)}}{\langle\tau_M\rangle}}\theta\left(\tau_M-\sum_{i=1}^n\tau_m^{(i)}\right)\right].\end{aligned} \tag{C·2}$$

Finally, $p^*_{Mm}(\tau_M|\tau_m^{(1)},\ldots,\tau_m^{(n)})$ is derived from Eqs. (C·1) and (C·2) as

$$p^*_{Mm}(\tau_M|\tau_m^{(1)},\ldots,\tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\left[\frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-\frac{\tau_M-\sum_{i=1}^n\tau_m^{(i)}}{\langle\tau_M\rangle}}\theta\left(\tau_M-\sum_{i=1}^n\tau_m^{(i)}\right)+e^{-\frac{\tau_M-\sum_{i=1}^n\tau_m^{(i)}}{\langle\tau_m\rangle}}\delta\left(\tau_M-\sum_{i=1}^n\tau_m^{(i)}\right)\right]. \tag{C·3}$$

This is different from Eq. (25), which reflects the difference whether the position of lower intervals is specified.

Appendix D: Derivation of Eq. (30)

First, we substitute Eqs. (13)–(15) into Eq. (29)

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

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

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

Finally, Eq. (30) is obtained by rearranging the above equations.

Appendix E: Derivation of Eq. (35)

In this appendix, $P(\tau_m^{(1)},\ldots,\tau_m^{(l)}|\tau_M)$ is derived for the uncorrelated time series. First, we divide the summation in Eq. (34) into two parts:

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

This equation is further rewritten by substituting Eqs. (13)–(15) in the same way as in Appendix B. The second term on the r.h.s. except for the step function is

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

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

Therefore

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

Finally, because $\tau_M \gneqq \sum_{i=1}^{l}\tau_m^{(i)}$ holds for $l < n$ by the condition of Eq. (27)

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

Appendix F: Derivation of Eq. (40)

In this appendix, the approximation function of the inverse probability density function for the uncorrelated time series [Eq. (40)] is derived. By substituting Eqs. (11), (16), and (35) into Eq. (37)

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

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 ($\{\cdots\}$ and $[\cdots]$) in Eq. (F·1) are respectively transformed as

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

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

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

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

$$= \left(A_{\Delta m}\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^n \prod_{i=1}^{n} e^{\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}. \tag{F·3}$$

Finally, Eq. (40) is derived by substituting Eqs. (F·2) and (F·3) into Eq. (F·1).

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 [Eq. (24)] and Eq. (37), i.e., the approximations made on Eq. (24) that correspond to the assumptions made in Sect. 4.2 to derive Eq. (37).

The summation in Eq. (24) can be decomposed into

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

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

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

The first term on the r.h.s. of Eq. (G·1) is equivalent to $-(n-1)\prod_{i=1}^{n}P_i$ (Appendix E), and then, this term formally coincides with the correction term in Eq. (37). Therefore, the second term corresponds to the kernel part [Eq. (38)].

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 Eq. (G·1) 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 $(i-j+1)$, $\rho_{mM}(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)})$.

These constraints are relaxed in the derivation in Sect. 4.2. In the view in Sect. 4.2, the upper intervals of length τ_M are collected and the new time series is generated as shown in Fig. 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 Eq. (G·1), i.e., the correction term in Eq. (37). (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 [Eq. (G·2)]. In addition, the constraint on the number of division can be eliminated by taking the average [Eq. (G·3)].

$$\rho_{mM}\left(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right) \approx \rho_{mM}(\tau_m^{(j)}|i-j+1,\tau_M) \tag{G·2}$$

$$\approx \frac{\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(j)}|i,\tau_M)}{\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M)} \quad \text{(G·3)}$$
$$= p_{mM}(\tau_m^{(j)}|\tau_M).$$

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

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

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

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 [Eq. (25)] and the inter-event time distribution $[p_M(\tau_M)]$

$$D(p_{Mm}\|p_M) := \int_T^{\infty} |p_\theta(\tau_M,T)-p_M(\tau_M)|^2\, d\tau_M, \quad \text{(H·1)}$$

where

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

By substituting these functions in Eq. (H·1), the distance is derived as

$$D(p_{Mm}\|p_M) = \frac{\langle\tau_M\rangle C_1^2}{4}+\frac{C_1C_2(T)}{2}+\frac{C_2(T)^2}{2\langle\tau_M\rangle}, \quad \text{(H·2)}$$

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 $\langle D'(p_{Mm}^{\text{approx}}\|p_M)\rangle$ and $\langle D''(p_{Mm}^{\text{approx}}\|p_M)\rangle$ at Large T

In this appendix, we examine the cause of the separation between $\langle D'(p_{Mm}^{\text{approx}}\|p_M)\rangle$ and $\langle D''(p_{Mm}^{\text{approx}}\|p_M)\rangle$ at long-elapsed time T. Let us compare Fig. 7 to Figs. S6(a) and S6(c)[43] for $\mathcal{N}=10^5$. The separation is suppressed compared to that shown in Fig. 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 (45), and to the improvement of the calculation of distance in Eq. (47).

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

$$\boldsymbol{P}_1' = \left[[P_{1,jk}]_{j=j_{\min}^{(k)},\ldots,j_{\max}^{(k)}}\right]_{k=k_{\min},\ldots,k_{\max}-l_c}. \quad \text{(I·1)}$$

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

Figures S6(b) and S6(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 Figs. S6(a) and S6(c),[43] 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.

*tanaka.hiroki.43s@st.kyoto-u.ac.jp
†umeno.ken.8z@kyoto-u.ac.jp

1) C. H. Scholz, *The Mechanics of Earthquakes and Faulting* (Cambridge University Press, Cambridge, U.K., 2002) 2nd ed., pp. 408–412.
2) The Headquarters for Earthquake Research Promotion [https://www.jishin.go.jp/reports/research_report/choukihyoka_01b/] [in Japanese], last access: November 2023.
3) P. Bak, K. Christensen, L. Danon, and T. Scanlon, Phys. Rev. Lett. **88**, 178501 (2002).
4) Á. Corral, Phys. Rev. E **68**, 035102 (2003).
5) Á. Corral, Physica A **340**, 590 (2004).
6) J. Davidsen and C. Goltz, Geophys. Res. Lett. **31**, L21612 (2004).
7) E. Lippiello, C. Godano, and L. de Arcangelis, Phys. Rev. Lett. **98**, 098501 (2007).
8) Y. Aizawa, Bussei Kenkyu **97**, 309 (2011).
9) Y. Aizawa, T. Hasumi, and S. Tsugawa, Nonlinear Phenom. Complex Syst. **16**, 116 (2013).
10) Y. Aizawa and S. Tsugawa, *Nonlinear Phenomena in Complex Systems: From Nano to Macro Scale* (Springer, Dordrecht, 2014) p. 21.
11) B. Gutenberg and C. F. Richter, Bull. Seismol. Soc. Am. **34**, 185 (1944).
12) F. Omori, J. Coll. Sci., Imp. Univ. Tokyo, Jpn. **7**, 111 (1894).
13) T. Utsu, Geophys. Mag. **30**, 521 (1961).
14) Á. Corral, Phys. Rev. Lett. **92**, 108501 (2004).
15) G. Molchan, Pure Appl. Geophys. **162**, 1135 (2005).
16) S. Hainzl, F. Scherbaum, and C. Beauval, Bull. Seismol. Soc. Am. **96**, 313 (2006).
17) A. Saichev and D. Sornette, Phys. Rev. Lett. **97**, 078501 (2006).
18) A. Saichev and D. Sornette, J. Geophys. Res.: Solid Earth **112**, B04313 (2007).
19) D. Sornette, S. Utkin, and A. Saichev, Phys. Rev. E **77**, 066109 (2008).
20) S. Touati, M. Naylor, and I. G. Main, Phys. Rev. Lett. **102**, 168501 (2009).
21) M. Bottiglieri, L. de Arcangelis, C. Godano, and E. Lippiello, Phys. Rev. Lett. **104**, 158501 (2010).
22) E. Lippiello, Á. Corral, M. Bottiglieri, C. Godano, and L. de Arcangelis, Phys. Rev. E **86**, 066119 (2012).
23) L. de Arcangelis, C. Godano, J. R. Grasso, and E. Lippiello, Phys. Rep. **628**, 1, Chap. 3 (2016).
24) Y. Ogata, J. Am. Stat. Assoc. **83**, 9 (1988).

25) Y. Ogata, Tectonophysics **169**, 159 (1989).
26) Y. Ogata, Ann. Inst. Stat. Math. **50**, 379 (1998).
27) Y. Ogata, K. Katsura, and M. Tanemura, J. R. Stat. Soc., Ser. C **52**, 499 (2003).
28) Y. Ogata and J. Zhuang, Tectonophysics **413**, 13 (2006).
29) Y. Ogata, Earth Planets Space **74**, 110 (2022).
30) A. Helmstetter and D. Sornette, J. Geophys. Res.: Solid Earth **107**, 2237 (2002).
31) A. Helmstetter and D. Sornette, Geophys. Res. Lett. **30**, 2069 (2003).
32) A. Helmstetter and D. Sornette, J. Geophys. Res.: Solid Earth **108**, 2457 (2003).
33) A. Helmstetter, D. Sornette, and J.-R. Grasso, J. Geophys. Res.: Solid Earth **108**, 2046 (2003).
34) T. Utsu and A. Seki, Jishin (Journal of the Seismological Society of Japan. 2nd ser.) **7**, 233 (1955) [in Japanese].
35) A. Helmstetter, Phys. Rev. Lett. **91**, 058501 (2003).
36) A. Helmstetter and D. Sornette, Geophys. Res. Lett. **30**, 1576 (2003).
37) D. Vere-Jones, Int. J. Forecasting **11**, 503 (1995).
38) H. Tanaka and Y. Aizawa, J. Phys. Soc. Jpn. **86**, 024004 (2017).
39) H. Tanaka and K. Umeno, ESS Open Archive **December 16** (2021) [DOI: 10.1002/essoar.10509635.1].
40) Y. Tohru and O. Atsuyuki, Sogo Toshi Kenkyu (Comprehensive urban studies) **43**, 99 (1991) [in Japanese].
41) D. J. Webb, J. Theor. Biol. **43**, 277 (1974).
42) Y. Y. Kagan, Bull. Seismol. Soc. Am. **94**, 1207 (2004).
43) (Supplemental Material) Supplemental figures on numerical results are provided online.
44) T. Omi and S. Nomura, *Tenkatei no Jikeiretsukaiseki* (Time Series Analysis for Point Processes) (Kyoritsu, Tokyo, 2021) Chap. 7 [in Japanese].
45) Y. Ogata, K. Katsura, H. Tsuruoka, and N. Hirata, Seismol. Res. Lett. **89**, 1298 (2018).
46) E. Lippiello, L. de Arcangelis, and C. Godano, Phys. Rev. Lett. **100**, 038501 (2008).
47) E. Lippiello, C. Godano, and L. de Arcangelis, Geophys. Res. Lett. **39**, L05309 (2012).
48) H. Tanaka and K. Umeno, J. Phys. Soc. Jpn. **92**, 113001 (2023).