Ph.D. Thesis

Analysis of Hierarchical Structure of Seismic Activity: Bayesian Approach to Forecasting Earthquakes

Hiroki Tanaka

Department of Applied Mathematics and Physics
Graduate School of Informatics
Kyoto University

Thesis supervisor:
Ken Umeno

February 2024

Contents

Chapter 1

Introduction

Some parts of this chapter are based on the introductory parts of Refs. [1, 2, 3, 4], each corresponds to papers 1, 2, 3, and 4 in the list of author's papers, respectively.

1.1 Seismic Activity, Laws, and Modelling

An earthquake is an instantaneous fracture process on a fault plane to release the strain energy slowly accumulated originating from plate tectonics [Section 1.8 in Ref. [5]]. The basic understanding of earthquake occurrence can be simplified in this way. However, the interactions between earthquakes, i.e., the effect of an earthquake on other earthquakes' occurrence at other faults, typically due to the Coulomb stress change by an earthquake that causes the generation or suppression of subsequent earthquakes, make the occurrence of earthquakes complicated [Section 4.5 in Ref. [6], Section 5 in Ref. [7]]. As a result of such interactions, the collective occurrence pattern of earthquakes, referred to as seismicity or seismic activity, shows complexity; some studies have indicated that the networks generated from seismic activity manifest complexity, such as scale-free and small-world natures [8, 9, 10, 11].

In analyzing seismic activity, each earthquake is characterized by a few representative quantities; in most cases, it is a combination of an earthquake's time of occurrence, spatial location such as hypocenters or epicenters[1], and the magnitude to quantify the size of an earthquake [12, 13] [2]. Information on such representatives is summarized and published in seismic catalogs [7, 12, 13]. Seismic catalogs include local catalogs for earthquakes in a local area, such as the JMA catalog of earthquakes around Japan [16, 17] and the Southern California catalog [18, 19], and global catalogs that include earthquakes on a global scale, such as the Preliminary Determination of Earthquakes (PDE) catalog [20] and the global Centroid Moment Tensor (CMT) catalog [21, 22, 23].

Figure 1.1 shows an example of the space-time-magnitude information using the local JMA catalog in the area around the mainshock of the Tohoku earthquake in March 2011 (140° – 146°E and 35° – 42°N), from 01/01/2000 – 03/31/2022 [16, 17]. As this figure shows, seismic activity can be expressed as a point process with marks representing magnitude values in space-time [Section 4 in Ref. [7]] as conceptually drawn in Fig. 1.2. Figure 1.1(a) shows the earthquakes' ($\geq M3$) epicenters in this space-time domain by colored circles; the color and the size of circles reflect the value of each earthquake's magnitude. Also, Figure 1.1(b) is the projection of hypocenters onto a cross-section of longitude × depth. The magnitude of each earthquake is indicated by color and symbol size as in Fig. 1.1(a). Hypocenters whose depth increases as the longitude decreases and approaches the Japanese island indicate the subduction of the Pacific plate under the Okhotsk plate [pp. 65-70 in Ref. [24]]. Figures 1.1(c) – (e) show the time series of magnitude; each earthquake is represented as a bar at the occurrence time with magnitude value as the height. In Figs. 1.1(c) – (e), also the cumulative number of earthquakes ($N(t)$ against time since the origin time t) with magnitude $\geq M3$ are shown. As shown in Figs. 1.1(c) and (e), after the large earthquake ($M9$ Tohoku mainshock), seismicity is activated, and many earthquakes follow around the

[1] The epicenter is the latitude and longitude of the starting location of a rupture. The hypocenter further includes depth information [p.2 in Ref. [5]] [12].

[2] Magnitude has several definitions, including based on the body or surface wave in seismic waves [Section 5.2 in Ref. [5]] [12, 13]. Such magnitudes become inaccurate by saturation at a large value, though the moment magnitude can avoid such saturation [Section 5.2 in Ref. [5]][13, 14, 15].

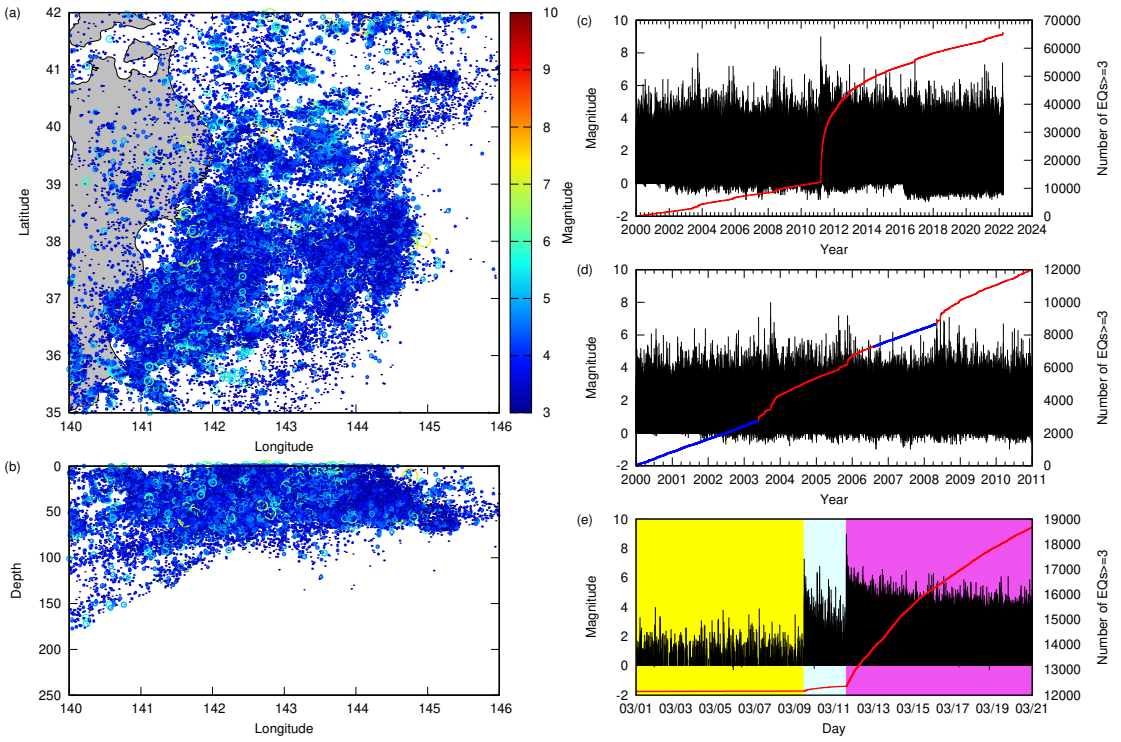

Figure 1.1: Seismic activity in the area 140° – 146°E and 35° – 42°N, from 01/01/2000 – 03/31/2022, drawn using the JMA catalog [16, 17]. (a) Distribution of epicenters of earthquakes in the above space-time domain. Only the earthquakes with magnitude $\geq M3$ are shown. The size of the circle and color reflect the magnitude of each earthquake. The map was drawn using the file in [http://www.gnuplotting.org/plotting-the-world-revisited/]. (b) Distribution of hypocenters with magnitude $\geq M3$. (c) Time series of magnitude of earthquakes in the spatial domain. The cumulative number of earthquakes $\geq M3$ from 01/01/2000 is also shown by the red curve. (d) and (e) show the enlarged versions of (c) for the period 01/01/2000 – 12/31/2010 and 03/01/2011 – 03/20/2011, respectively. In (d), the cumulative number of earthquakes in the time domains (01/01/2000 – 05/26/2003 and 08/01/2006 – 04/30/2008) in which seismic activity can be regarded as stationary is shown by thick blue lines (the same as JS in Chapter 5). In (e), the time domain of aftershock activity of the M9 mainshock is shown by magenta, the foreshock activity following the M7.3 earthquake by cyan, and before the foreshock domain by yellow (After Refs. [1, 25, 26]).

mainshock rupture zone; such subsequent earthquakes are called aftershocks [pp. 3-4, p.189 in Ref. [5]]. Occasionally, as shown in Fig. 1.1(e), a large earthquake (and its following seismic activity) is observed prior to the mainshock; these are called foreshocks, though they do not always precede the mainshock [pp. 3-4 in Ref. [5]]. It is only after the occurrence of a mainshock (i.e., the occurrence of a subsequent larger earthquake than foreshocks) that such earthquakes are confirmed to be foreshocks, and it is a crucial open problem from the viewpoint of forecasting whether the foreshocks and mainshock-aftershocks or swarms can be distinguished in real-time [pp. 322-323 in Ref. [5], pp. 116-117 in Ref. [24]]. Figure 1.1(d) shows seismic activity before the Tohoku mainshock. Several mainshock-aftershock sequences can be identified, though other than such activations, the activity is stationary with a nearly constant occurrence rate (the slope of the cumulative number of earthquakes is almost constant, which is prominent in the time domains indicated by bold blue lines in Fig. 1.1(d)).

Figure 1.3 shows another example of the epicenter distribution and magnitude time series for the case of the global PDE catalog [20], from 01/01/1973 to 12/31/2023. Figure 1.3(a) shows that most large earthquakes occur in a limited area at the boundary of plates. By limiting the earthquakes to those with large magnitudes, aftershock activities observed in the local catalog are less noticeable, and the occurrence rate of earthquakes with $\geq M5$ seem almost stationary as in Fig. 1.3(b). In this thesis, the

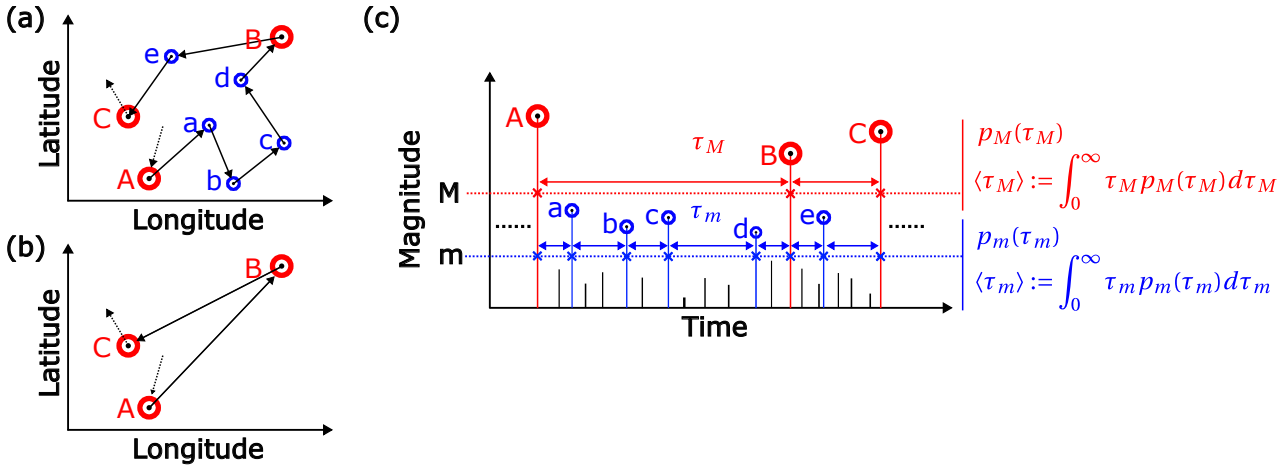

Figure 1.2: Conceptual diagram of a marked point process to represent seismic activity. (a, b) Jump process of epicenters in a spatial domain at different cut-off magnitudes (m for (a) and $M(:= m + \Delta m,\ \Delta m > 0)$ for (b)). In the figure, A, B, and C are events with magnitude $\geq M$ and a, b, ..., e are between m and M. With different cut-off magnitudes, the jump process is different, and thus, the epicentral network looks very different. (c) Corresponding time series of magnitude. A point (indicated by $\times$ symbol) process is determined by setting a cut-off magnitude. The interval between points $\times$ is called the inter-event time and expressed by τ; for the point process at cut-off magnitude m (M), the inter-event time is represented by τ_m (τ_M), and its probability density, the inter-event time distribution, is by $p_m(\tau_m)$ ($p_M(\tau_M)$); their average is expressed by $\langle\tau_m\rangle$ ($\langle\tau_M\rangle$). The inter-event time intervals at the lower cut-off magnitude (m) can be obtained as the partition of the divided inter-event time interval at the upper cut-off magnitude (M).

seismic time series with an almost constant occurrence rate as in whole activity in Fig. 1.3(b) and the blue-colored parts in Fig. 1.1(d) are referred to as the stationary time series, the aftershocks following major earthquakes as the aftershock sequence, and their mixture or superposition as in whole activity in Fig. 1.1(d) as the mixture time series [3].

The seismic activities shown in these figures are apparently complex and random. However, seismology has found some characteristic patterns and statistical laws in seismic activity. Spatiotemporal pattern includes the aftershocks, the swarms [Section 7.3 in Ref. [5]], the emergence of the seismic gap in space and seismic quiescence in time [Section 7.6A in Ref. [5]], doughnut-shaped pattern [Section 7.6A in Ref. [5]], and activation by foreshocks [Section 7.2 in Ref. [5]] (as shown in Fig. 1.1(e)) around the future rupture zone before large earthquakes [Section 7.2.2 in Ref. [6]] [27].

The most notable statistical laws are the Omori – Utsu (OU hereafter) [28, 29, 30] and Gutenberg – Richter (GR hereafter) [31] laws. The OU law is about the temporal property of seismicity; the number of aftershocks per day at time t from the mainshock ($R(t)$) obeys inverse power-law [28, 29, 30]:

$$R(t) = \frac{K}{(t+c)^p}, \tag{1.1}$$

where K, c ($\lesssim 1$ day), and p (≈ 1) are parameters determined for each aftershock sequence, in particular c and p are referred to as the c-value and p-value, respectively. Figure 1.4(a) shows an example of the OU law for the aftershocks ($\geq M4$) of Tohoku mainshock in Fig. 1.1. This example indicates that aftershocks occurred at a high occurrence rate immediately after the mainshock ($t \lesssim c$ (≈ 0.362) days), and as time passes ($t \gg c$), the occurrence rate decays in a power-law manner ($R(t) \approx K/t^p$). As suggested by such a power-law decay, an aftershock activity can continue for a long period of time [30]. It has widely been observed that major earthquakes are followed by aftershocks whose decay rate obeys OU law [29, 30].

The other recognized seismic law, the GR law, is about the frequency of magnitude, which denotes that the distribution function of magnitude ($P(M)$, the probability for magnitude $\geq M$, Prob(magnitude $\geq$

[3] This category of seismic time series and their judgment way is based on the proceeding studies on the scaling universality of the inter-event time distributions in a seismic activity described in Section 1.3.1.

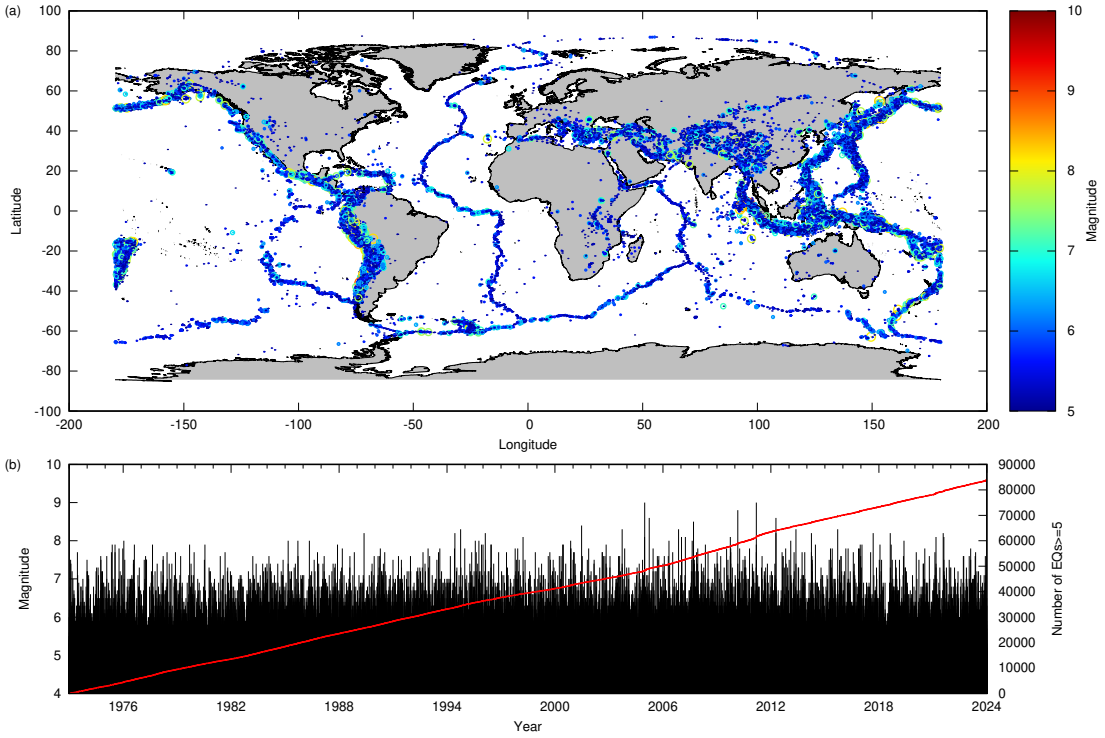

Figure 1.3: Global seismic activity in the PDE catalog [20], during 01/01/1973 – 12/31/2023. (a) Epicenter distribution of earthquakes $\geq M5$. The magnitude of each earthquake is indicated by the size of the circle and color. The file in [http://www.gnuplotting.org/plotting-the-world-revisited/] was used to draw the map. (b) Time series of magnitude for earthquakes $\geq M5$ and cumulative number of earthquakes $\geq M5$ from 01/01/1973.

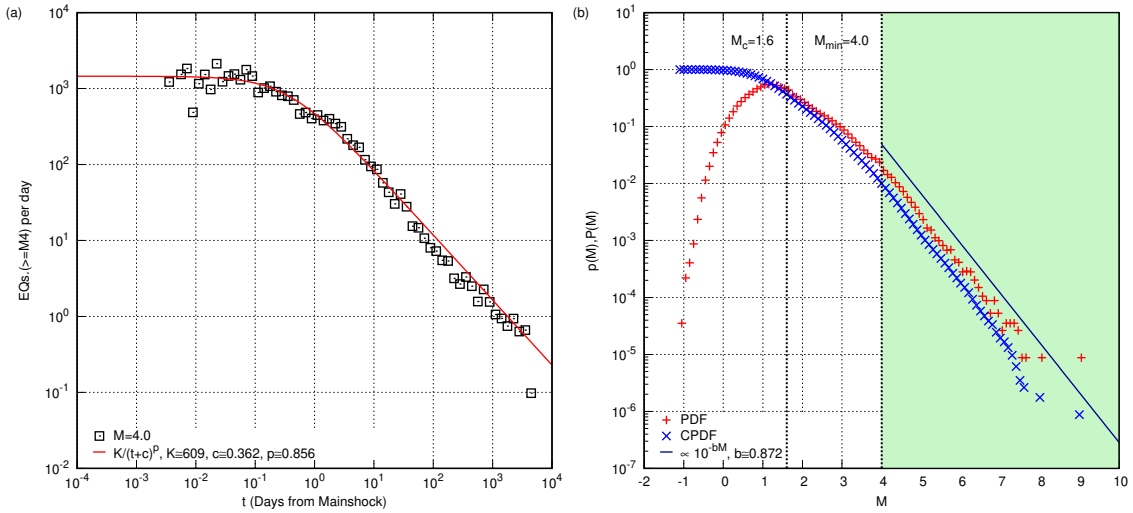

Figure 1.4: An example of (a) the OU law and (b) the GR law in the JMA catalog. (a) The frequency ($R(t)$) of aftershocks (earthquakes $\geq M4$ within spatial domain 140° – 146°E, 35° – 42°N, and time domain since the mainshock time until 03/31/2022) versus time t since the Tohoku mainshock occurrence. The fitting curve by the OU law in Eq. (1.1) is shown by the red curve. (b) (Red, + symbols) The probability density of magnitude ($p(M)$) and (blue, × symbols) the distribution of magnitude ($P(M)$) for earthquakes within 140° – 146°E and 35° – 42°N and from 01/01/2000 to 03/31/2022. The GR law in Eq. (1.2) is shown by the solid line. Although the completeness magnitude M_c calculated by the MBASS method [32] was 1.6, the author judged the magnitude range in which $p(M)$ obeys the straight line suggested by the GR law as $\geq M4$ (indicated by the green-colored background). The b-value (≈ 0.872) was determined for magnitude $\geq M4$ by the maximum likelihood method [33, 34].

M)) obeys the exponential distribution [31]:

$$P(M) = 10^{-b(M-M_{\min})}. \tag{1.2}$$

Here $M_{\min}$ is the minimum magnitude to be considered, and b (≈ 1 for ordinary seismic activity) is a parameter characterizing the magnitude distribution called the b-value [Section 5.4B in Ref. [5]]. Figure 1.4(b) shows an example of the probability distribution of magnitudes for the time series of the JMA catalog in Fig. 1.1(c). The probability density function of magnitudes $p(M)$ ($P(M) = \int_M^\infty p(m)dm$) is also shown in the figure [4]. In this figure, the magnitude distribution shows an exponential GR decay for $M \geq 4$, though it does not follow the exponential distribution for smaller magnitudes. Such deviations from the GR law at small magnitude are interpreted as indicating that such small earthquakes are undetected [35], i.e., the seismic catalog is incomplete at the magnitude range [35, 36]. Therefore, when analyzing seismic catalogs, the data analysis is limited to the range of magnitudes for which the magnitude frequency obeys the GR law; the infimum magnitude following the GR law is regarded to be the completeness magnitude (M_c), and earthquakes larger than or equal to M_c are considered to be complete [35]. The setting of M_c is crucial in order to correctly capture the nature of the seismic activity, removing artifacts caused by catalog incompleteness [35, 36]. There have been proposed several methods to determine M_c based on the GR law [35] [5]. Among them, the result of M_c calculated by the MBASS (Median-Based Analysis of the Segment Slope) method [32] ($M_c = 1.6$) is shown in Fig. 1.4(b). However, looking at Fig. 1.4(b), it appears that the linear decay of the exponential distribution suggested by the GR law seems to follow in the larger magnitude range ($M \geq 4$). Thus, in this thesis, the range following the GR law is often determined by appearance; Chapter 2 determines the magnitude range only by sight, and Chapter 5 determines it in a larger magnitude than the completeness magnitude determined by the objective MBASS method.

[4] The GR law denotes $p(M) \propto 10^{-bM}$.

[5] As M_c reflects the ability of seismic networks to detect earthquakes, there are methods to determine M_c from the observation networks [35, 37].

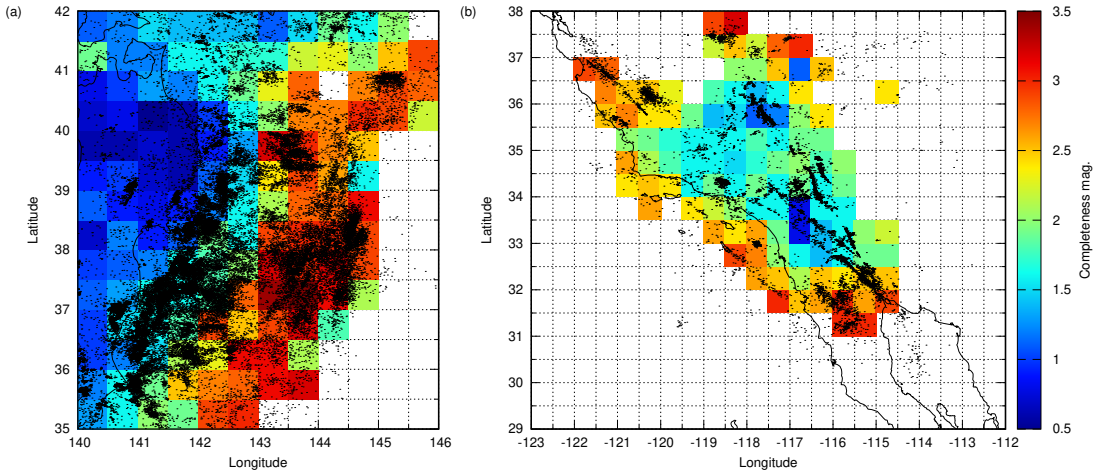

Figure 1.5: Examples of spatial distribution of completeness magnitude M_c calculated using the MBASS method [32]. Referring to Ref. [40], the spatial domain was divided into regions of $0.5° \times 0.5°$, and the results of M_c calculated by applying the MBASS method to cells containing at least 100 earthquake data are shown. (a) Earthquakes recorded in the JMA catalog within 140° – 146°E, 35° – 42°N, and during 01/01/2000 – 03/31/2022, and (b) in the Southern California catalog [18, 19] within 112° – 123°W, 29° – 38°N, and during 01/01/1981 – 03/31/2022. Black dots are epicenters of earthquakes $\geq M3$ for reference. The maps were drawn using the file in [http://www.gnuplotting.org/plotting-the-world-revisited/].

The completeness magnitude can vary in time [36]; M_c will go down as the detection capability increases with the development of seismic observation networks [36], and it is known that the detection ability decreases immediately after the mainshock due to the effect of the mainshock seismic waves [13] and thus M_c temporally increases [36, 38, 39]. M_c also varies spatially [36]; Fig. 1.5 shows the spatial distribution of M_c calculated using the MBASS method for the JMA catalog and the Southern California catalog. The calculations are made for cells containing 100 earthquakes or more, referring to Ref. [40]. As Fig. 1.5 shows, M_c is higher at sea where there are no ground stations and at the edges of the observation network; the detectability decreases in the regions where observation stations are scarce, and thus M_c increases [36]. Therefore, when conducting catalog analysis, it is necessary to avoid such regions with high M_c or set high M_c [36].

In the determined magnitude range in which the GR law holds, the maximum likelihood method is optimal to calculate the b-value [12]; when the lower bound of magnitude range is given, the method in Ref. [33] should be used, and in the case when the lower and upper bounds are given, the method in Ref. [34] can be applicable. In Fig. 1.4(b), the GR law with the b-value calculated by the maximum likelihood method is also shown.

As the magnitude is the logarithm of the energy released by an earthquake [14, 41], the GR law indicates the power-law decay of the energy distribution [pp. 147-148 in Ref. [5]]. Therefore, the above-mentioned two laws represent widely observed scale-free nature characteristic of seismic activity, and the physical mechanisms to explain them have been considered; simple physical and automaton models exhibiting self-organized criticality [42, 43] have been attempted to explain them; for example, the Spring-Block model [44, 45] since Burridge and Knopoff [46] and the cellular automaton model such as the Olami – Feder – Christensen model [47, 48, 49].

Meanwhile, other attempts have been made to model seismic activity as shown in Fig. 1.2 using stochastic models [7, 50, 51], and a simple non-stationary Poisson process combining the empirical OU and GR law called the Epidemic Type Aftershock Sequence (ETAS) model was proposed [27, 50, 52, 53, 54, 55]. Figures 1.1(c – e) suggested that a mainshock can excite aftershocks; the ETAS model assumes that every earthquake can excite its subsequent aftershocks at the rate of the OU law with the weight given by the Utsu – Seki law [56] (or productivity law [57], the number of aftershocks $\propto 10^{\alpha M}$) [50]. Thus, the ETAS model considers seismic activity as a superposition of mainshock-aftershock sequences and expresses it

as an inhomogeneous marked Poisson process with the occurrence rate ($\lambda(t)$) at time t given by [50, 58]:

$$\lambda(t) = \lambda_0 + \sum_{j:t_j<t} \frac{K10^{\alpha(M_j-M_0)}}{(t-t_j+c)^{\theta+1}}, \tag{1.3}$$

where $\{t_j, M_j\}$ ($j \in \mathbb{N}$) are the time and magnitude of events before time t. M_0 represents the minimum magnitude and $(\lambda_0, K, \alpha, c, \theta(=p-1))$ are the parameters to characterize the activity; in particular, λ_0 represents the constant rate for background seismicity [50, 58]. The combination of the remaining parameters as well as the b-value in the GR law yields the branching ratio $n_{\rm br} = \frac{K}{\theta c^\theta}\frac{b}{b-\alpha}$ (when $\theta > 0$) [58] that determines the stationarity of the time series (if $n_{\rm br} < 1$, the time series does not diverge [p. 81 in Ref. [59]]) as well as the ratio of aftershocks generated by mainshocks [60]. While the occurrence rate depends on the past activity as Eq. (1.3), the magnitude of each event is determined independently and randomly by the GR law [50]. This means that the correlation between seismic activity and its subsequent major earthquake occurrence is not assumed.

Although the ETAS model is a simple stochastic process, it is capable of representing the nature of actual seismic activity well [50, 58, 61, 62, 63]. Estimating of the model parameter values from past seismic activity, the ETAS model provides a standard for seismic activity at a region, which enables us to detect relative anomalous activity [27, 50]. The ETAS model has been extended to space-time versions by adding a term on spatial correlations in Eq. (1.3) [52, 54, 64]. Furthermore, the HIST-ETAS model [53, 55], which divides the space into smaller parts, yields the spatial distribution of background seismicity rate (λ_0) [7]. As will be described in Section 1.2.1, the conditional intensity function can evaluate the seismic risk at a certain time, location, and magnitude [7]. Thus, the ETAS model is valuable for seismic risk assessment, but the ETAS model is also useful in examining statistical properties of seismicity due to its high reproducibility of seismic activity, as will be described in Section 1.3.1.

1.2 Probabilistic Forecasting of Earthquake Occurrence

As large earthquakes can cause severe damage to society, seismic risk assessment is an important issue. Probabilistic forecasting of significant earthquake occurrence is one of the approaches to this problem. In this approach, the view of seismic activity is fundamental; the earthquake occurrence is forecasted probabilistically by utilizing the data in seismicity represented by the marked point process as Fig. 1.2. In the following subsections, two primary approaches are described.

1.2.1 Point Process Approach

The conditional intensity function [p.9 in Ref. [7]] [51, 55] $R(t, \boldsymbol{x}, M|H_t)$ given by modeling a marked point process yields the probability of occurrence time (t), spatial position ($\boldsymbol{x}$), and magnitude (M) of a future event using a set of parameter values determined by the history of seismic activity (H_t) which includes the information on small-scale earthquakes $\geq M_0$ [2, 7, 27, 51, 50, 52, 54, 55, 59, 65]. For the ETAS model, the conditional intensity function is given as [7] [6]:

$$\begin{aligned} R(t, M|H_t) &= \lambda(t|H_t)p(M) \\ &\propto \left\{ \lambda_0 + \sum_{j:t_j<t} \frac{K10^{\alpha(M_j-M_0)}}{(t-t_j+c)^{\theta+1}} \right\} 10^{-b(M-M_0)}. \end{aligned} \tag{1.4}$$

Numerous studies have been conducted using such an approach, and projects are underway to develop a better forecasting model [7]. However, as typically indicated in Eq. (1.4), the conditional intensity function is in most point process modeling expressed as the product of two functions; a function that depends only on time and spatial position and another function only on magnitude [7, 51, 65] [Section 6.3.1 in Ref. [59]]. This is based on the already-mentioned assumption that a seismicity pattern does not affect the spatiotemporal occurrence of its subsequent large-size earthquake. This means that although the seismic risk at a spatiotemporal domain can be evaluated depending on past seismic activity, the

[6] Although this is the conditional intensity function for time and magnitude, if Eq. (1.3) includes a function for space, one can obtain the conditional intensity function for space-time-magnitude [pp.9-10 in Ref. [7]].

GR law independently determines how large the event will occur. Considering correlations between a significant shock and seismicity ahead of it [27, 40, 65, 66, 67, 68] is important for improving forecasting [65]. Thus, it is meaningful to prepare a theoretical framework to manage such inter-event correlations toward improving forecasting.

1.2.2 Renewal Process Approach

Although the above-mentioned point process approaches have been actively studied for better earthquake forecasting [7], the currently used seismic risk assessment is based on a simpler viewpoint of the renewal process [69] (see Fig. 1.2(c)). In this approach, a point process at a certain large cut-off magnitude [7] is regarded as a renewal process, and the risk of occurrence of earthquakes $\geq$ that cut-off magnitude is evaluated by a hazard function of the inter-event time distribution determined by past seismic activity [51, 59, 70].

The inter-event time distribution is defined as a probability density of a length between adjacent points in the point process determined by setting a cut-off magnitude for the marked point process (see Fig. 1.2(c)). Hereafter, for the cut-off magnitude m (M), the inter-event times are denoted by using a variable τ_M (τ_m), and the inter-event time distribution it follows by $p_M(\tau_M)$ ($p_m(\tau_m)$). Also, the average of the inter-event time distribution is expressed as:

$$\langle \tau_m \rangle := \int_0^\infty \tau_m p_m(\tau_m) d\tau_m, \tag{1.5}$$

$$\langle \tau_M \rangle := \int_0^\infty \tau_M p_M(\tau_M) d\tau_M. \tag{1.6}$$

The hazard function ($\phi_M(t)$) for the event with magnitude $\geq M$ is defined as follows [pp.50-51 in Ref. [59]]:

$$\phi_M(t) = \frac{p_M(t)}{\int_t^\infty p_M(\tau) d\tau}. \tag{1.7}$$

The Headquarters for Earthquake Research Promotion basically publishes the probabilistic risk of earthquake occurrence based on the hazard function calculated above [69].

However, this approach has a disadvantage that it cannot utilize the abundant information of earthquakes below the cut-off magnitude for forecasting, which can be utilized in the point process approach in Section 1.2.1. As the GR law suggests, the number of earthquakes exponentially increases as the magnitude decreases. Because the information included in seismic catalogs is limited, utilizing such abundant information for forecasting is an important point. Therefore, considering the theoretical framework to incorporate the information on small-scale earthquakes into this renewal process approach is significant to improve the earthquake forecasting method currently used.

1.3 Properties of Inter-event Time Distribution in Seismic Activity

As described in Section 1.2.2, the inter-event time distribution yields fundamentally important information for probabilistic forecasting of earthquakes. On the other hand, the inter-event time distribution in seismic activity has been studied in the context of a better understanding of complex seismic activity, such as its connection with other statistical laws and properties [25, 71, 72, 73, 74, 75, 26]. In this section, two properties of the inter-event time distributions of earthquakes are introduced.

[7] This thesis distinguishes the threshold magnitude (M) set for a marked point process as follows. If the events with a magnitude lower than threshold magnitude M are discarded, then such threshold is called the cut-off magnitude. The cut-off magnitude is used when considering seismic time series in which the magnitude values are recorded in one or two decimal places (In Chapter 5, the calculated moment magnitude in the CMT catalog is output to one decimal place, and the cut-off magnitude is set as in other catalog data). On the other hand, if the events with a magnitude smaller than or equal to the threshold magnitude M are discarded, that is called the magnitude threshold, which is used in theoretical analysis and numerical analysis using synthetic time series such as the ETAS model in which the magnitude value is a real number. Although such a distinction does not make much difference, we make them just in case.

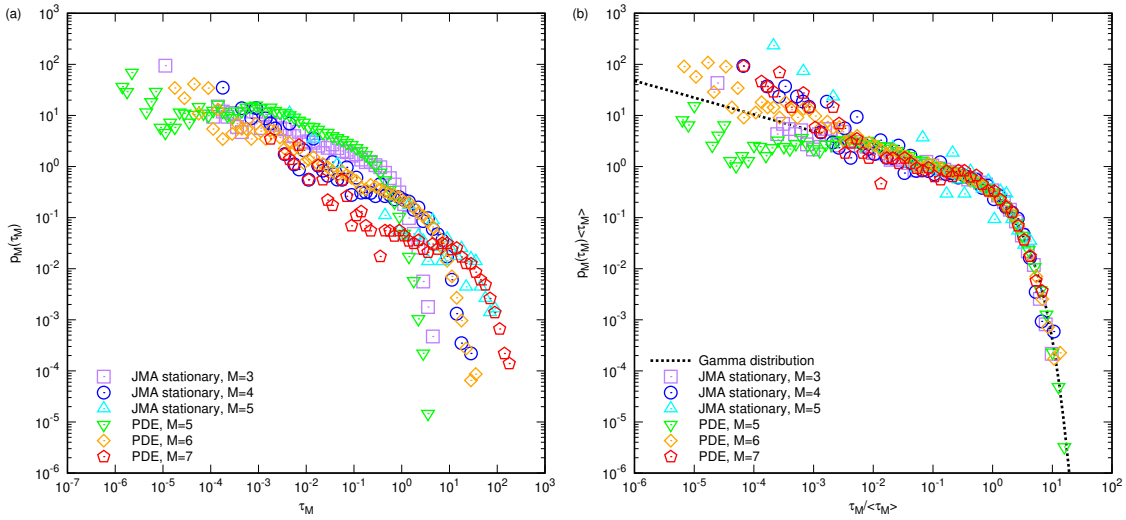

Figure 1.6: Examples of scaling collapses of the inter-event time distributions for stationary time series in the JMA and PDE catalogs. (a) The inter-event time distributions ($p_M(\tau_M)$) for different cut-off magnitudes. (b) Rescaled inter-event time distribution by its average ($\langle\tau_M\rangle$). The dotted black curve shows the gamma distribution in Eq. (1.9).

1.3.1 Scaling Universality of Inter-event Time Distribution in Seismicity

Bak et al. pointed out a universality in the inter-event time distribution in seismicity after scaling by the GR and OU laws and the fractality of epicenters [71]. Although its universality was questioned [76], Corral found another scaling universality in the inter-event time distribution [77] in the process of examining the Bak et al.'s universality [72, 73]. Corral's scaling universality indicates that the inter-event time distribution at the cut-off magnitude M in a spatial domain S (temporarily represented as $p_M^S(\tau_M)$ with superscript S) collapses to around a universal (generalized) gamma distribution ($f(y)$) independent of M and S, except for short time intervals ($y \lesssim 10^{-2}$) [77, 78], after rescaled by the occurrence rate (R_M^S):

$$p_M^S(\tau_M) = R_M^S f\left(R_M^S \tau_M\right), \tag{1.8}$$

$$f(y) \approx C y^{\gamma-1} e^{-y/B}, \tag{1.9}$$

where $(\gamma, B, C) \approx (0.67, 1.58, 0.5)$ [77]. This scaling universality was shown for stationary seismic time series (almost constant occurrence rate, $R_M^S \approx R$ (const.), which is the average occurrence rate over the stationary time series [77])[72, 73, 77, 79, 80] and aftershock sequences (for time t from the mainshock so large $t \gg c$ that the OU occurrence rate in Eq. (1.1) can be approximated by power-law $R_M^S(t) \propto t^{-p}$) [77, 80]. As the occurrence rate is almost constant in the stationary time series, the scaling factors in these time ranges can be consolidated into the occurrence rate at a given moment; thus, this scaling universality implies that the instantaneous occurrence rate governs the temporal characteristic of seismic activity [77].

Figure 1.6 illustrates examples of scaling in Eq. (1.8) for the stationary time domains in the JMA catalog and the PDE catalog (blue-colored time domain in Fig. 1.1(d) and whole time series in Fig. 1.3(b), see also Fig. 5.7 in Chapter 5). As these time series can be regarded as stationary, the scaling factor of the inter-event time distribution is the inverse of the average interval ($R_M = \langle\tau_M\rangle^{-1}$ for the cut-off magnitude M). The disparate inter-event time distributions for each cut-off magnitude and the spatial domain collapse around the gamma distribution in Eq. (1.9).

Subsequent studies of Corral's catalog analysis critically examined this scaling universality theoretically and numerically using the ETAS model [78, 81, 82, 83, 84, 85] and elevated it to include multiple time scales other than the occurrence rate [86, 87, 88].

The author notes that the treatment of the time series differs in these subsequent studies from Corral's original method; these studies regarded the stationarity as the branching ratio [58] [p. 81 in Ref. [59]]] is less than one, and the universality was discussed for the mixture time series that can include major

aftershock activities using the average occurrence rate for entire time series as the scaling factor, leading to approximate- or non-universality. However, according to the original treatment of seismic time series, scaling of the inter-event time distribution should be considered for stationary time series and aftershock sequences by different scaling factors to discuss universality. Rescaling by the average of the entire mixture time series overestimates the number of short inter-event times in rescaling inter-event time distributions because the short intervals in major aftershock sequences are not transformed by the OU law. Also, the scaling universality for aftershock sequences by the OU rate has rarely been considered (excluding, for example, Ref. [89]) since Corral pointed it out.

1.3.2 Dependence of Weibull Parameters on Cut-off Magnitude

It was shown by the catalog analysis that the inter-event time distributions in seismic activity can be fit by the superposition of the Weibull distribution and the log-Weibul distribution [90, 91]. Their proportion varies with the cut-off magnitude in a way that as it gets larger, the proportion of the Weibull distribution becomes dominant [90, 91]. Thus, the main contribution for the inter-event time distribution is the Weibull distribution, and the logWeibull distribution works as compensation [25, 26]. The shape of the Weibull probability density function ($p_M^W(\tau_M)$, or its residual distribution function $P_M^W(\tau_M)$ and cumulative distribution $Q_M^W(\tau_M) := 1 - P_M^W(\tau_M)$) varies with the cut-off magnitude M in the following way [25, 26, 90, 91]:

$$p_M^W(\tau_M) = \left(\frac{\alpha_M}{\beta_M}\right)\left(\frac{\tau_M}{\beta_M}\right)^{\alpha_M - 1} \exp\left[-\left(\frac{\tau_M}{\beta_M}\right)^{\alpha_M}\right], \tag{1.10}$$

$$P_M^W(\tau_M) = \int_{\tau_M}^{\infty} p_M^W(s)ds = e^{-\left(\frac{\tau_M}{\beta_M}\right)^{\alpha_M}}, \tag{1.11}$$

$$Q_M^W(\tau_M) = \int_0^{\tau_M} p_M^W(s)ds = 1 - e^{-\left(\frac{\tau_M}{\beta_M}\right)^{\alpha_M}}. \tag{1.12}$$

The Weibull parameters $\{\alpha_M, \beta_M\}$ characterize the temporal property of the renewal process at the cut-off magnitude M, and their dependence on M shows the characteristic of the temporal hierarchy in seismic activity reflecting the correlations between events at different scales [25, 26, 75]. Such a hierarchical structure was examined by catalog analysis for mixture time series as well as the aftershock sequence [25, 26]. For the aftershock sequence, the moving ensemble of 100 days long was used to see the temporal variation of the Weibull distribution; in each moving ensemble, the inter-event time distribution was fit by the Weibull distribution and the hierarchical structure captured by the dependence of Weibull parameters on cut-off magnitude was examined [25, 26]. The unification of the inter-event time statistics and the magnitude frequency can be represented as, using a constant $k_{\rm EQ}$ independent of cut-off magnitude M [25, 26, 75]:

$$\langle\tau_m\rangle P(m) = \langle\tau_M\rangle P(M) = e^{-k_{EQ}}, \tag{1.13}$$

and using Weibull statistics and the GR law [25, 26, 75]:

$$\beta_M \Gamma\left(1 + 1/\alpha_M\right) 10^{-bM} = e^{-k_{EQ}}, \tag{1.14}$$

where the representation of the average of the Weibull distribution $\langle\tau_M\rangle = \beta_M\Gamma\left(1 + 1/\alpha_M\right)$ ($\Gamma(\cdot)$ is the gamma function) is used. Equation (1.14) indicates the unification of the Weibull statistics and the GR law [25, 26, 75]. Equation (1.14) indicates that the hierarchical structure captured by the Weibull distribution can be expressed as [25, 26, 75]:

$$\begin{aligned} \alpha_M &= f_\alpha(M, b), \\ \beta_M &= f_\beta(M, b), \end{aligned} \tag{1.15}$$

and from Eq. (1.15), the relation between Weibull parameters is:

$$\beta_M = f_\beta(f_\alpha^{-1}(\alpha_M)) = g(\alpha_M, b). \tag{1.16}$$

The relation in Eq. (1.16) (or Eq. (1.15)) is called the "multi-fractal relation" and its universal features have been reported [25, 26]. The multi-fractal relation is a new formula to combine two important seismic

laws: the GR law for the magnitude and the Weibull law for the inter-event times [25, 26, 75], and Eq. (1.16) suggests that the GR parameter b is determined by the Weibull parameters (α_M, β_M) [25, 75].

However, as with the GR and OU laws, the multi-fractal relation, as well as the Weibull distribution, is an empirical property, and it is necessary to promote a more fundamental understanding of multi-fractal relation.

1.4 Purpose and Structure of Thesis

As described in Section 1.2, there is room for improvement in the approaches of probabilistic forecasting of future earthquake occurrence; Section 1.2.1 pointed out the problem in point process approach that the correlation between seismic activity and subsequent large earthquake is not assumed, and Section 1.2.2 pointed out that the practical renewal process approach cannot take into consideration the abundant information on small-scale earthquakes. The latter approach is based on the inter-event time distribution in seismic activity, which, as mentioned in Section 1.3, is important not only as a basis of earthquake forecasting but understanding statistical properties of seismic activity. Two properties related to the universality in the hierarchy of the inter-event time statistics were presented with their respective problems.

To better understand seismic activity and improve probabilistic forecasting, (1) this thesis presents a method to analyze the hierarchical nature of seismic activity represented by marked point processes, i.e., the conditional probability between inter-event times at different cut-off magnitude and the integral equation that connects the inter-event time distributions at different cut-off magnitude via the conditional probability, and (2) examines the hierarchical structure using several seismic catalogs to understand seismic activity, in particular the inter-event time statistics. Furthermore, (3) this thesis presents a Bayesian approach to improve the renewal process approach for forecasting earthquakes based on that conditional probability, which also possibly enhances the point process approach, and (4) examines it using synthetic and seismic time series.

Chapter 2 is related to the points (1) and (2). This chapter introduces the conditional probability and the integral equation. The properties related to the conditional probability are analyzed using several seismic catalogs, and an attempt is made to derive the multi-fractal relation from the integral equation.

Chapter 3 is related to the points (3) and (4). Bayes' theorem about the conditional probability and its extension, Bayesian updating, are considered analytically for a simple point process. This Bayesian method incorporates the information on small-scale events into the forecasting by the renewal process approach. The Bayesian updating is numerically applied to the ETAS time series to examine its effectiveness for forecasting future large-scale event timing. Possible improvement by the Bayesian framework of the point process approach is also discussed.

Chapter 4 is related to the point (3). This chapter extends the Bayesian method considered in Chapter 3 for magnitude-time marked point process to the spatiotemporal version theoretically.

Chapter 5 is related to the points (2), (3), and (4). The simplest case of the Bayesian method (Bayes' theorem) is examined using stationary and aftershock sequences, though these are transformed by occurrence rates. The probability densities related to Bayes' theorem are analyzed in such seismic time series with weak inter-event correlations, and the scaling universality in the inter-event time distribution is discussed in view of the temporal hierarchy in seismic time series. Possible ways to apply the Bayesian method for forecasting are discussed based on the results of catalog analysis.

Finally, Chapter 6 summarizes the chapters and discusses the future prospects.

Chapter 2

Introduction of the Method to Characterize the Hierarchical Structure of Seismic Time Series and the Analysis of the Hierarchy of Weibull Statistics of Inter-event Time Statistics

This chapter is based on and updated from Ref. [1], which is paper 1 in the list of author's papers, and Ref. [92].

2.1 Introduction

This chapter introduces the central tools throughout the thesis: the conditional probability that characterizes the temporal hierarchy in marked point processes, and the integral equation that has this conditional probability at its kernel and the inter-event time distributions at its solution. Properties of the conditional probability are examined using two seismic catalogs, and the multi-fractal relation is derived from the integral equation with such properties.

2.2 Conditional Probability and Integral Equation

This chapter considers the time series of shocks as represented in Fig. 1.2(c). Let us define the conditional probability $p_{mM}(\tau_m|\tau_M)$ which expresses the probability density of the inter-event time intervals τ_m in the upper interval τ_M, where the magnitude m is always smaller than M, i.e., $m < M(:= m+\Delta m,\ \Delta m > 0)$. The conditional probability defined here plays an important role in characterizing the correlation between shock sequences with two different cut-off magnitudes, m and M. If we assume that there exist the stationary inter-event time distributions $p_m(\tau_m)$ and $p_M(\tau_M)$ in each magnitude level, $p_{mM}(\tau_m|\tau_M)$ satisfies the following integral equation:

$$N_m p_m(\tau_m) = N_M \int_{\tau_m}^{\infty} \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M) p_M(\tau_M) d\tau_M. \tag{2.1}$$

Here N_m and N_M stand for the total number of inter-event time intervals at the threshold level m and M, respectively. Also, $\langle\langle \tau_m \rangle\rangle_{\tau_M}$ represents the conditional average of the inter-event time intervals at the lower cut-off magnitude in an upper interval of length τ_M:

$$\langle\langle \tau_m \rangle\rangle_{\tau_M} = \int_0^{\infty} \tau_m \; p_{mM}(\tau_m|\tau_M) d\tau_m. \tag{2.2}$$

Equation (2.1) can be expressed as follows using the GR law ($N_M/N_m = 10^{-b\Delta m}$ (This transformation 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. [8]):

$$p_m(\tau_m) = 10^{-b\Delta m} \int_{\tau_m}^{\infty} \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M) p_M(\tau_M) d\tau_M. \tag{2.3}$$

In the present chapter, we refer to Eq. (2.3) as the embedding equation for the magnitude sequences of earthquakes since the equation characterizes the hierarchical structure of the point process at the lower cut-off magnitude (τ_m's), which is embedded in the large scale interval τ_M. In fact, Eq. (2.3) combines the two inter-event time distributions, $p_m(\tau_m)$ and $p_M(\tau_M)$, via the conditional probability $p_{mM}(\tau_m|\tau_M)$, and one can easily understand that Eq. (2.3) describes the higher correlation between two point processes obtained by different cut-off magnitudes m and M. Thus, the embedding equation (2.3) allows us to discuss the characteristics of the inter-event time distribution from the viewpoint of the hierarchical nature of the inter-event times quantified by the conditional probability [9].

The conditional probability $p_{mM}(\tau_m|\tau_M)$ can be decomposed into two parts depending on the condition (1) $\tau_m < \tau_M$ and (2) $\tau_m = \tau_M$. These cases correspond to the two patterns to obtain the lower inter-event time interval of length τ_m by decreasing the threshold level from M to m, as shown in Fig. 2.1. Therefore, the conditional probability $p_{mM}(\tau_m|\tau_M)$ is represented as:

$$p_{mM}(\tau_m|\tau_M) = \Xi_{mM}(\tau_m, \tau_M)\theta(\tau_M - \tau_m) + \Delta_{mM}(\tau_M)\delta(\tau_M - \tau_m), \tag{2.4}$$

where $\theta(x)$ is a unit step function ($\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ for $x \leq 0$) and $\delta(x)$ is the Dirac's delta function. The first term $[\Xi_{mM}(\tau_m, \tau_M)\theta(\tau_M - \tau_m)]$ corresponds to the case (1) in Fig. 2.1 and $[\Delta_{mM}(\tau_M)\delta(\tau_M - \tau_m)]$ to the case (2) in Fig. 2.1. By considering the normalization condition of $p_{mM}(\tau_m|\tau_M)$, $\Delta_{mM}(\tau_M)$ is expressed by $\Xi_{mM}(\tau_m, \tau_M)$ as:

$$\Delta_{mM}(\tau_M) = 1 - \int_0^{\tau_M} \Xi_{mM}(\tau_m, \tau_M) d\tau_m. \tag{2.5}$$

Therefore, $\Xi_{mM}(\tau_m, \tau_M)$ is the only factor to determine the form of $p_{mM}(\tau_m|\tau_M)$. Finally, the embedding equation (2.1) is rewritten into the following form:

$$\begin{aligned} N_m p_m(\tau_m) = N_M \int_{\tau_m}^{\infty} & \frac{\tau_M \ \Xi_{mM}(\tau_m, \tau_M) p_M(\tau_M)}{\int_0^{\tau_M} x \ \Xi_{mM}(x, \tau_M) dx + \tau_M \Delta_{mM}(\tau_M)} d\tau_M \\ & + N_M \frac{\tau_m p_M(\tau_m) \Delta_{mM}(\tau_m)}{\int_0^{\tau_m} x \ \Xi_{mM}(x, \tau_m) dx + \tau_m \Delta_{mM}(\tau_m)}. \end{aligned} \tag{2.6}$$

This equation shows that the interrelation between two inter-event time distributions $p_m(\tau_m)$ and $p_M(\tau_M)$ are completely determined by a characteristic function $\Xi_{mM}(\tau_m, \tau_M)$. In other words, the correlation between two shocks with different threshold levels m and M is expressed by $\Xi_{mM}(\tau_m, \tau_M)$.

In the following sections, the property of the conditional probability is examined using seismic catalogs to consider the multi-fractal relation from the viewpoint of the hierarchical structure of seismic time series. In the next section, we introduce the information on seismic catalogs to be used in this chapter.

2.3 Seismic Catalogs and Statistical Properties

2.3.1 Catalog Information

The three seismic time series (named PDE, JS, and JA, respectively) are numerically analyzed in the present chapter [16, 17, 20] (see Table 2.1) [10]. The first one (PDE) is for the worldwide global case, and the second (JS) and the third (JA) are the data in the local Fukushima – Miyagi area in Japan; (JS) is the mixture time series before the large earthquake (Tohoku EQ at March 11, 2011), and (JA) for

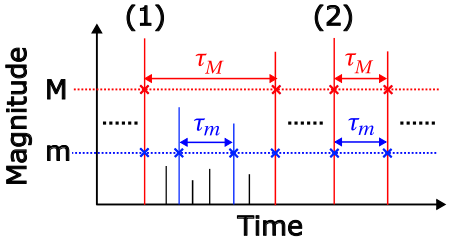

Figure 2.1: Schematic of the two patterns that yields an inter-event time interval at the lower cut-off magnitude (m) of length τ_m. (1) The lower interval of length τ_m is obtained as a piece of its upper interval (of length τ_M) divided by the event inserted by lowering the cut-off magnitude. In this case, $\tau_m < \tau_M$. (2) No event is inserted in the upper interval when lowering the cut-off magnitude. In this case, $\tau_m = \tau_M$.

Table 2.1: Catalog information on the time series used in Chapter 2.

Catalog	Name	Spatial domain	Time domain	b-value (in [1] [11])	b-value (MLE)
PDE	PDE	Whole world	01/01/1973 – 12/31/2002	1.05	1.042 ($M \in [5, 7.5]$)
JMA	JS	$140°E$ – $150°E$	10/01/1997 – 02/28/2011	0.84	0.768 ($M \in [3, 5]$)
	JA	$35°N$ – $42°N$	03/11/2011 – 03/11/2014	0.95	0.878 ($M \in [4, 5]$), 0.725 ($M \in [3, 5]$)

the aftershock sequence of the Tohoku EQ. As will be described later, the aftershock sequence (JA) is analyzed using moving ensembles with 100-day time intervals.

2.3.2 Statistical Properties of the Seismic Time Series

Magnitude frequency

Figure 2.2 shows the magnitude frequency for the three time series. One can see deviations from the GR law in Fig. 2.2(c) because the ensemble is not stationary in the JA case. The GR law can be judged to hold in the magnitude range $M \in [5, 7.5]$ for PDE, $M \in [3, 5]$ for JS, and $M \in [4, 5]$ for JA. The b-values calculated within these magnitude ranges by the maximum likelihood estimate [33] are summarized in Table 2.1 and indicated in Fig. 2.2. Although the case of JA seems complete for $M \geq 4$, we also used $M \geq 3$ data where the GR law is approximately satisfied in the analysis of the hierarchical structure of seismic activity (the b-value calculated in this range is also in Table 2.1). Much detailed analysis of non-stationary data was reported in the preceding papers [25]. In this Chapter, the cut-off magnitude is set to be within the same magnitude range as that used in the calculation of the b-values by the maximum likelihood estimate (i.e., $M \in [5, 7.5]$, $M \in [3, 5]$, and $M \in [4, 5]$ for PDE, JS, and JA, respectively). The cut-off magnitude is set in the increment of 0.1 from the respective minimum magnitude.

Occurrence rate of aftershocks

Figure 2.3 shows the time evolution of the occurrence rate of aftershocks since the Tohoku mainshock time (t days) for JA. The OU power-law decay ($\propto t^{-p}$) is observed for M (cut-off magnitude)≥ 3 and t longer than $1 \sim 10$ days. For this aftershock sequence, we denote the time series in 100-day time window ($(t_0, t_0 + 100]$) with the initial time $t_0 = 100,\ 200,\ ...,\ 500$ (days from the mainshock time) by

[8] Note here is from the author's paper [2], paper 2 in the list of author's papers.
[9] This sentence is from the author's paper [4], paper 4 in the list of author's papers.
[10] The data in the seismic catalogs are at the time of Ref. [1].
[11] The b-values in Ref. [1] were obtained as follows. For PDE, b-value was calculated by fitting the probability density $(\log_{10} p(j\Delta M)$ $(j \in \mathbb{Z})$, the endpoint of the bin width of magnitude $(\Delta M = 0.1)$ as a representative point (whereas in Fig. 2.2, the representative point is at $(j + 0.5)\Delta M$ for the probability density)) in the range $M \in [5, 7]$. For JS and JA, b-value was calculated by fitting $\log_{10} P(j\Delta M)$ in the range $3 \leq M$ and $M \in [4, 8]$, respectively.

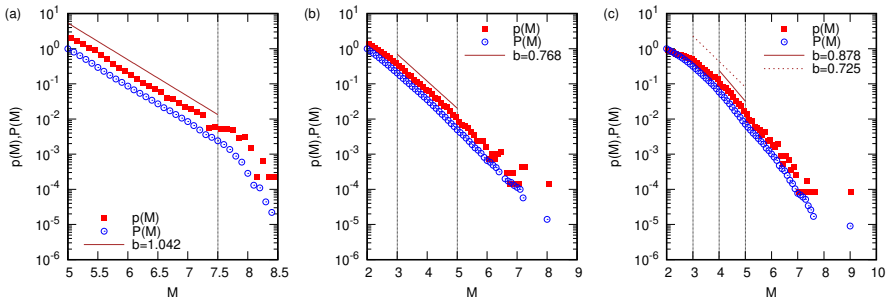

Figure 2.2: The magnitude frequency ($p(M)$) and the distribution function of magnitudes ($P(M)$) for the time series of (a) PDE, (b) JS, and (c) JA.

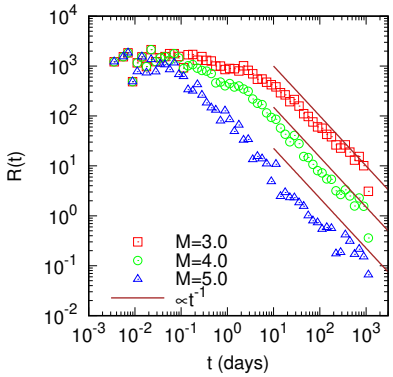

Figure 2.3: Magnitude dependence of the time evolution of the occurrence rate of aftershocks of the Tohoku mainshock to the time since the mainshock. The solid lines show the OU law with $p = 1$ for reference. See Ref. [25] for details.

JA($t_0 = 100$), JA($t_0 = 200$), ..., JA($t_0 = 500$), respectively, and examine their property in each time window.

Inter-event time distributions

Figure 2.4 shows the residual probabilities ($P_M(\tau_M) := \int_{\tau_M}^{\infty} p_M(x)dx$) of the inter-event time distribution at several cut-off magnitudes for our data-analysis and their fittings by the Weibull distribution (Eq. (1.12)) for our data-analysis [12]. Figure 2.4 shows that the Weibull distribution describes the inter-event time distribution, in particular at large intervals. Note that in the calculation of the inter-event time distribution, we treated the seismic catalogs similarly as the former studies [25, 26], i.e., the minimum cut-off time interval τ_{min} is introduced to remove the short intervals from the population to avoid the uncertainty coming from the time resolution in each data catalog; a similar idea is used in the inter-event time statistics of the earthquakes [71, 77, 90]. The parameter τ_{min} used in the catalog analysis is summarized in Table 2.2.

Multi-fractal relation

The Weibull parameters ($\{\alpha_M, \beta_M\}$) at each cut-off magnitude (M) that can be obtained by the fittings in the previous subsection yield the multi-fractal relation for each time series. Figure 2.5 shows the multi-fractal diagrams [25, 75, 26]. The Weibull fittings of the inter-event time distribution are performed up to the upper limit of the magnitude range to analyze determined in the previous subsection, and several Weibull parameter values at large cut-off magnitudes are added to those in Ref. [1]. α_M and β_M show

[12] In the calculation of the residual probabilities of the inter-event time distribution, the bin width ($\Delta\tau$) was set to 0.0025 for PDE and JA, and 0.01 for JS as in Ref. [1]. The same number of points such that $P_M(j\Delta\tau) \equiv 0$ are recorded at the tail of the distribution as the number of bins with $p_M(j\Delta\tau) > 0$.

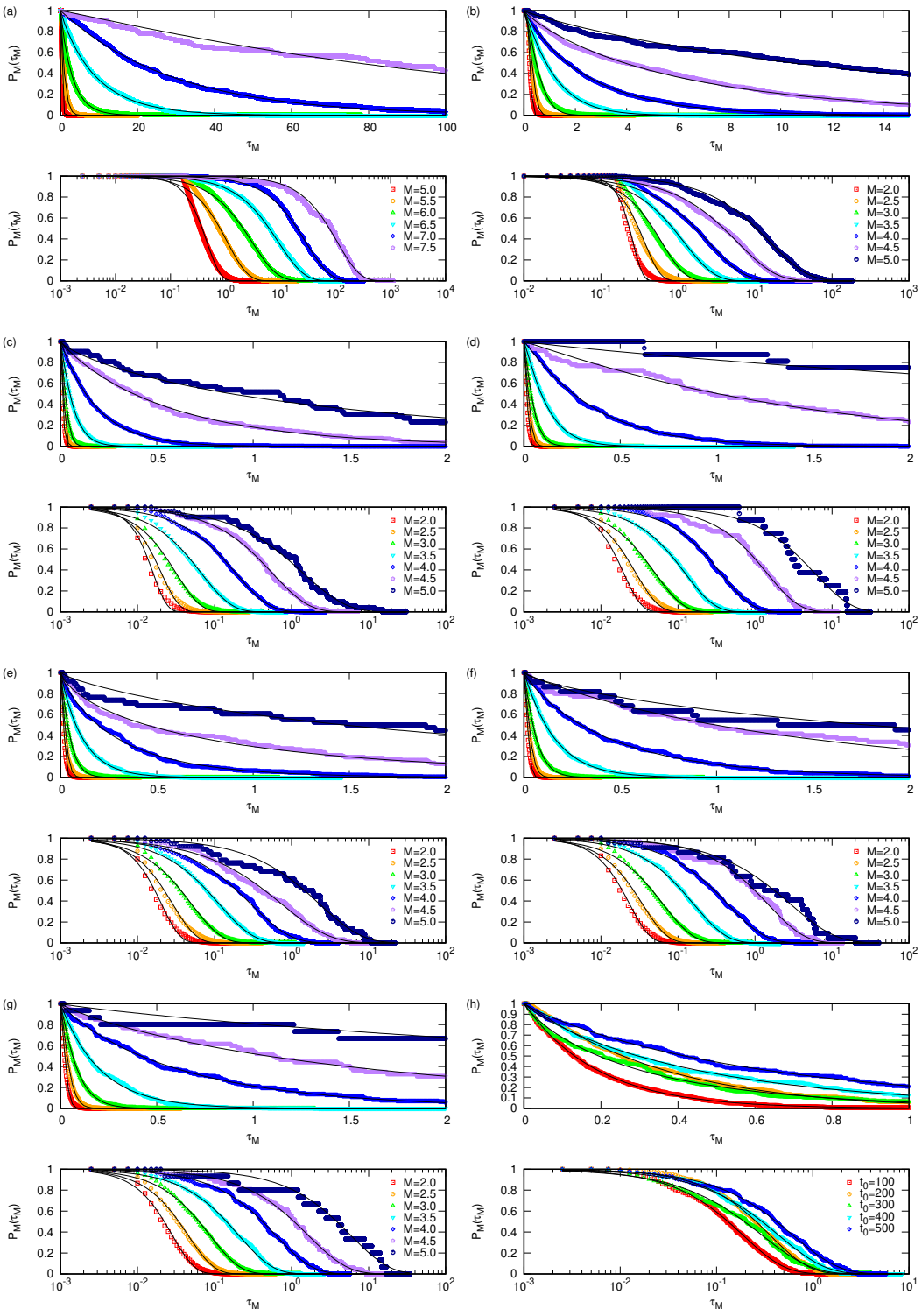

Figure 2.4: The residual distributions ($P_M(\tau_M)$) of the inter-event times and their fittings by the Weibull distribution $P_M^W(\tau_M)$ for (a) PDE, (b) JS, (c) JA($t_0 = 100$), (d) JA($t_0 = 200$), (e) JA($t_0 = 300$), (f) JA($t_0 = 400$), and (g) JA($t_0 = 500$). (h) shows $P_M(\tau_M)$ at $M = 4$ with its fitting Weibull distribution for each time window $t_0 = 100$ to $t_0 = 500$ in JA. The upper panels show the distributions at short inter-event times, and the lower panels show the entire distribution with the logarithmic scale of the inter-event times.

monotonically decreasing and increasing trends with respect to the cut-off magnitude M, respectively, of which empirical forms can be approximated by:

$$\alpha_m \approx \lambda 10^{-\mu m} + \nu, \tag{2.7}$$

$$\log(\beta_m) \approx -p + q10^{rm}, \tag{2.8}$$

where (λ, μ, ν) and (p, q, r) are the fitting parameters in each case. The parameter values (for four significant digits) are summarized in Table 2.3.

Figure 2.6 shows the multi-fractal relation corresponding to each case in Fig. 2.5 [13]. In the stationary (PDE) and mixture (JS) cases, the multi-fractal relation seems to be close, though the multi-fractal diagrams ($f_\alpha(M, b)$ and $f_\beta(M, b)$ in Eq. (1.15)) are different as shown in Figs. 2.5(a) and (b). Furthermore, in the case of the non-stationary (aftershock) regime (JA($t_0 = 100$), ..., JA($t_0 = 500$)), the multi-fractal relation in each moving ensemble seems to be around an invariant curve though the difference in those parameters (α_M, β_M) as shown in Figs. 2.5(c – g). The universality conjecture for the multi-fractal relation have been discussed in preceding papers [25, 75, 26], but the theoretical background still remains open.

2.4 Analysis of the Conditional Probability and the Integral Equation using Seismic Catalog Data

This section introduces two phenomenological coefficients to demonstrate the correlation in the time series of earthquakes; one is the magnitude factor $\eta_{mM}(\tau_m)$ which can describe the influence from the larger magnitude M to the lower magnitude m, and the other is the temporal factor $a_{mM}(\tau_M)$ which describes the influence from the upper interval τ_M (for larger shock) to the lower interval τ_m (for smaller shock). In what follows, one can see that those two factors $\eta_{mM}(\tau_m)$ and $a_{mM}(\tau_M)$ obey certain scaling features respectively, and the embedding equation can be rewritten by using those correlation coefficients $\eta_{mM}(\tau_m)$ and $a_{mM}(\tau_M)$.

2.4.1 Magnitude Correlation Coefficient $\eta_{mM}(\tau_m)$ and its Scaling Property

Here, we derive a characteristic parameter to describe the magnitude correlation $\eta_{mM}(\tau_m)$. The first term on the right-hand side (r.h.s.) of Eq. (2.6) corresponds to the part of θ-function in Eq. (2.4), and the quantity multiplied by $d\tau_m$ represents the total number of the intervals at the lower cut-off magnitude with the length within $[\tau_m, \tau_m + d\tau_m)$ obtained as a piece of division of the upper interval as shown in Fig. 2.1(1). Denote the second term on the r.h.s. in Eq. (2.6) by $N_\delta(\tau_m)$, and its ratio to the total quantity of the left-hand side (l.h.s.) of Eq. (2.6) ($N_m p_m(\tau_m)$) by $R_\delta(\tau_m)$, i.e.:

$$N_\delta(\tau_m) := N_M \frac{\tau_m p_M(\tau_m) \Delta_{mM}(\tau_m)}{\int_0^{\tau_m} x\, \Xi_{mM}(x, \tau_m) dx + \tau_m \Delta_{mM}(\tau_m)}, \tag{2.9}$$

and:

$$R_\delta(\tau_m) := \frac{N_\delta(\tau_m)}{N_m p_m(\tau_m)}. \tag{2.10}$$

Table 2.2: Conditions used in the statistical analysis.

Time series	Inter-event time distribution	ICC		MCC	
	$\tau_{\min}$ (minutes)	$\tau_{\min}$ (minutes)	$N^M_{\min}$	$\tau_{\min}$ (minutes)	$N^m_{\min}$
PDE	240	240	20	2.0	100
JS	240	240	20	2.0	100
JA	14.4	14.4	20	0.0	75

[13] In Ref. [1], several α_M values in JA($t_0 = 100$) were incorrectly set to slightly different values of about four decimal places when moving them from the multi-fractal diagram to multi-fractal relations, which is corrected in Figs. 2.6, 2.11, and 2.14.

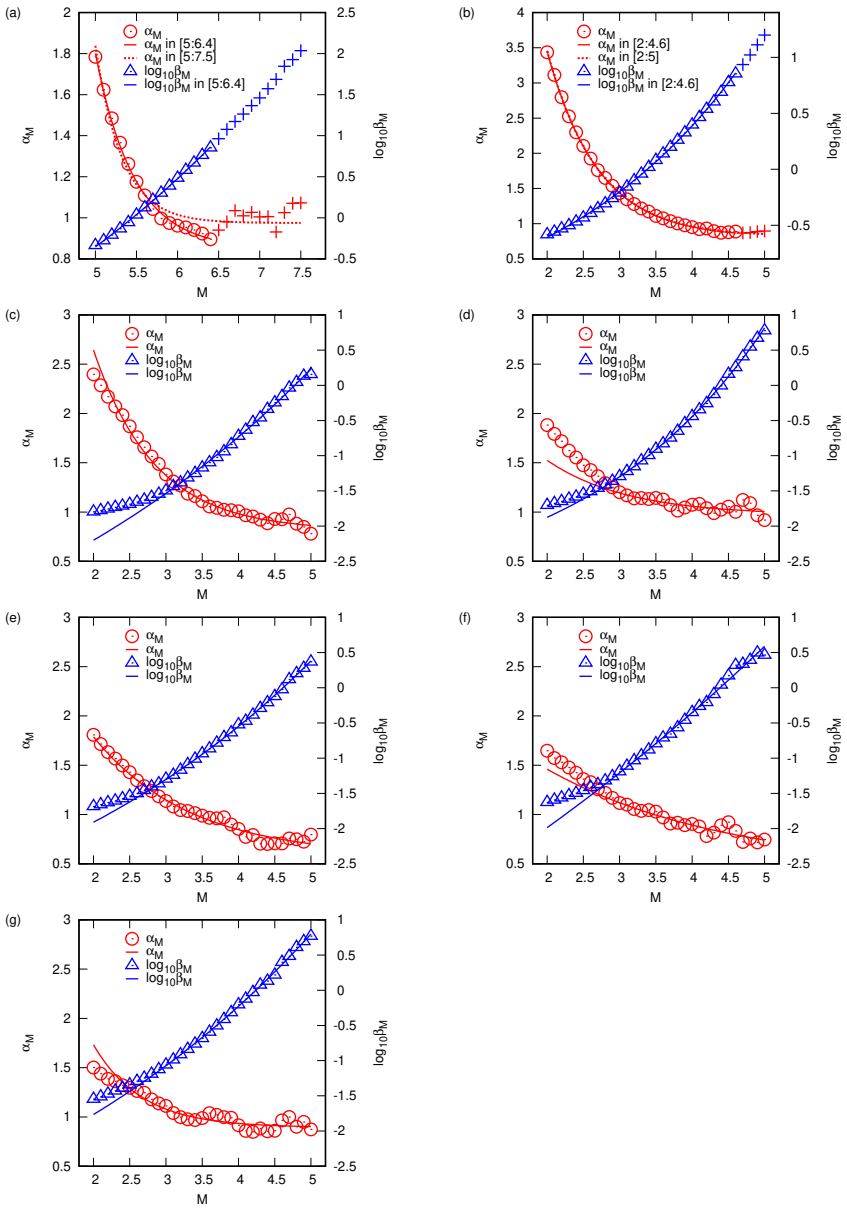

Figure 2.5: Multi-fractal diagrams for each time series; (a) PDE, (b) JS, (c) JA($t_0 = 100$), (d) JA($t_0 = 200$), (e) JA($t_0 = 300$), (f) JA($t_0 = 400$), and (g) JA($t_0 = 500$). The red-colored (dotted) curves show the fitting of α_M by Eq. (2.7) and blue-colored curves the fitting of $\log_{10}\beta_M$ by Eq. (2.8). In (a), the results of Weibull parameters in the magnitude range $M \in [5, 6.4]$ analyzed in Ref. [1] are indicated using $\bigcirc$ and $\triangle$, and the results newly added in the range $M \in [6.5, 7.5]$ are by $+$ and $\times$. The fittings by Eqs. (2.7) and (2.8) in the range $[5, 6.4]$ are shown using solid curves, and in the range $[5, 7.5]$ by the dotted curve. In (b), the results of Weibull parameters in $M \in [2, 4.6]$ analyzed in Ref. [1] are indicated by $\bigcirc$ and $\triangle$ with their fitting functions shown by solid curves. Also, the newly added results in the magnitude range $M \in [4.7, 5]$ are indicated by $+$ and $\times$ symbols with the fitting curve in the range $M \in [2, 5]$ shown by the dotted curve.

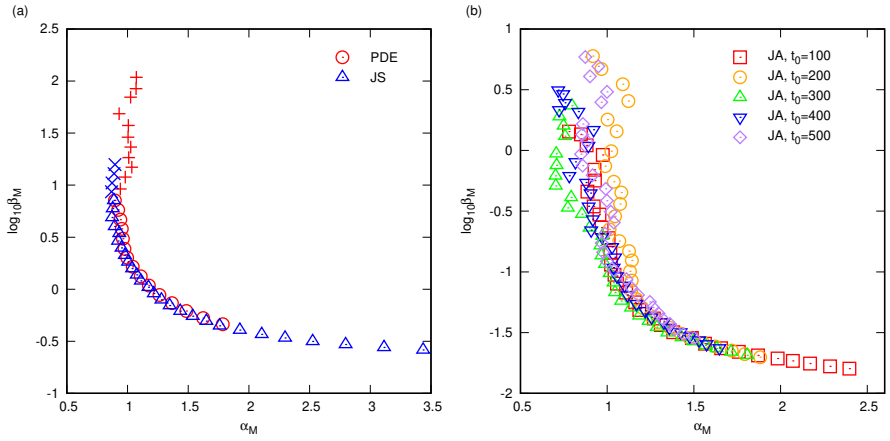

Figure 2.6: Multi-fractal relations obtained from Fig. 2.5; cases of (a) stationary (PDE) and mixture (JS), and (b) non-stationary (JA($t_0 = 100$), ..., JA($t_0 = 500$)). Cases for PDE and JS are close, and in the aftershock region, the multi-fractal relation for five moving ensembles seems around an invariant curve, though the moving time t_0 differs [26].

Let us see that the relative fraction parameter $R_\delta(\tau_m)$ is an important quantity to estimate the effect of the correlation between different magnitude scales m and M. Then the original embedding equation (Eq. (2.1)) is rewritten into:

$$N_m p_m(\tau_m) R_\delta(\tau_m) = N_M \frac{\tau_m p_M(\tau_m) \Delta_{mM}(\tau_m)}{\int_0^{\tau_m} x\ \Xi_{mM}(x, \tau_m) dx + \tau_m \Delta_{mM}(\tau_m)}. \tag{2.11}$$

If the magnitude of each shock is independently random [14][15], $R_\delta(\tau_m)$ is evaluated in the following way. In the case of Fig. 2.1(1), the magnitude of both ends of the events that sandwich τ_m have to be larger than m. On the other hand, if the interval τ_m satisfies the case of Fig. 2.1(2), the magnitudes of two successive shocks are both larger than M. Therefore, if there is no correlation between the time interval

Table 2.3: Fitting parameters in data-analysis (for four significant digits).

Time series	λ	μ	ν	p	q	r
PDE ($M \in [5, 6.4]$)	4.191×10^4	0.9290	0.8510	3.850	1.203	0.09267
PDE ($M \in [5, 7.5]$)	3.812×10^6	1.329	0.9748			
JS ($M \in [3, 4.6]$)	46.10	0.6205	0.8029	1.524	0.4393	0.1597
JS ($M \in [3, 5]$)	48.32	0.6315	0.8190			
JA ($t_0 = 100$)	19.65	0.5152	0.8117	5.461	2.248	0.08072
JA ($t_0 = 200$)	3.424	0.3945	0.9685	3.086	0.5549	0.1694
JA ($t_0 = 300$)	5.425	0.3254	0.5751	4.494	1.693	0.09188
JA ($t_0 = 400$)	2.296	0.1760	0.4390	9.822	6.483	0.04114
JA ($t_0 = 500$)	16.96	0.6528	0.8932	4.964	2.160	0.08525

[14] Some earthquake models use this hypothesis to explain the real seismicity, see Ref. [50].

[15] Note that the discussion here implicitly assumes not only independence between magnitudes but also independence between magnitude and inter-event time. The parameter $R_\delta(\tau_m)$ is a quantity that can reflect the correlation between magnitudes, and that is the reason we referred to it as the magnitude correlation coefficient, though it also can include the correlation between magnitude and inter-event time. For example, in the case of the ETAS model, although the magnitudes are generated totally randomly, Eq. (2.12) holds only for the stationary Poisson process (when $\lambda(t) \equiv \lambda_0$ in Eq. (1.3)); in other cases, the time interval following an event is affected by the magnitude of that event, and Eq. (2.12) does not hold reflecting such a correlation.

τ_m and the magnitudes of the two intervening events, the parameter $R_\delta(\tau_m)$ does not depend on the variable τ_m, and $R_\delta(\tau_m)$ is described only by the magnitude distribution given in the GR law of Eq. (1.2):

$$R_\delta(\tau_m) = \frac{P(M)^2}{P(m)^2} = 10^{-2b\Delta m}, \tag{2.12}$$

where the GR law ($N_M/N_m = P(M)/P(m) = 10^{-b\Delta m}$) is used. Equation (2.11) is rewritten into:

$$p_m(\tau_m) = 10^{b\Delta m} \frac{\tau_m p_M(\tau_m)\Delta_{mM}(\tau_m)}{\int_0^{\tau_m} x\ \Xi_{mM}(x,\tau_m)dx + \tau_m \Delta_{mM}(\tau_m)}, \tag{2.13}$$

where $N_m/N_M = 10^{b\Delta m}$ is used again. This equation is a simplified form of the embedding equation applied to the case where no correlation exists between different magnitudes.

However, the data analysis in many cases shows us that the parameter $R_\delta(\tau_m)$ deviates from Eq. (2.12). This suggests the possibility that the magnitudes of successive two shocks are not independent but correlated. In fact, such magnitude correlation has been pointed out in the study of actual earthquakes [40, 68], and it has been expected to obtain the more precise statistical law from the correlation analysis in the inter-event time statistics.

Here we introduce a correction factor $\eta_{mM}(\tau_m)$ to describe the correlation instead of Eq. (2.12):

$$R_\delta(\tau_m) = 10^{-2b\Delta m}\eta_{mM}(\tau_m), \tag{2.14}$$

and consider the embedding equation in the following form:

$$p_m(\tau_m)\eta_{mM}(\tau_m) = 10^{b\Delta m} \frac{\tau_m p_M(\tau_m)\Delta_{mM}(\tau_m)}{\int_0^{\tau_m} x\ \Xi_{mM}(x,\tau_m)dx + \tau_m \Delta(\tau_m)}. \tag{2.15}$$

The factor $\eta_{mM}(\tau_m)$ is referred to as the magnitude correlation coefficient (MCC) in the present chapter because it can reflect a correlation between two different magnitudes (m and M) and expresses the correction in the GR law.

Figure 2.7 shows that the coefficient $\eta_{mM}(\tau_m)$ satisfies the following empirical scaling property [16],

$$\eta_{mM}(\tau_m) = 1 + \overline{\eta}\left(\frac{\tau_m}{\langle\tau_M\rangle}\right)\Delta m, \tag{2.16}$$

where $\overline{\eta}(x)$ is a characteristic factor depending on the rescaled parameter, i.e., $x = \frac{\tau_m}{\langle\tau_M\rangle}$. In the calculation of the MCC, we consider the cases where the number of samples (N) is large enough to obtain the stable evaluation $\overline{\eta}(\tau_m/\langle\tau_M\rangle)$; the MCC is obtained for $N \geq N^m_{\min}$ (at τ_m) in Table 2.2. The values of the cut-off interval $\tau_{\min}$ are fixed constant in each calculation.

Based on the scaling property suggested in Eq. (2.16), the scaling function $\overline{\eta}(x)$ is calculated for the mixed population of the samples at the cut-off magnitude pairs (m, M) with the same Δm [17]. This method enables us to obtain a clearer form of the scaling function, avoiding the shortage of the sample number, in particular at large Δm. In this calculation, no threshold is set for time intervals ($\tau_{\min}$). As Fig. 2.7(a, b) shows, such calculated $\overline{\eta}(x)$ is almost the same as those from a specific cut-off magnitudes

[16] Equation (2.14) includes the GR law, though the b-value does not come into play calculating $\eta_{mM}(\tau_m)$ for a specific pair of (m, M), because in the actual calculation $(\langle\tau_m\rangle/\langle\tau_M\rangle)^2$ with numerically obtained $\langle\tau_m\rangle$ and $\langle\tau_M\rangle$ are used instead of $10^{-2b\Delta m}$. In this chapter, the b-values calculated by the maximum likelihood estimate are used differently from Ref. [1], though the results of $\overline{\eta}(x)$ are the same. However, for the calculation using mixed samples for the (m, M)-pairs with the same Δm, the b-values are necessary and those calculated by the maximum likelihood estimate are used.

[17] The similar approach is used in Chapter 5. The outline of the method used here is as follows. Let us define the transformation of the lower inter-event times $\varsigma := \log_{10}\frac{\tau_m}{\langle\tau_M\rangle}$. Let $\{m_0, m_1, ..., m_n\}$ be the cut-off magnitudes to be set for each time series ($m_j = m_0 + 0.1j$). For a given Δm, the lower and upper cut-off magnitudes can be $\{m_0, m_0 + \Delta m\}, ..., \{m_{n-10\Delta m}, m_n\}$. Let $\hat{N}_{m_j, m_j+\Delta m}(\varsigma)$ denote the number of such transformed lower inter-event time intervals whose length within $[\varsigma, \varsigma + \Delta\sigma]$ when the upper and the lower cut-off magnitudes are $\{m_j, m_j + \Delta m\}$. Also, let $\hat{N}^\delta_{m_j, m_j+\Delta m}(\varsigma)$ denote the number of such lower intervals obtained in the manner shown in Fig. 2.1(2). Thus, for a Δm, the coefficient $\overline{\eta}(\varsigma)$ can be calculated as: $\overline{\eta}(\varsigma) = \frac{\left(\frac{\sum_{j=0}^{n-10\Delta m}\hat{N}^\delta_{m_j,m_j+\Delta m}(\varsigma)}{\sum_{j=0}^{n-10\Delta m}\hat{N}_{m_j,m_j+\Delta m}(\varsigma)}\right)10^{2b\Delta m}-1}{\Delta m}$.

pair (m, M), indicating that the scaling relation in Eq. (2.16) is reasonable. However, such consistency is not observed in the cases of JA; this is considered to be due to the effect of the b-value. In this calculation, the b-value calculated in the range $M \in [3:5]$ was used, though, as can be seen in Fig. 2.2(c), the linearity of $\log_{10} p(M)$ in this magnitude range is not good. This is likely to affect the result of $\overline{\eta}(x)$ obtained from mixed samples. Moreover, $\overline{\eta}(x)$ shows gradual deviation as Δm gets large even in Figs. 2.7(a, b), suggesting the scaling relation in Eq. (2.16) is not perfect in particular for large Δm.

Figure 2.7 also compares the results with those for the time series with reshuffled magnitudes; in particular, Figs. 2.7(a,b) shows that the magnitude correlations are reflected in the scaling function $\overline{\eta}_{mM}(x)$. The scaling function $\overline{\eta}_{mM}(x)$ shows the tendency that it takes larger positive value at short intervals (x) and gradually decreases to around a negative value (around $[-1, 0]$) as the length of the interval (x) gets longer. Such a tendency is relatively clearly observed in the case of PDE and JS. In particular, in the case of JA, the results obtained for a specific combination of (m, M) seem to show the above-described tendency partially. Although the results calculated for mixed populations for (m, M) combinations with the same Δm are vague, these also take value around $[-1, 0]$ at large x. Figure 2.7(h), which is the result of mixed populations for all time windows of JA, shows a relatively clear trend similar to that of PDE and JS.

In the stationary and mixture cases (PDE, JS) treated in this study, the scaling form of Eq. (2.16) reveals almost the same behavior and is surmised to be universal. Even in the non-stationary (aftershock) case (JA), the factor $\overline{\eta}(\tau_m/\langle\tau_M\rangle)$ seems to be described by a similar curve in all cases for the moving ensembles at different time.

2.4.2 Inter-event Time Correlation Coefficients $a_{mM}(\tau_M)$ and its Scaling Property

This subsection derives another phenomenological coefficient $a_{mM}(\tau_M)$, which describes the correlation between two inter-event times τ_m and τ_M. Here we assume that the conditional probability $p_{mM}(\tau_m|\tau_M)$ approaches to the stationary one $p_m(\tau_m)$ when the interval τ_M (or the difference $\tau_M - \tau_m$) goes to large enough, i.e., $\tau_M \to \infty$ (or $\Delta m = M - m \to \infty$), the conditional probability is approximated by $p_m(\tau_m)$ similarly to Eq. (2.4):

$$p_{mM}(\tau_m|\tau_M) \approx p_m(\tau_m)\theta(\tau_M - \tau_m) + P_m(\tau_m)\delta(\tau_M - \tau_m), \tag{2.17}$$

or:

$$Q_{mM}(\tau_m|\tau_M) \approx Q_m(\tau_m)\theta(\tau_M - \tau_m) + \theta(\tau_m - \tau_M), \tag{2.18}$$

where $Q_{mM}(\tau_m|\tau_M)$ is the cumulative distribution of the conditional probability; $Q_{mM}(\tau_m|\tau_M) := \int_0^{\tau_m} p_{mM}(x|\tau_M)dx$.

The examples of the comparison between $Q_{mM}(\tau_m|\tau_M)$ and $Q_m(\tau_m)$ for selected pairs of cut-off magnitudes for each time series shown in Fig. 2.8 suggest that the above approximation (2.18) seems satisfactory when Δm and τ_M are large enough. When the above conditions break, we surmise here that Eq. (2.18) is extended to the following generalized form:

$$Q_{mM}(\tau_m|\tau_M) = a_{mM}(\tau_M)Q_m(\tau_m)\theta(\tau_M - \tau_m) + \theta(\tau_m - \tau_M), \tag{2.19}$$

where $a_{mM}(\tau_M)$ is a correction factor which obeys $a_{mM}(\tau_M) \to 1$ as Δm goes to large, and $a_{mM}(\tau_M) \to 0$ as $\Delta m \to 0$. By integrating Eq. (2.19), one can obtain the factor $a_{mM}(\tau_M)$ directly from the data analysis, i.e.:

$$a_{mM}(\tau_M) = \frac{\tau_M - \langle\langle\tau_m\rangle\rangle_{\tau_M}}{\int_0^{\tau_M} Q_m(x)dx}. \tag{2.20}$$

The factor $a_{mM}(\tau_M)$ is referred to as the inter-event time correlation coefficient (ICC) in the present chapter because it describes a correlation among different inter-event times (τ_m and τ_M). Figure 2.9 shows the scaling property of the ICC:

$$a_{mM}(\tau_M) = \overline{a}_{\Delta m}\left(\frac{\tau_M}{\langle\tau_M\rangle}\right), \tag{2.21}$$

[18] In Ref. [1], the range of $\tau_M/\langle\tau_M\rangle$ to show was $\tau_M/\langle\tau_M\rangle \in [1, 10]$. Figure 2.8 shows wider range than it.

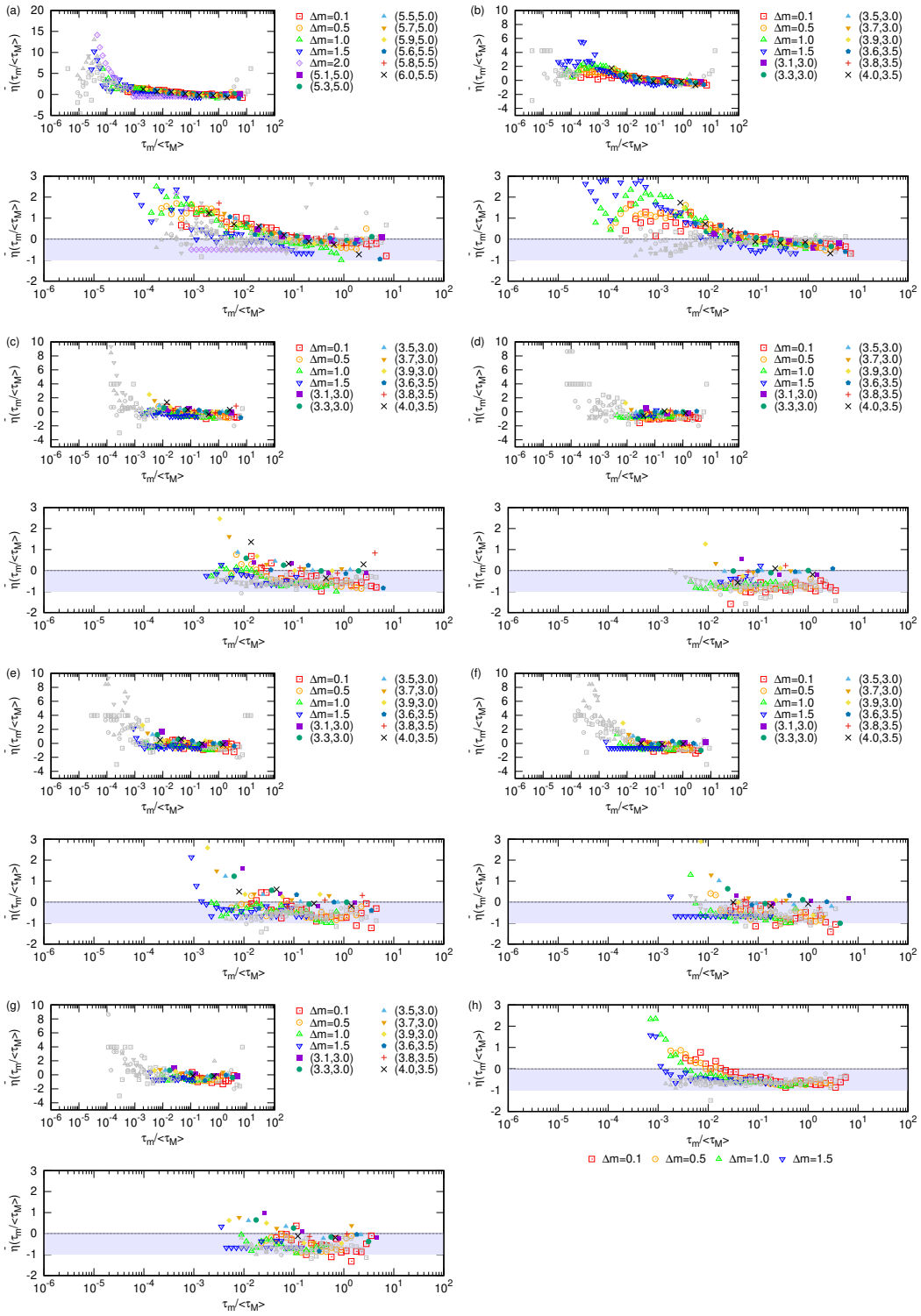

Figure 2.7: The scaling in Eq. (2.16) for (a) PDE, (b) JS, (c) JA($t_0 = 100$), (d) JA($t_0 = 200$), (e) JA($t_0 = 300$), (f) JA($t_0 = 400$), and (g) JA($t_0 = 500$). (h) shows the result of the calculation using all the time series from JA($t_0 = 100$) to JA($t_0 = 500$). (Upper panels) Rescaled correction function $\overline{\eta}(\tau_m/\langle\tau_M\rangle)$ calculated at some (m, M)-pairs and computed for mixed samples of (m, M) with the same Δm as the population. The gray-colored symbols are the results with the number of samples N being less than $N^m_{\min}$. Some very large symbols are out of the range shown in the figure. (Lower panels) The enlarged view of the results in the upper panel satisfying $N \geq N^m_{\min}$. The gray-colored symbols show the results for the time series with magnitudes shuffled. The blue-colored band indicates the approximate range for $\overline{\eta}(1)$.

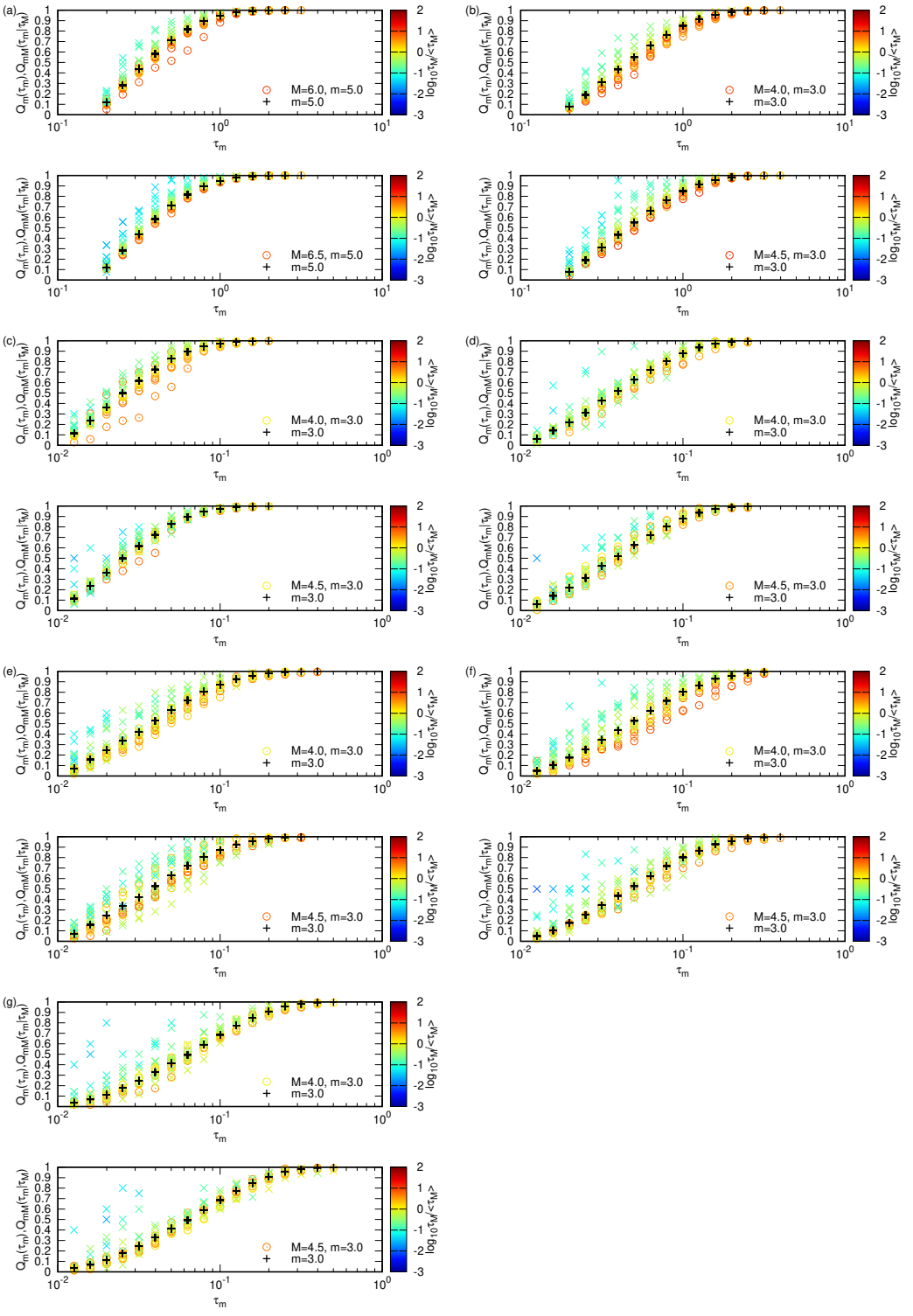

Figure 2.8: Comparison between $Q_{mM}(\tau_m|\tau_M)$ and $Q_m(\tau_m)$ at two cut-off magnitudes in (a) PDE, (b) JS, (c) JA($t_0 = 100$), (d) JA($t_0 = 200$), (e) JA($t_0 = 300$), (f) JA($t_0 = 400$), and (g) JA($t_0 = 500$). Equation (2.18) is approximately satisfactory for $\Delta m > 1.0$ and $\tau_M/\langle\tau_M\rangle > 1$ [18].

which suggests that $a_{mM}(\tau_M)$ does not explicitly depend on the magnitude (m and M) but a function of the difference Δm and is uniquely determined by a rescaled variable $\tau_M/\langle\tau_M\rangle$. Note that in the calculation of the ICC, we consider the cases where the number of samples (N) is large enough ($N \geq N^M_{\min}$ (at τ_M) in Table 2.2) to obtain the stable evaluation of ICC ($\overline{a}_{\Delta m}(\tau_M/\langle\tau_M\rangle)$).

Based on the above-suggested scaling property, we calculated the scaling function using the samples for the cut-off magnitude pairs (m, M) with the same Δm as a mixed population in the same manner as the MCC [19]. The resulting scaling functions are almost the same as those for the specific pairs (m, M), which supports the validity of the scaling property in Eq. (2.21). In particular, similarly to the MCC, the scaling functions $\overline{a}_{\Delta m}(\tau_M/\langle\tau_M\rangle)$ for JA are nearly the same regardless of t_0.

Figure 2.10 shows the Δm-dependence of $a_{\Delta m}(1)$ for each time series calculated for the mixed populations of (m, M) pairs with the same Δm; the method to use mixed populations enabled us to determine in more detail such a Δm dependence. As shown in the figure, the Δm dependence of $a_{\Delta m}(1)$ can be fit by the exponential function $1 - \exp(-\rho\Delta m)$, and the derivative of the scaling function $\overline{a}'_0(1)$ can be estimated as ρ. Table 2.4 summarizes the estimated value of ρ ($= \overline{a}'_0(1)$) for each time series.

The embedding equation (Eq. (2.15)) is finally rewritten into the following by using Eqs. (2.18) – (2.21),

$$p_m(\tau_m)\left(1+\overline{\eta}\left(\frac{\tau_m}{\langle\tau_M\rangle}\right)\Delta m\right) = 10^{b\Delta m}\frac{\tau_m\left(1-\overline{a}_{\Delta m}\left(\frac{\tau_m}{\langle\tau_M\rangle}\right)Q_m(\tau_m)\right)}{\tau_m-\overline{a}_{\Delta m}\left(\frac{\tau_m}{\langle\tau_M\rangle}\right)\int_0^{\tau_m}Q_m(x)dx}p_M(\tau_m). \tag{2.22}$$

The point is that the original embedding equation is expressed by two empirical coefficients $\overline{\eta}\left(\frac{\tau_m}{\langle\tau_M\rangle}\right)$ and $\overline{a}_{\Delta m}\left(\frac{\tau_M}{\langle\tau_M\rangle}\right)$ which reflect the magnitude correlation and the inter-event time correlation, respectively. The next section derives the multi-fractal relation from Eq. (2.22).

2.5 Approach to the Multi-fractal Relation

2.5.1 Derivation of the Multi-fractal Relation

The multi-fractal relation is considered to be a new statistical law that represents the important combination of the GR law and the inter-event time statistics. In this section, using the properties found in Section 2.4 on the conditional probability, we discuss the theoretical background of the multi-fractal relation from the embedding equation and give an approximate calculation of the multi-fractal relation to compare with the results of data analysis in Section 2.3.2.

Table 2.4: Estimated value of $\overline{a}'_{\Delta m}(1)$ based on the fitting of $\overline{a}_{\Delta m}(1)$ by the exponential function (up to four decimal places).

Time series	Estimate of $\overline{a}'_0(1)$
PDE	7.1175
JS	4.4046
JA ($t_0 = 100$)	6.0932
JA ($t_0 = 200$)	6.3019
JA ($t_0 = 300$)	4.9168
JA ($t_0 = 400$)	5.8595
JA ($t_0 = 500$)	5.2134

[19] The calculation used here is as follows. First, we numerically obtain $\langle\langle\tau_m\rangle\rangle_{\tau_M}$ and $\int_0^{\tau_M}Q_m(x)dx$. Let $n_j(\sigma)$ be the number of the upper intervals within the transformed length $[\sigma, \sigma+\Delta\sigma]$ (where $\sigma := \log_{10}\frac{\tau_M}{\langle\tau_M\rangle}$ and $\Delta\sigma = 0.1$) for a given Δm and a pair of the cut-off magnitudes $\{m_j, m_j+\Delta m\}$. Thus, compute the addition of $a^k_{m_j m_j+\Delta m}(\sigma)$ for these upper intervals in the time series for all the pairs of $\{m_j, m_j+\Delta m\}$ with the same Δm and take the average; $a_{\Delta m}(\sigma) = \frac{\sum_{j=0}^{n-10\Delta m}\sum_{k=0}^{n_j(\sigma)}a^k_{m_j m_j+\Delta m}(\sigma)}{\sum_{j=0}^{n-10\Delta m}\sum_{k=0}^{n_j(\sigma)}1}$.

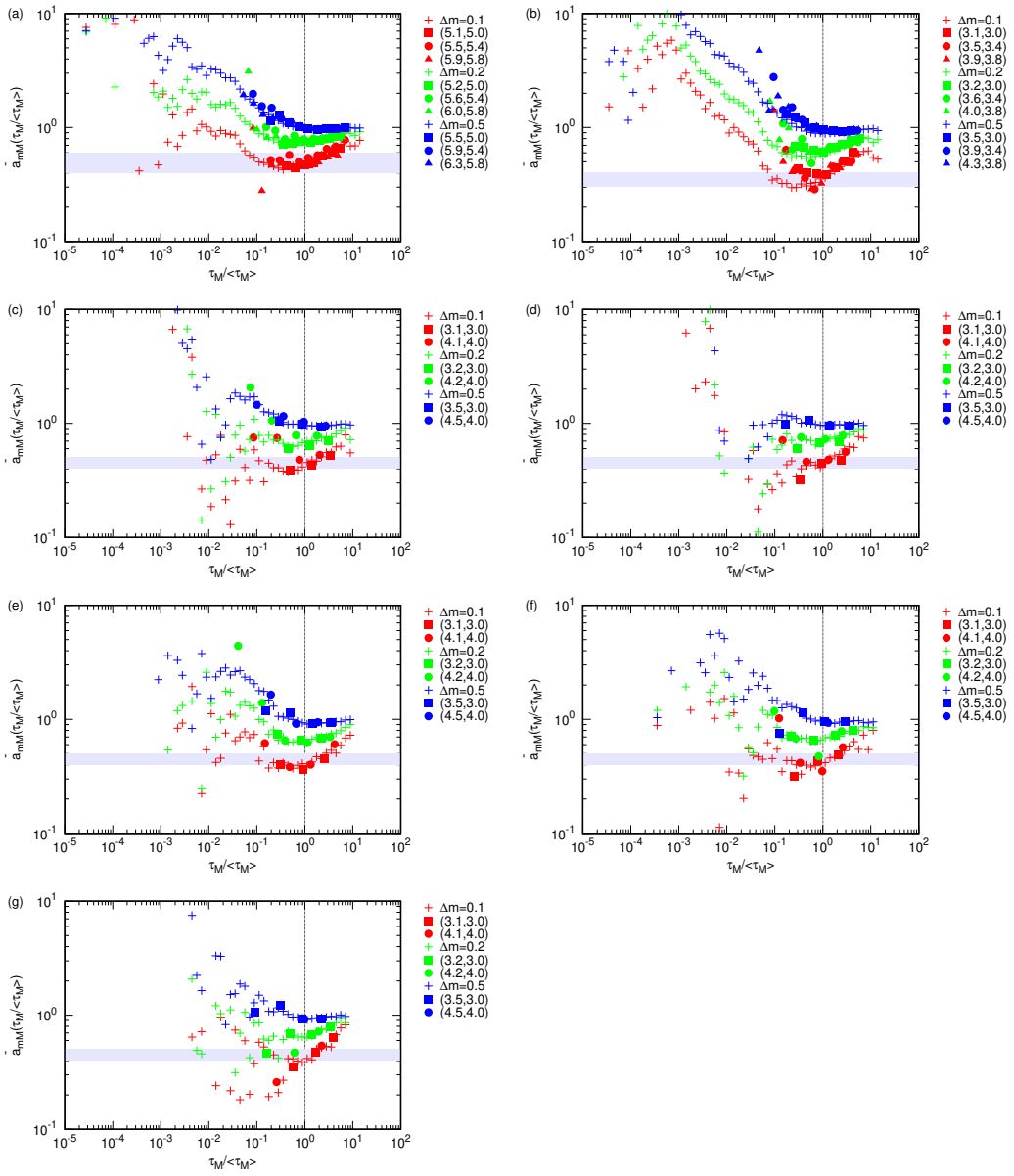

Figure 2.9: Rescaled correction coefficient $\overline{a}_{\Delta m}(\tau_M/\langle\tau_M\rangle)$ for (a) PDE, (b) JS, (c) JA($t_0 = 100$), (d) JA($t_0 = 200$), (e) JA($t_0 = 300$), (f) JA($t_0 = 400$), and (g) JA($t_0 = 500$). The results for specific pairs of cut-off magnitudes are shown by filled symbols and for the mixed populations of (m, M) with the same Δm by empty symbols. The blue-colored band shows a rough range of the $a_{0.1}(1)$ value for each time series.

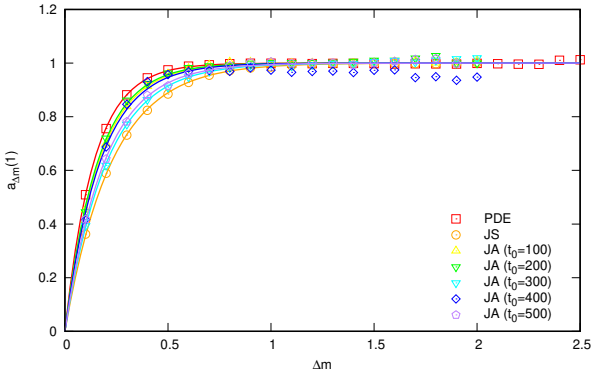

Figure 2.10: Δm dependence of $\overline{a}_{\Delta m}(1)$ and the fitting by the exponential function for each time series.

Let us start with Eq. (2.22). Data analysis has suggested that the Weibull distribution can approximate the inter-event time statistics. Especially, in the case when τ_m ($< \tau_M$) is large enough, not only $p_M(\tau_M)$ but also $p_m(\tau_m)$ can be adjusted by the Weibull distribution. Therefore, our purpose is to derive a relation between two Weibull parameters (α_m, β_m) from the embedding equation (Eq. (2.22)) by substituting the Weibull distributions $p_m(\tau_m) \approx p_m^W(\tau_m)$ and $p_M(\tau_M) \approx p_M^W(\tau_M)$ in Eq. (1.10). Equation (2.22) is rewritten as follows:

$$\begin{aligned}
&\left(\frac{\alpha_m}{\beta_m}\right)\left(\frac{\tau_m}{\beta_m}\right)^{\alpha_m-1} e^{-\left(\frac{\tau_m}{\beta_m}\right)^{\alpha_m}}\left[1+\overline{\eta}\left(\frac{\tau_m}{\langle\tau_M\rangle}\right)\Delta m\right] \\
&= \frac{\beta_M\Gamma\left(1+\frac{1}{\alpha_M}\right)}{\beta_m\Gamma\left(1+\frac{1}{\alpha_m}\right)} \frac{\tau_m\left[1-\overline{a}_{\Delta m}\left(\frac{\tau_m}{\langle\tau_M\rangle}\right)\left(1-e^{-\left(\frac{\tau_m}{\beta_m}\right)^{\alpha_m}}\right)\right]}{\tau_m-\overline{a}_{\Delta m}\left(\frac{\tau_m}{\langle\tau_M\rangle}\right)\left[\tau_m-\frac{\beta_m}{\alpha_m}\gamma\left(\frac{1}{\alpha_m},\left(\frac{\tau_m}{\beta_m}\right)^{\alpha_m}\right)\right]}\left(\frac{\alpha_M}{\beta_M}\right)\left(\frac{\tau_m}{\beta_M}\right)^{\alpha_M-1} e^{-\left(\frac{\tau_m}{\beta_M}\right)^{\alpha_M}},
\end{aligned} \tag{2.23}$$

where $\gamma(x,y)$ is the lower incomplete gamma function.

One can say that Eq. (2.23) approximates at the arbitrary time τ_m if the Weibull approximation is satisfied, but in our data analysis we have to be careful that the Weibull approximation breaks in the case of small magnitude as well as in the case of small τ_m, where it is known that the inter-event time statistics reveals a gradual change towards the Log-Weibull distribution from the Weibull distribution [90, 91]; consequently, we have to treat Eq. (2.23) under the following two conditions: (1) m is large enough ($m \leq M$, $m \to M$) and (2) τ_m ($\leq \tau_M$) is also large enough (i.e., $\tau_m \approx \langle\tau_M\rangle$). Under these two conditions, Eq. (2.23) can be rewritten into the multi-fractal relation between two Weibull parameters α_m and β_m (see Appendix A):

$$\begin{aligned}
\ln\left(\frac{\beta_m}{\beta_{m_0}}\right) = \int_{m_0}^{m} dm \frac{1}{(\alpha_m-1)-\alpha_m\Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}}\Bigg\{-\overline{\eta}(1)+\overline{a}_0'(1)\Bigg[e^{-\Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}} \\
-\frac{\gamma\left(\frac{1}{\alpha_m},\Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}\right)}{\alpha_m\Gamma\left(1+\frac{1}{\alpha_m}\right)}\Bigg]+\alpha_m'\left[\frac{1}{\alpha_m}-\frac{\psi\left(1+\frac{1}{\alpha_m}\right)}{\alpha_m^2}+\left(1-\Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}\right)\log\Gamma\left(1+\frac{1}{\alpha_m}\right)\right]\Bigg\},
\end{aligned} \tag{2.24}$$

where we need the scaling forms obtained in Section 2.2 (Eqs. (2.16) and (2.21)), $\psi(x)$ is the digamma function, and $\alpha_m' := d\alpha_m/dm$, $\beta_m' := d\beta_m/dm$. m_0 is a reference point of the integration; in our analysis, m_0 is fixed to $m_0 = 5.0$ for PDE, 3.0 for JS, and 3.0 for JA, respectively, considering the Weibull approximation is satisfied at these cut-off magnitudes.

Figure 2.11 shows the numerical results of Eq. (2.24) calculated with the fitting parameter function Eq. (2.7) of α_M with the parameter values of $\{\lambda, \mu, \nu\}$ in Table 2.3 and the representative values of $\overline{\eta}(1)$ and $\overline{a}_0'(1)$ determined from the numerical results in Section 2.2; as shown by the blue-colored bands in Figs. 2.7 and 2.9, these parameter values may fluctuate within a certain range; from the results in Fig.

2.7, $\overline{\eta}(1)$ may deviate within the range $[-1, 0]$. Also, from the results in Fig. 2.9 and Table 2.4, $\overline{a}'_0(1)$ may deviate about $[4, 7.2]$ for PDE, $[3, 4.5]$ for JS, and $[4, 6.4]$ for JA. Thus, it is important to consider the impact of these fluctuations on the numerically computed multi-fractal relations. However, we determined a representative of these parameter values from these ranges to be $(\overline{\eta}(1), \overline{a}'_0(1)) = (-0.5, 5.0)$ for PDE, $(-0.3, 4.0)$ for JS, and $(-0.4, 5.0)$ for JA($t_0 = 100 \sim 500$), respectively (the value of $\overline{a}'_0(1)$ is roughly inferred by the value of $\Delta m = 0.1$). The effect of the changes in these values on the numerical results of the multi-fractal relation will be discussed later.

The results in Fig. 2.11 show that the approximation given by Eq. (2.24) describes qualitatively the tendency of the multi-fractal relation obtained by catalog analysis for either stationary (PDE), mixture (JS), and non-stationary (JA) cases. In particular, in the non-stationary (aftershock) sequence, it is striking that the theoretically derived multi-fractal relations in each moving ensemble seem to obey nearly the same invariant curve. However, these derived multi-fractal relation curves are imperfect in that they gradually separate from the results of catalog analysis at large cut-off magnitudes; the cause of which is discussed in the following.

Figures 2.12 – 2.14 show the results of the additional numerical calculations using the parameter values of $\overline{\eta}(1)$ and $\overline{a}'_0(1)$ in the above range. These figures indicate that the effect from $\overline{a}'_0(1)$ is significant; the larger values of $\overline{a}'_0(1)$, which is obtained from the more detailed estimation of $\overline{a}'_0(1)$ in Fig. 2.10, can better describe the shape of the multi-fractal relations at large Δm. On the other hand, the difference at small Δm seems to widen. The value of $\overline{\eta}(1)$ also affects the multi-fractal relation; from Fig. 2.7, a slight tendency that $\overline{\eta}(1)$ gets smaller as Δm gets larger can be observed, and this tendency does not compensate but rather increase the difference between the theoretical curve and the catalog analysis result. Therefore, although the fluctuations in the scaling functions related to the conditional probability can qualitatively better describe the hyperbolic-shaped multi-fractal relation, it is difficult to explain in full detail; this suggests the necessity of the detailed consideration of the conditional probability, especially the assumptions put on the two correlation coefficients.

In order to analyze the properties of the conditional probability and the embedding equation, we introduced these two correlation coefficients as a stepping stone, though, in fact, they have problems. The MCC, as already pointed out, is not the smallest unit of the inter-correlation between earthquakes in that it can include other correlations than magnitude-magnitude correlation. Also, the scaling property of the MCC is empirical, and the mixing method of the samples with the same Δm suggests the tendency of the scaling function to gradually deviate for large Δm. The other coefficient ICC is assumed to be the univariate function, though the analytically derived conditional probability in later Chapter 3 for the case of the stationary marked Poisson process is the bivariate function [20], and thus it is a rough approximation. These points can cause quantitative deviations in the theoretically derived multi-fractal relation from the result of the catalog analysis that can not be compensated by the parameter values.

Nevertheless, the derived multi-fractal relation based on these assumptions can qualitatively explain the hyperbolic shape, and it is shown that the hierarchical structure of the seismic activity is reflected in the multi-fractal relation, including its universality such as the time-independence in aftershock sequence. For a detailed quantitative description of the multi-fractal relation, a more accurate understanding of the conditional probability and, furthermore, derivation of its functional form is necessary. The following chapters gradually clarify the properties and the functional form of this conditional probability; Chapter 3 derives its functional form for the stationary marked Poisson process, and Chapter 5 for time series with weak inter-event correlations.

[20] The analytical result in Chapter 3 shows that the assumption put on the ICC (Eq. (2.19)) is incorrect; a_{mM} is a bivariate function $a_{mM}(\tau_m, \tau_M)$, and this assumption leads to $\beta_m = \beta_{m_0} 10^{(e-2)b(m-m_0)}$ for the stationary marked Poisson process, which is actually $\beta_m \propto 10^{b(m-m_0)}$. Therefore, the assumption causes the discrepancy between the theoretical result and catalog analysis by the factor of $\sim 10^{0.71}$, shortening the estimate of the β_m value. This simple result suggests that the quantitative discrepancy between the theoretical results and catalog analysis in the multi-fractal relation can be caused by the assumptions put on the ICC and possibly MCC, and therefore, a detailed analysis of the conditional probability starting from a simple time series is necessary.

[21] In Ref. [1], the theoretical curve of PDE was incorrectly drawn in the magnitude range $[5, 7.5]$. The theoretical curves of the multi-fractal relation for the magnitude ranges of $[5, 6.4]$ and $[5, 7.5]$ are correctly shown in Fig. 2.11(a).

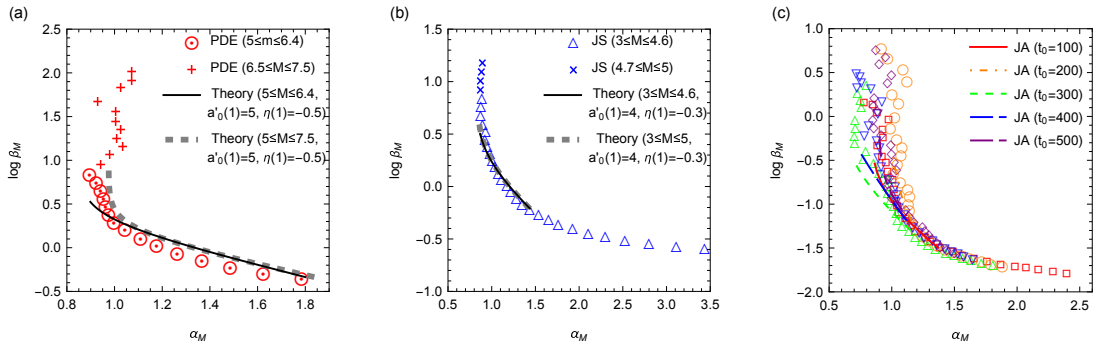

Figure 2.11: The multi-fractal relation obtained by (symbols) the catalog analysis and ((dotted) curves) Eq. (2.24) calculated using the fitting curve of α_M and the representative values of $(\overline{\eta}(1), \overline{a}_0'(1))$ for (a) PDE, (b) JS, and (c) JA. (a) The symbol $\odot$ shows the $\{\alpha_M, \beta_M\}$ obtained by catalog analysis in the range $M \in [5, 6.4]$ analyzed in Ref. [1] and $+$ are in the range $M \in [6.5, 7.5]$. The solid (dotted) curve shows the numerical calculation results using the fitting function of α_M only within the magnitude range $[5, 6.4]$ ($[5, 7.5]$) in which the fluctuation in numerical data analysis results is less. The integral in Eq. (2.24) was performed in the range $M \in [5, 6.4]$ [21] ($[5, 7.5]$). (b) The $\triangle$ symbols are the results of catalog analysis in the range $M \in [3, 4.6]$ analyzed in Ref. [1] and $\times$ in the range $M \in [4.7, 5]$. The solid (dotted) curve shows Eq. (2.24) using the fitting function of α_M within the magnitude range $[3, 4.6]$ ($[3, 5]$) and calculated in the integration in Eq. (2.24) for $m \in [3, 4.6]$ ($m \in [3, 5]$). (c) The symbols show the multi-fractal relation obtained by catalog analysis (the same symbol as in Fig. 2.5) and the (dotted) curves show Eq. (2.24) calculated within $M \in [3, 5]$, which is the same as the magnitude range of the fitting of α_M.

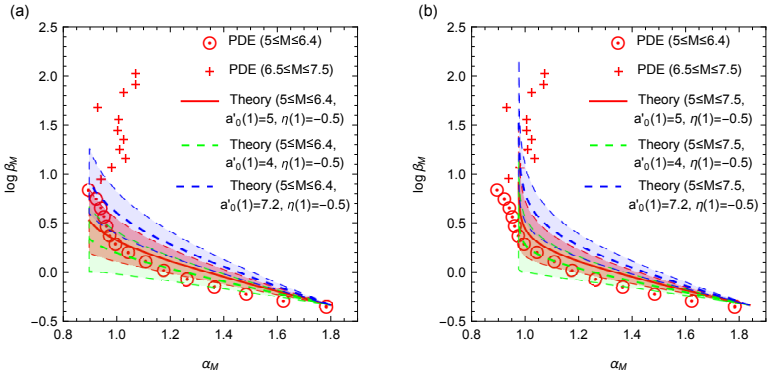

Figure 2.12: Multi-fractal relations calculated from Eq. (2.24) with the parameter values within the range $\overline{a}_0'(1) \in [4, 7.2]$ and $\overline{\eta}(1) \in [-1, 0]$ and the fitting function of α_M in (a) $M \in [5, 6.4]$ and (b) $M \in [5, 7.5]$. The results for three values of $\overline{a}_0'(1) = 4, 5, 7.2$ with $\overline{\eta}(1) = -0.5$ are shown by (dotted) curves, and from these, the range for varying the value of $\overline{\eta}(1)$ is shaded.

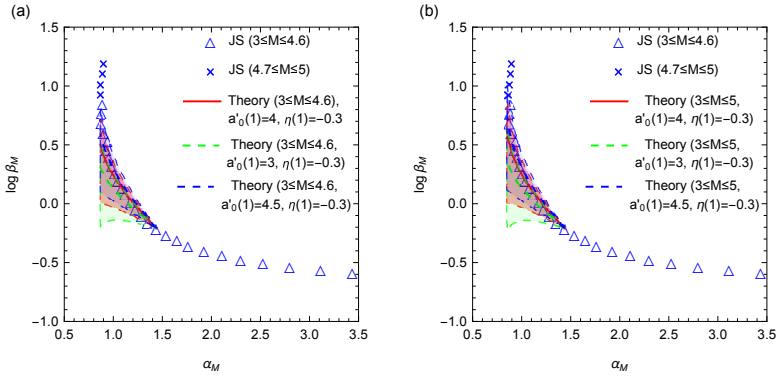

Figure 2.13: Equation (2.24) for JS with $\overline{a}_0'(1) \in [3, 4.5]$ and $\overline{\eta}(1) \in [-1, 0]$. The fitting function of α_M in the magnitude range (a) $M \in [3, 4.6]$ and (b) $M \in [3, 5]$. The three cases of $\overline{a}_0'(1)$ with $\overline{\eta}(1) = -0.3$ are shown by (dotted) curves, and the ranges when $\overline{\eta}(1)$ varies in $[-1, 0]$ are shown by the shaded area.

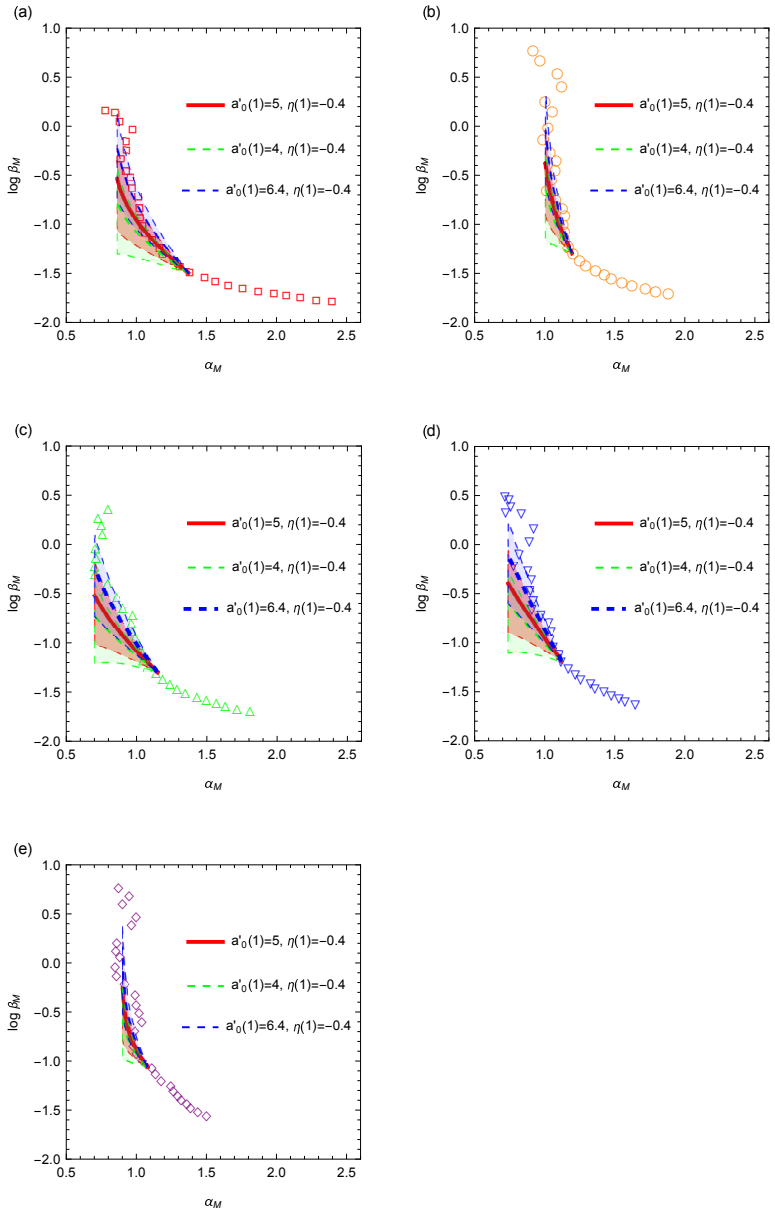

Figure 2.14: Equation (2.24) for five time windows of JA; (a) JA($t_0 = 100$), (b) JA($t_0 = 200$), (c) JA($t_0 = 300$), (d) JA($t_0 = 400$), and (e) JA($t_0 = 500$). The parameter values are set in the range $\overline{a}'_0(1) \in [4, 6.4]$ and $\overline{\eta}(1) \in [-1, 0]$. The results for three values of $\overline{a}'_0(1)$ with $\overline{\eta}(1) = -0.4$ are shown by (dotted) curves and the range Eq. (2.24) sweeps when $\overline{\eta}(1)$ varies within $[-1, 0]$ is shown by shaded region.

2.6 Summary and Discussions

In this chapter, we developed the Embedding Equation Theory (EET) to derive detailed information about the inter-event time statistics of seismicity. The conditional probability is described by two correlation coefficients; one is the magnitude coefficient, and another is the inter-event time coefficient, and the scaling forms of those coefficients are numerically determined by using the seismic data in the PDE catalog and the JMA catalog. Furthermore, it is striking that the scaling form of each coefficient seems to be universal not only in the stationary and mixture ensembles but also in the non-stationary (aftershock) ensembles.

The empirical scaling laws obtained by the EET enable us to derive a new statistical feature in seismic activity. Indeed, EET yields a theoretical foundation for the multi-fractal relation, and it is shown that the approximate calculation of the multi-fractal relation reveals qualitative agreement with the numerical results, which are obtained under the condition that the inter-event time distribution obeys the Weibull statistics.

However, theoretically derived multi-fractal relations do not give a quantitatively full explanation of the catalog analysis, and it is important to further understand the nature of the conditional probability that characterizes the hierarchical structure of seismic time series and to clarify its functional form. In the subsequent chapters, the properties of the conditional probability are examined through the consideration of the Bayesian method to improve the probabilistic forecasting method of earthquakes.

Chapter 3

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

This chapter is based on Ref. [2], which is paper 2 in the list of author's papers, and Ref. [93].

3.1 Introduction

Chapter 2 introduced the conditional probability in marked point processes that yields the statistical relation between inter-event times at different magnitude thresholds (m and M ($= m + \Delta m$), $\Delta m > 0$). It can be considered that the information on the intervals at the lower magnitude threshold (m) can be utilized for estimating the length of the upper interval through the conditional probability. This chapter considers Bayes' theorem and Bayesian updating on the intervals at different magnitude thresholds and presents the results of the numerical analysis related to the inverse probability density function.

In Section 3.2, we derive Bayes' theorem for the inter-event time intervals at different magnitude thresholds in the marked point process. In Section 3.3, the inverse probability density function is derived for the stationary marked Poisson process that corresponds to the background seismicity of the ETAS model ($\lambda(t) \equiv \lambda_0$ in Eq. (1.3)). In Section 3.4, the Bayesian updating method is considered for the stationary marked Poisson process, and the inverse probability density and its approximation functions are derived. These functions are calculated numerically and compared in Section 3.5. In Section 3.6, 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, Section 3.7 presents additional discussions and conclusions. As so many variables and constants appear in this chapter, they are listed in Table 3.1, including those already described in previous Chapters.

Table 3.1: Variables and constants used in Chapter 3.

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.
$p_m(\tau_m)$, $p_M(\tau_M)$	Inter-event time distributions of τ_m, τ_M.

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 ETAS model to determine history-dependence.
M_0	The minimum magnitude in ETAS time series.
$p(M)$	Probability density of magnitude being M.
b	b-value of the GR 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.
$\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$	n consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$.
$N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$	Number of combinations of an upper interval of length τ_M and $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$.
$p_{Mm}(\tau_M\|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$	Inverse probability denisity function of an upper interval length given $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ in it.
$\mathcal{T}$	Sum of the lengths of consecutive lower intervals.
$N'_{mM}(\tau_m^{(1)}, \cdots, \tau_m^{(n)}\|\tau_M)$	Number of $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ in the new time series in Fig. 3.4(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.
$P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}\|\tau_M)$	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.
$p_{Mm}(\tau_M\|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$	Inverse probability density function of an upper interval length given $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$.
$p_{Mm}^{\mathrm{approx}}(\tau_M\|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$	Approximation function of $p_{Mm}(\tau_M\|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$.
$p_{Mm}^{\mathrm{kernel}}(\tau_M\|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$	Kernel part of the approximation function $p_{Mm}^{\mathrm{approx}}(\tau_M\|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$.
$p_{Mm}^{\mathrm{correct}}(\tau_M\|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$	Correction term in the approximation function $p_{Mm}^{\mathrm{approx}}(\tau_M\|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$.
$\tau_{m,j}$, $\tau_{M,k}$	Discretized intervals defined by Eq. (3.41).
$j_{\min}$, $j_{\max}$ ($j_{\min}^{(k)}$, $j_{\max}^{(k)}$)	Minimum and maximum of j in Eq. (3.41) (with explicit dependence on k).
$k_{\min}$, $k_{\max}$	Minimum and maximum of k in Eq. (3.41).
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. (3.42).

$\boldsymbol{p}_{mM}$, $\boldsymbol{P}_1$	Numerically calculated conditional probability density function and $P_1(\tau_m\|\tau_M)$ defined by Eq. (3.42).
$\mathcal{N}$	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. (3.45).
$D'(\cdot\|\|\cdot)$	Distance calculated by Eq. (3.46) with Eqs. (3.39) and (3.40).
$D''(\cdot\|\|\cdot)$	Distance calculated by Eq. (3.46) with Eqs. (3.36) and (3.37).
T	Elapsed time from the event with magnitude above M.
$\hat{\tau}_M^{\max}$	Maximum peak time in Eq. (3.24) which is discretized in Eq. (3.41).
$\tau_M^{\max,\mathrm{approx}}$	Maximum peak time in Eq. (3.36).
$k^{\max}$, $k^{\max,\mathrm{approx}}$	The k corresponds to $\hat{\tau}_M^{\max}$ and $\tau_M^{\max,\mathrm{approx}}$ by Eq. (3.41).
$\tau_M^{\max}$ ($\tau_M^{\max,\mathrm{n}}$)	Maximum peak time of the kernel part in Eq. (3.37) (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. (3.41) (at the n-th update).
$\tau_M^{\max,\mathrm{L}}$, $\tau_M^{\max,\mathrm{R}}$	Maximum peak time in Eq. (3.36) 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. (3.41).
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. (3.48).
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. (3.51).
$\Delta \log_{10} R_n$, $\Delta k_n^{\max}$	Variations of $\log_{10} R_n$ and $k^{\max,\mathrm{n}}$ defined by Eqs. (3.52) and (3.53).
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.

3.2 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. 3.1). Hereafter, we express this $N_{mM}(\tau_M, \tau_m)$ as the number of the pairs of the intervals such that the length of the upper interval *is* τ_M and the length of the lower interval *is* τ_m, for simplicity, and other numbers of the intervals are expressed in the same way. $N_{mM}(\tau_M, \tau_m)$ can be represented in two ways:

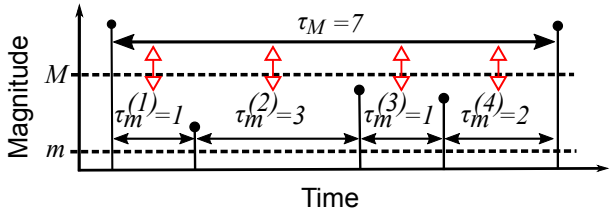

Figure 3.1: Schematic of the approach to count the number of pairs of upper and lower intervals whose lengths are τ_M and τ_m, respectively. Four pairs are shown in the figure, and $N_{mM}(7,1) = 2$, $N_{mM}(7,2) = 1$, and $N_{mM}(7,3) = 1$.

1) Derive $N_{mM}(\tau_M, \tau_m)$ by counting the cumulative total number of the upper intervals of length τ_M that include the lower interval of length τ_m (Fig. 3.2(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:

$$N_{mM}(\tau_M, \tau_m) = N_m p_m(\tau_m) p_{Mm}(\tau_M|\tau_m) d\tau_m d\tau_M. \tag{3.1}$$

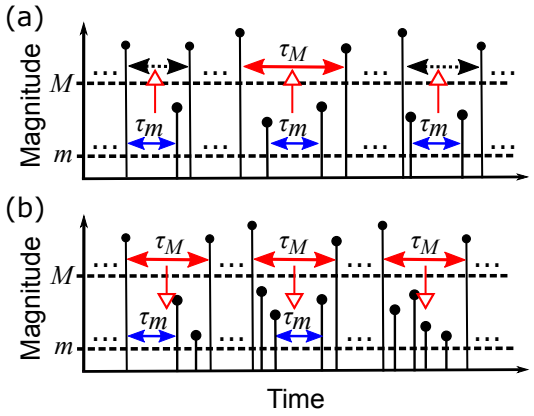

Figure 3.2: Schematic of the two approaches for calculating $N_{mM}(\tau_M, \tau_m)$. (a) The first approach involves counting the cumulative total number of the upper intervals of length τ_M that include the lower interval of length τ_m. (b) The second approach involves counting the number of the lower intervals of length τ_m included in the upper interval of length τ_M.

2) Derive $N_{mM}(\tau_M, \tau_m)$ by counting the total number of the lower intervals of length τ_m included in the upper interval of length τ_M (Fig. 3.2(b)). The number of the upper intervals of length τ_M in the time series is $N_M p_M(\tau_M)d\tau_M$. Therefore, the number of the lower intervals included in these upper intervals is:

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

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

$$N_{mM}(\tau_M, \tau_m) = N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M) d\tau_m d\tau_M. \tag{3.2}$$

From Eqs. (3.1) and (3.2):

$$N_m p_m(\tau_m) p_{Mm}(\tau_M|\tau_m) = N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M). \tag{3.3}$$

By using $N_M/N_m = \langle\tau_m\rangle/\langle\tau_M\rangle$, Eq. (3.3) is rewritten as:

$$p_{Mm}(\tau_M|\tau_m) = \left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \right) \frac{p_{mM}(\tau_m|\tau_M) p_M(\tau_M)}{p_m(\tau_m)}. \tag{3.4}$$

Equation (3.4) 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 [94]:

$$\begin{aligned} z_m(\tau_m) &:= \frac{\tau_m}{\langle\tau_m\rangle} p_m(\tau_m), \\ z_M(\tau_M) &:= \frac{\tau_M}{\langle\tau_M\rangle} p_M(\tau_M), \\ z_{mM}(\tau_m|\tau_M) &:= \frac{\tau_m}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M). \end{aligned} \tag{3.5}$$

These functions satisfy the normalization condition of the probability density function. Equations (2.3) and (3.4) are simplified as [94]:

$$z_m(\tau_m) = \int_0^\infty z_{mM}(\tau_m|\tau_M) z_M(\tau_M) d\tau_M, \tag{3.6}$$

$$p_{Mm}(\tau_M|\tau_m) = \frac{z_{mM}(\tau_m|\tau_M) z_M(\tau_M)}{z_m(\tau_m)}. \tag{3.7}$$

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

3.3 Bayes' Theorem for Stationary Marked Poisson Process

In this section, we derive $p_{Mm}(\tau_M|\tau_m)$ for a stationary marked Poisson process generated by the ETAS model with $\lambda(t) \equiv \lambda_0$ in Eq. (1.3). In this case, the magnitudes and inter-event times obey the following probability density functions independently.

$$p(M) \propto 10^{-bM}, \tag{3.8}$$

$$p_M(\tau_M) = \frac{1}{\langle \tau_M \rangle} e^{-\frac{\tau_M}{\langle \tau_M \rangle}}. \tag{3.9}$$

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

$$p_{mM}(\tau_m|\tau_M) = \frac{\sum_{i=1}^\infty i\ \rho_{mM}(\tau_m|i,\tau_M)\Psi_{mM}(i|\tau_M)}{\sum_{i=1}^\infty i\ \Psi_{mM}(i|\tau_M)}, \tag{3.10}$$

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 stationary Poisson process, 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:

$$\Psi_{mM}(i|\tau_M) = \frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right)^{i-1}}{(i-1)!} e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}, \tag{3.11}$$

where:

$$\begin{aligned} A_{\Delta m} &:= \frac{\langle\tau_M\rangle}{\langle\tau_m\rangle} - 1 \\ &= 10^{b\Delta m} - 1. \end{aligned} \tag{3.12}$$

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

$$\rho_{mM}(\tau_m|1,\tau_M) = \delta(\tau_M - \tau_m). \tag{3.13}$$

For $i \geq 2$ [95, 96],

$$\rho_{mM}(\tau_m|i,\tau_M) = \frac{(i-1)}{\tau_M}\left(1 - \frac{\tau_m}{\tau_M}\right)^{i-2}\theta(\tau_M - \tau_m). \tag{3.14}$$

From Eqs. (3.11) – (3.14), $p_{mM}(\tau_m|\tau_M)$ is derived as (Appendix B):

$$p_{mM}(\tau_m|\tau_M) = \frac{e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m) + \frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\left(A_{\Delta m}\frac{\tau_M-\tau_m}{\langle\tau_M\rangle}+2\right)\theta(\tau_M-\tau_m)}{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}+1\right)}. \tag{3.15}$$

This conditional probability composed of Eqs. (3.11) – (3.14) certainly has exponential distributions as the solution of Eq. (2.3) (Appendix B).

Second, we derive $p_{Mm}(\tau_M|\tau_m)$. From Eqs. (3.9) and (3.15), $p_{Mm}(\tau_M|\tau_m)$ is obtained as (Appendix B):

$$p_{Mm}(\tau_M|\tau_m) = \frac{e^{-\frac{\tau_M-\tau_m}{\langle\tau_m\rangle}}\delta(\tau_M-\tau_m) + \frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-\frac{\tau_M-\tau_m}{\langle\tau_M\rangle}}\left(A_{\Delta m}\frac{\tau_M-\tau_m}{\langle\tau_M\rangle}+2\right)\theta(\tau_M-\tau_m)}{\left(A_{\Delta m}+1\right)^2}. \tag{3.16}$$

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

$$\tau_M^{\max} = \tau_m + \langle\tau_M\rangle\left(1-\frac{2}{A_{\Delta m}}\right), \tag{3.17}$$

when the next condition is satisfied (Appendix B).

$$\Delta m > \frac{\log_{10}3}{b}. \tag{3.18}$$

3.4 Bayesian Updating for Stationary Marked Poisson Process

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

3.4.1 Inverse Probability Density Function

As in Section 3.2, we derive the inverse probability density function by expressing the total number of combinations of the upper interval of length τ_M and the consecutive lower intervals of lengths $\{\tau_m^{(1)},\cdots,\tau_m^{(n)}\}$ included in it denoted by $N_{mM}(\tau_M,\tau_m^{(1)},\cdots,\tau_m^{(n)})$ in two ways.

First, we derive $N_{mM}(\tau_M,\tau_m^{(1)},\cdots,\tau_m^{(n)})$ by counting the cumulative total number of the upper intervals of length τ_M that include the consecutive lower intervals of lengths $\{\tau_m^{(1)},\cdots,\tau_m^{(n)}\}$ (Fig. 3.3(a)). We begin with the case $n=2$. The intervals in the stationary marked Poisson process 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. 3.3(a)). In that case, the magnitude of the event sandwiched between the two lower intervals is larger than M. In the stationary marked Poisson process, the proportion that the consecutive lower intervals belong to the same upper interval equals to the probability that the magnitude of the event sandwiched between the two lower intervals is smaller than M. It is given by the GR law as:

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

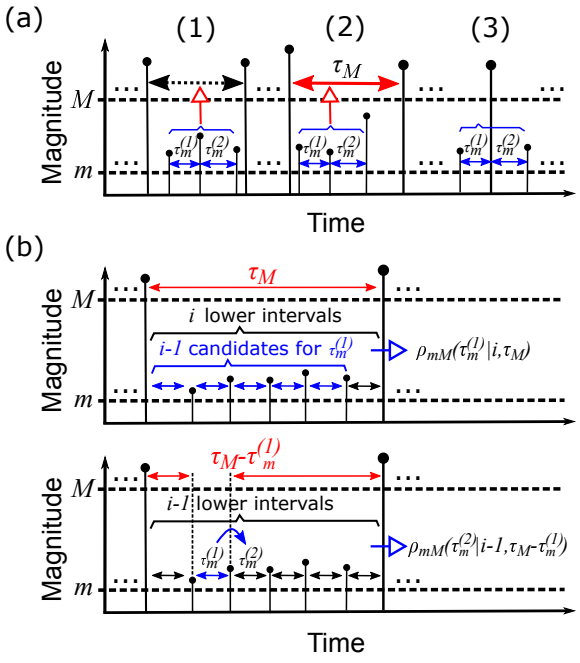

Figure 3.3: Schematic of the two approaches to calculate $N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)})$. (a) The first approach involves counting the cumulative total number of the upper intervals of length τ_M that include the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \tau_m^{(2)}\}$. (b) 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.

Therefore,

$$N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)}) = N_m \left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right) p_m(\tau_m^{(1)}) p_m(\tau_m^{(2)}) p_{Mm}(\tau_M | \tau_m^{(1)}, \tau_m^{(2)}) d\tau_m^2 d\tau_M. \tag{3.19}$$

Equation (3.19) can be generalized for $n(\geq 2)$ consecutive lower intervals.

$$N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) = N_m \left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1} \left(\prod_{i=1}^{n} p_m(\tau_m^{(i)})\right) p_{Mm}(\tau_M | \tau_m^{(1)}, \cdots, \tau_m^{(n)}) d\tau_m^n d\tau_M. \tag{3.20}$$

Second, we derive $N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)})$ by counting the total number of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ included in the upper interval of length τ_M (Fig. 3.3(b)). To this end, we start with the case $n = 2$ again (Fig. 3.3(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):

$$N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)}) = N_M p_M(\tau_M) d\tau_M \sum_{i=2}^{\infty} (i-1) \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \rho_{mM}(\tau_m^{(2)}|i-1, \tau_M - \tau_m^{(1)}) d\tau_m^2. \tag{3.21}$$

Equation (3.21) is generalized for the case $n(\geq 2)$ lower intervals as:

$$N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) = N_M p_M(\tau_M) d\tau_M \times \sum_{i=n}^{\infty} (i-n+1) \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{n} \rho_{mM}\left(\tau_m^{(j)} | i-j+1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)}\right) d\tau_m^n. \tag{3.22}$$

From Eqs. (3.20) and (3.22), $p_{Mm}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)})$ is derived as:

$$p_{Mm}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\frac{1}{\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}}\frac{p_M(\tau_M)}{\prod_{i=1}^n p_m(\tau_m^{(i)})}$$
$$\times\sum_{i=n}^{\infty}(i-n+1)\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(1)}|i,\tau_M)\prod_{j=2}^{n}\rho_{mM}\left(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right). \quad (3.23)$$

Furthermore, the explicit form of the inverse probability density function is derived by substituting Eqs. (3.9) and (3.11) – (3.14) into Eq. (3.23) as (Appendix C):

$$p_{Mm}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)}) = \left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^2\left\{e^{-\frac{\tau_M-\sum_{i=1}^n\tau_m^{(i)}}{\langle\tau_m\rangle}}\delta\left(\tau_M-\sum_{i=1}^n\tau_m^{(i)}\right)\right.$$
$$\left.+\frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-\frac{\tau_M-\sum_{i=1}^n\tau_m^{(i)}}{\langle\tau_M\rangle}}\left[\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left(\tau_M-\sum_{i=1}^n\tau_m^{(i)}\right)+2\right]\theta\left(\tau_M-\sum_{i=1}^n\tau_m^{(i)}\right)\right\}. \quad (3.24)$$

Equation (3.24) includes the case $n=1$ (Eq. (3.16)). In addition, Eq. (3.24) is identical to Eq. (3.16) when τ_m is replaced with $\mathcal{T}:=\sum_{i=1}^n\tau_m^{(i)}$; this implies that the occurrence pattern of small events does not affect that of upper intervals. This seems natural for the stationary Poisson process.

The same property as Eqs. (3.17) and (3.18) holds for $p_{Mm}(\tau_M|\tau_m^{(1)},\cdots,\tau_m^{(n)})$; it has a peak at:

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

under the condition

$$\Delta m > \frac{\log_{10}3}{b}. \quad (3.25)$$

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. (3.24) (Appendix D). 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 [38]. In that case, the lower intervals should be considered not from the leftmost one but from somewhere else. Therefore, in the present chapter, 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.

3.4.2 Approximation Function of Inverse Probability Density Function

Equation (3.23) indicates that new information on the lower intervals cannot be added by the product of the conditional probabilities as is usual in Bayesian updating. In this subsection, we derive its approximation function with a convenient form applicable to the time series with correlations between events.

To this end, we use the approximate derivation of $N_{mM}(\tau_M,\tau_m^{(1)},\cdots,\tau_m^{(n)})$ described below instead of the second approach for deriving Eq. (3.22). In the following, the upper and the lower consecutive intervals are assumed to satisfy:

$$\tau_M \geq \sum_{i=1}^n\tau_m^{(i)}. \quad (3.26)$$

First, consider the case $n=2$. There are $N_M p_M(\tau_M)d\tau_M$ upper intervals of length τ_M in the time series. These upper intervals are as shown in Fig. 3.4(a), and we use them to generate a new time series by connecting them in the order of appearance as in Fig. 3.4(b). Let the number of the consecutive

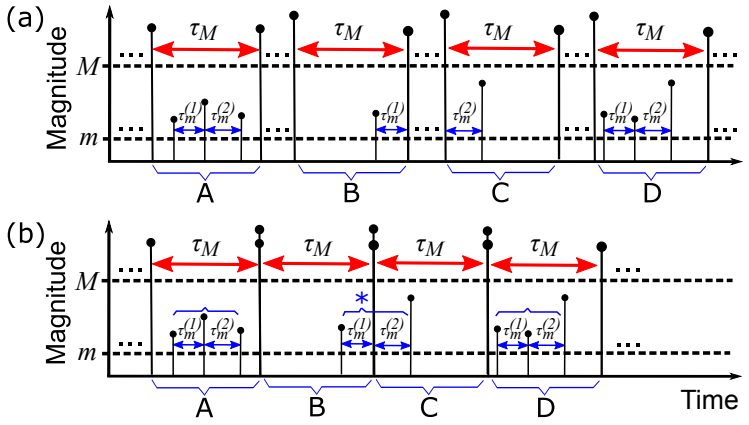

Figure 3.4: Schematic of another approach to count the total number of consecutive lower intervals of lengths $\tau_m^{(1)}$ and $\tau_m^{(2)}$ included in the upper interval of length τ_M. (a) First, pick up all upper intervals of length τ_M from the time series. (b) Second, generate new time series 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 $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$.

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:

$$N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M) \approx N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle}_{\tau_M} p_{mM}(\tau_m^{(1)}|\tau_M) p_{mM}(\tau_m^{(2)}|\tau_M) d\tau_m^2 d\tau_M. \tag{3.27}$$

$N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$ is not equivalent to $N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)})$ because $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$ includes cases where the two consecutive lower intervals do not belong to the same upper interval (the case indicated by $*$ in Fig. 3.4(b)). Therefore, it is necessary to count such cases in the time series, and subtract them from $N'_{mM}(\tau_m^{(1)}, \tau_m^{(2)}|\tau_M)$.

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

$$\begin{aligned} P^{\mathrm{R}}(\tau_m|\tau_M) &= P^{\mathrm{L}}(\tau_m|\tau_M) \\ &= \sum_{i=1}^{\infty} \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m|i, \tau_M). \end{aligned} \tag{3.28}$$

Here, the probability density for the rightmost lower interval is denoted by $P^{\mathrm{R}}(\tau_m|\tau_M)$, and the leftmost by $P^{\mathrm{L}}(\tau_m|\tau_M)$. Equation (3.28) can be explicitly written using Eqs. (3.11) – (3.14) as (Appendix E):

$$\begin{aligned} P^{\mathrm{R}}(\tau_m|\tau_M) &= P^{\mathrm{L}}(\tau_m|\tau_M) \\ &= e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}} \delta(\tau_M - \tau_m) + \frac{A_{\Delta m}}{\langle\tau_M\rangle} e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}} \theta(\tau_M - \tau_m). \end{aligned} \tag{3.29}$$

By using $P^{\mathrm{L}}(\tau_m|\tau_M)$ and $P^{\mathrm{R}}(\tau_m|\tau_M)$, the number of cases to exclude can be expressed for a sufficiently large N_M (because $N_M p_M(\tau_M)d\tau_M$ in Eq. (3.30) is precisely $N_M p_M(\tau_M)d\tau_M - 1$) as:

$$N_M p_M(\tau_M) P^{\mathrm{R}}(\tau_m^{(1)}|\tau_M) P^{\mathrm{L}}(\tau_m^{(2)}|\tau_M) d\tau_m^2 d\tau_M. \tag{3.30}$$

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

$$\begin{aligned} &N_{mM}(\tau_M, \tau_m^{(1)}, \tau_m^{(2)}) \\ &\quad \approx N_M p_M(\tau_M) \left(\frac{\tau_M}{\langle\langle \tau_m \rangle\rangle}_{\tau_M} p_{mM}(\tau_m^{(1)}|\tau_M) p_{mM}(\tau_m^{(2)}|\tau_M) - P^{\mathrm{R}}(\tau_m^{(1)}|\tau_M) P^{\mathrm{L}}(\tau_m^{(2)}|\tau_M) \right) d\tau_m^2 d\tau_M. \end{aligned} \tag{3.31}$$

Next, we consider the case $n(\geq 3)$. Equation (3.27) is generalized as:

$$N'_{mM}(\tau_m^{(1)}, \cdots, \tau_m^{(n)}|\tau_M) \approx N_M p_M(\tau_M) \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle}_{\tau_M} \left(\prod_{i=1}^{n} p_{mM}(\tau_m^{(i)}|\tau_M) \right) d\tau_m^n d\tau_M. \tag{3.32}$$

From this $N'_{mM}(\tau_m^{(1)}, \cdots, \tau_m^{(n)}|\tau_M)$, the cases wherein the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ are not included in the same upper interval need to be excluded. Considering the condition of Eq. (3.26), a sequence of consecutive lower intervals is divided by only one boundary event with a magnitude above M (Fig. 3.5). Let $P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M)$ be the probability that the rightmost or leftmost lower intervals of the upper interval of length τ_M is $\{\tau_m^{(1)}, \cdots, \tau_m^{(l)}\}$ $(l \geq 2)$. Then, as the position of the rightmost or leftmost lower intervals is confirmed among the $i(\geq l)$ lower intervals, $P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M)$ is:

$$P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M) = \sum_{i=l}^{\infty} \Psi_{mM}(i|\tau_M) \rho_{mM}(\tau_m^{(1)}|i, \tau_M) \prod_{j=2}^{l} \rho_{mM} \left(\tau_m^{(j)}|i-j+1, \tau_M - \sum_{k=1}^{j-1} \tau_m^{(k)} \right). \tag{3.33}$$

By substituting Eqs. (3.11) – (3.14) into Eq. (3.33) (Appendix F):

$$P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M) = \prod_{i=1}^{l} P_i(\tau_m^{(i)}|\tau_M), \text{ where } P_i(\tau_m^{(i)}|\tau_M) = \left(\frac{A_{\Delta m}}{\langle \tau_M \rangle} \right) e^{-A_{\Delta m} \frac{\tau_m^{(i)}}{\langle \tau_M \rangle}}. \tag{3.34}$$

There are $(n-1)$ possible choices for the boundary position of the consecutive lower intervals (Fig. 3.5(a)), each with an equal probability $\prod_{i=1}^{n} P_i$. The number of consecutive upper intervals in the new time series is almost $N_M p_M(\tau_M)d\tau_M$, and therefore, the number of cases to be excluded is:

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

Then,

$$N_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) \approx N_M p_M(\tau_M) \left[\frac{\tau_M}{\langle\langle \tau_m \rangle\rangle}_{\tau_M} \prod_{i=1}^{n} p_{mM}(\tau_m^{(i)}|\tau_M) - (n-1) \prod_{i=1}^{n} P_i(\tau_m^{(i)}|\tau_M) \right] d\tau_m^n d\tau_M. \tag{3.35}$$

Therefore, from Eqs. (3.20) and (3.35), the approximation function $(p_{Mm}^{\mathrm{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}))$ of the inverse probability density function is derived as:

$$\begin{aligned} p_{Mm}^{\mathrm{approx}}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = &\frac{\langle \tau_m \rangle}{\langle \tau_M \rangle} \frac{1}{\left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1}} \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle}_{\tau_M} \left(\prod_{i=1}^{n} \frac{p_{mM}(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})} \right) p_M(\tau_M) \\ &- \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle} \frac{(n-1)}{\left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1}} \left(\prod_{i=1}^{n} \frac{P_i(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})} \right) p_M(\tau_M). \end{aligned} \tag{3.36}$$

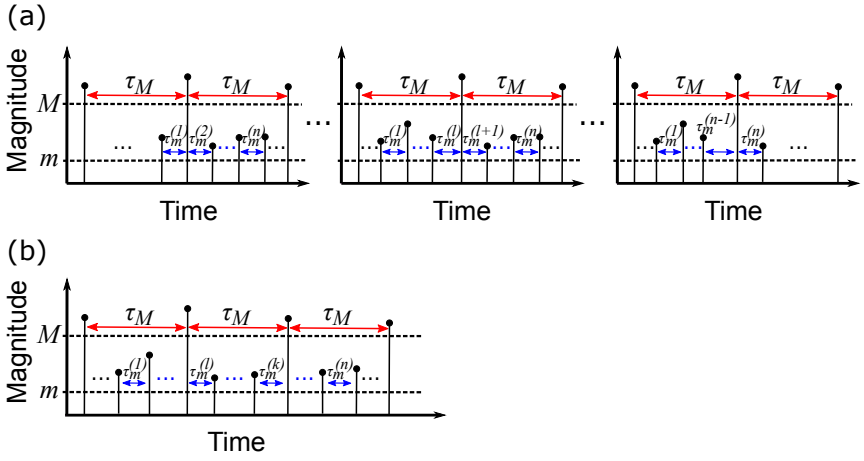

Figure 3.5: Schematic of the patterns of the consecutive lower intervals of lengths $\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ excluded from $N'_{mM}(\tau_m^{(1)}, \cdots, \tau_m^{(n)}|\tau_M)$. (a) There are $(n-1)$ ways to divide the sequence of lower intervals by the event with a magnitude greater than M at the boundary of the upper intervals of length τ_M. (b) The sequence can not be divided by more than one boundary according to condition (3.26).

Equation (3.36) is composed of two parts: the first term on the r.h.s. involves the product of the conditional probability density functions, and we refer to this part as the kernel part of the approximation function ($p_{Mm}^{\rm kernel}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$) hereafter.

$$p_{Mm}^{\rm kernel}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle} \frac{1}{\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^{n-1}} \frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} \left(\prod_{i=1}^{n} \frac{p_{mM}(\tau_m^{(i)}|\tau_M)}{p_m(\tau_m^{(i)})}\right) p_M(\tau_M). \tag{3.37}$$

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

$$\text{correction term} = (n-1) p_{Mm}^{\rm correct}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}), \tag{3.38}$$

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

Equation (3.36) can be explicitly written as (Appendix G):

$$\begin{aligned} p_{Mm}^{\rm approx}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) = {} & \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2}\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}+1\right) e^{-\frac{\tau_M-\sum_{i=1}^n \tau_m^{(i)}}{\langle\tau_M\rangle}} \\ & \times \prod_{i=1}^{n}\left[1-\left(\frac{\tau_m^{(i)}-\frac{\langle\tau_M\rangle}{A_{\Delta m}}}{\tau_M+\frac{\langle\tau_M\rangle}{A_{\Delta m}}}\right)\right] - \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2}\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)(n-1) e^{-\frac{\tau_M-\sum_{i=1}^n \tau_m^{(i)}}{\langle\tau_M\rangle}}. \end{aligned} \tag{3.39}$$

The kernel part is explicitly expressed as:

$$\begin{aligned} & p_{Mm}^{\rm kernel}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\ & = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle^2}\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}+1\right) e^{-\frac{\tau_M-\sum_{i=1}^n \tau_m^{(i)}}{\langle\tau_M\rangle}} \prod_{i=1}^{n}\left[1-\left(\frac{\tau_m^{(i)}-\frac{\langle\tau_M\rangle}{A_{\Delta m}}}{\tau_M+\frac{\langle\tau_M\rangle}{A_{\Delta m}}}\right)\right]. \end{aligned} \tag{3.40}$$

Note that functions (3.36) – (3.40) do not satisfy the normalization condition. Furthermore, in some cases, $p_{Mm}^{\rm approx}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$ in Eqs. (3.36) and (3.39) may take negative values when the correction term is larger than the kernel part. The relationship between the inverse probability density function and its approximation function is discussed in Appendix H.

3.5 Examination of Bayesian Updating Method in Stationary Marked Poisson Process

In this section, we compute the inverse probability density function given by Eq. (3.24) and the (part of) approximation function (Eqs. (3.36) – (3.40)) for the numerically generated stationary marked Poisson process, and we compare their properties. We examine the numerical method of Bayesian updating by changing some conditions to see its utility.

3.5.1 Time Series Generation and Bayesian Updating Methods

The stationary marked Poisson process can be numerically generated by setting $\lambda(t) \equiv \lambda_0$ in Eq. (1.3). In fact, it was numerically generated as the renewal process in which magnitudes and time intervals were generated randomly obeying Eqs. (3.8) and (3.9), 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. (3.25) 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 $\{\tau_m^{(1)}, \tau_m^{(2)}, \cdots, \tau_m^{(n)}\}$ into Eqs. (3.24), (3.39), and (3.40). The summation of the lower intervals at the n-th update $\sum_{i=1}^{n} \tau_m^{(i)}$ is equivalent to the elapsed time T from the previous event with a magnitude above M. Further, the updating 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. (3.36) and (3.37). First, we generated $\mathcal{N}$ time series each contains 10^5 events as sample data. From these sample data, we numerically obtained the statistics required for calculating Eqs. (3.36) and (3.37), i.e., $p_m(\tau_m)$, $p_M(\tau_M)$, $p_{mM}(\tau_m|\tau_M)$ and $P_i(\tau_m|\tau_M)$, and the average number of lower intervals 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:

$$\begin{aligned} \tau_{m,j} &:= 10^{(j+0.5)\Delta\tau_m}, \\ \tau_{M,k} &:= 10^{(k+0.5)\Delta\tau_M}, \end{aligned} \tag{3.41}$$

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

$$\begin{aligned} \boldsymbol{p}_m &= [p_{m,j}]_{j=j_{\min},\cdots,j_{\max}}, \\ \boldsymbol{p}_M &= [p_{M,k}]_{k=k_{\min},\cdots,k_{\max}}, \\ \boldsymbol{p}_{mM} &= \left[[p_{mM,jk}]_{j=j_{\min}^{(k)},\cdots,j_{\max}^{(k)}}\right]_{k=k_{\min},\cdots,k_{\max}}, \\ \boldsymbol{P}_1 &= \left[[P_{1,jk}]_{j=j_{\min}^{(k)},\cdots,j_{\max}^{(k)}}\right]_{k=k_{\min},\cdots,k_{\max}}. \end{aligned} \tag{3.42}$$

In Eq. (3.42), $j_{\min}, j_{\max}, k_{\min}$, and $k_{\max}$ represent the smallest and largest bin numbers of each distribution. For the statistics obtained as a matrix, the range of j depends on k, and this is indicated as $j_{\min}^{(k)}$ and $j_{\max}^{(k)}$. The ranges of j and k are different for distribution; however, the same symbol is used in Eq. (3.42). In this chapter, 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. K.1). 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. K.2 and K.3). Meanwhile, for $\tau_{M,k}$, the domain was extended using the average of the functions at $\{\tau_{M,k_{\min}}, \cdots, \tau_{M,k_{\min}+l_e-1}\}$ as the substitute for $\tau_{M,k}$ with $k < k_{\min}$, whereas using the functions at $\{\tau_{M,k_{\max}-l_e+1}, \cdots, \tau_{M,k_{\max}}\}$ as the substitute for $\tau_{M,k}$ with $k > k_{\max}$. We set $l_e = 5$ for $\Delta\tau_M = 0.1$ and $l_e = 20$ for $\Delta\tau_M = 0.025$.

Finally, for $\tau_{M,k}/\langle\langle\tau_m\rangle\rangle_{\tau_{M,k}}$, the interpolation and extrapolation procedures were conducted in the same way as $\boldsymbol{p}_M$, although the extrapolation functions were different (Fig. K.4).

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. (3.41). We set the range of k to be $-120 \leq k \leq 70$ for $\Delta\tau_M = 0.1$, and $-480 \leq k \leq 280$ for $\Delta\tau_M = 0.025$. Although this yielded the maximum range of Bayesian updating, the updating at the n-th step was performed within the range $\max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\} < \tau_M$. The properties of the inverse probability density function and the (part of) approximation function were examined within this range.

The kernel parts of the approximation functions were computed by calculating Eq. (3.37) in a step-by-step manner as:

$$\begin{aligned}
\ln p_{Mm}^{\text{kernel}}(\tau_{M,k}|\tau_m^{(1)}) &= \ln\left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\frac{\tau_{M,k}}{\langle\langle\tau_m\rangle\rangle_{\tau_{M,k}}}\right) + \ln p_{mM}(\tau_m^{(1)}|\tau_{M,k}) - \ln p_m(\tau_m^{(1)}) + \ln p_{M,k},\\
\ln p_{Mm}^{\text{kernel}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)}) &= -\ln\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln p_{mM}(\tau_m^{(2)}|\tau_{M,k}) - \ln p_m(\tau_m^{(2)}) + \ln p_{Mm}^{\text{kernel}}(\tau_{M,k}|\tau_m^{(1)}),\\
\ln p_{Mm}^{\text{kernel}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)},\tau_m^{(3)}) &= -\ln\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln p_{mM}(\tau_m^{(3)}|\tau_{M,k}) - \ln p_m(\tau_m^{(3)}) + \ln p_{Mm}^{\text{kernel}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)}),\\
&\vdots
\end{aligned} \tag{3.43}$$

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

$$\begin{aligned}
\ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)}) &= \ln\left(\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(1)}|\tau_{M,k}) - \ln p_m(\tau_m^{(1)}) + \ln p_{M,k},\\
\ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)}) &= -\ln\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(2)}|\tau_{M,k}) - \ln p_m(\tau_m^{(2)}) + \ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)}),\\
\ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)},\tau_m^{(3)}) &= -\ln\left(1-\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right) + \ln P_1(\tau_m^{(3)}|\tau_{M,k}) - \ln p_m(\tau_m^{(3)}) + \ln p_{Mm}^{\text{correct}}(\tau_{M,k}|\tau_m^{(1)},\tau_m^{(2)}),\\
&\vdots
\end{aligned} \tag{3.44}$$

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

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 $p_{\text{sup}} > \ln p_{Mm}^{\text{kernel}}, \ln p_{Mm}^{\text{correct}} > p_{\text{inf}}$. Here, $p_{\text{sup}}(= 600)$ and $p_{\text{inf}}(= -600)$ yielded the upper and lower limits of p_{Mm}^{kernel} and p_{Mm}^{correct} 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. (3.36) became negative were excluded.

Figure K.5 shows an example of Bayesian updating for the stationary marked Poisson process. The inverse probability density function given by Eq. (3.24) has a characteristic peak that is not observed in $p_M(\tau_M)$. The correction term makes the kernel part obtained from Eq. (3.40) closer to the inverse probability density function. Moreover, the numerical calculations based on Eqs. (3.36) and (3.37) with $\mathcal{N} = 10^3$ and $\Delta\tau_M = 0.1$ appear to be consistent with these results.

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

3.5.2 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.

Comparison by Distance

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

$$D(f||g) := \int_T^\infty |f(\tau_M) - g(\tau_M)|^2 \, d\tau_M. \tag{3.45}$$

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

$$D(f||g) \simeq \sum_{\substack{k;\tau_{M,k}>T \\ p_{\rm sup}>\ln f,\ln g>p_{\rm inf}}} |f(\tau_{M,k}) - g(\tau_{M,k})|^2 \, (\ln 10)\tau_{M,k}\Delta\tau_M. \tag{3.46}$$

$D(f||g)$ was calculated for each update throughout the 100 test data time series. If no k's satisfied $p_{\rm sup} > \ln f, \ln g > p_{\rm inf}$, 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 3.6(a) shows the average distance for the cases $f = p_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$, $p_{Mm}^{\rm approx}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$, $p_{Mm}^{\rm kernel}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$, and $g = p_M(\tau_M)$. In addition to the analytical calculation in Eq. (3.45) for $D(p_{Mm}||p_M)$, the results of the numerical integration of Eq. (3.46) are presented; the calculations using Eqs. (3.39) and (3.40) are indicated by $D'(\cdot||\cdot)$. The results of the calculation using Eqs. (3.36) and (3.37) with the numerical method in Section 3.5.1 with $\mathcal{N} = 10^3$ and $\Delta\tau_M = 0.1$ are presented by $D''(\cdot||\cdot)$. The results for $\mathcal{N} = 10^5$ with $\Delta\tau_M = 0.1$ and $\Delta\tau_M = 0.025$ are shown in Fig. K.6.

First, one can see that $\langle D'(p_{Mm}||p_M)\rangle$ is almost consistent with $\langle D'(p_{Mm}^{\rm approx}||p_M)\rangle$, which indicates that $p_{Mm}^{\rm approx}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$ derived in the previous section certainly approximates the inverse probability density function, regardless of the elapsed time (or regardless of the number of updates, because the occurrence rate is constant). However, these separate from $D(p_{Mm}||p_M)$ at around $T \sim 10^5$ and at a large T. As such separations disappear when $\Delta\tau_M = 0.025$ (Figs. K.6(c) and K.6(d)), this is attributed to the coarseness of the numerical integration.

Second, $\langle D'(p_{Mm}^{\rm kernel}||p_M)\rangle$ is nearly consistent with $\langle D''(p_{Mm}^{\rm kernel}||p_M)\rangle$. This suggests that the numerical updating method in Eq. (3.43) certainly calculates the kernel part. However, $\langle D''(p_{Mm}^{\rm approx}||p_M)\rangle$ gradually separates from $\langle D'(p_{Mm}^{\rm approx}||p_M)\rangle$ at a large T. This separation is more clearly illustrated in Fig. 3.6(b), which shows the average distances between $f = p_{Mm}^{\rm approx}, p_{Mm}^{\rm kernel}$ and $g = p_{Mm}$ calculated by Eq. (3.46). This separation can be attributed to the calculation of the correction term in Eq. (3.44), in particular to the fluctuation in the numerically obtained $\boldsymbol{P}_1$ (Appendix J).

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. (3.24) and its approximation function in Eq. (3.36) with the numerical updating method, each denoted by $\hat{\tau}_M^{\rm max}$ and $\tau_M^{\rm max,approx}$. Both functions are discretized as Eq. (3.41); the corresponding k in Eq. (3.41) is denoted by $\hat{k}^{\rm max}$ and $k^{\rm max,approx}$, respectively.

$\hat{k}^{\rm max}$ and $k^{\rm max,approx}$ 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 $\hat{k}^{\rm max}$ or $k^{\rm max,approx}$ was located at such edges,

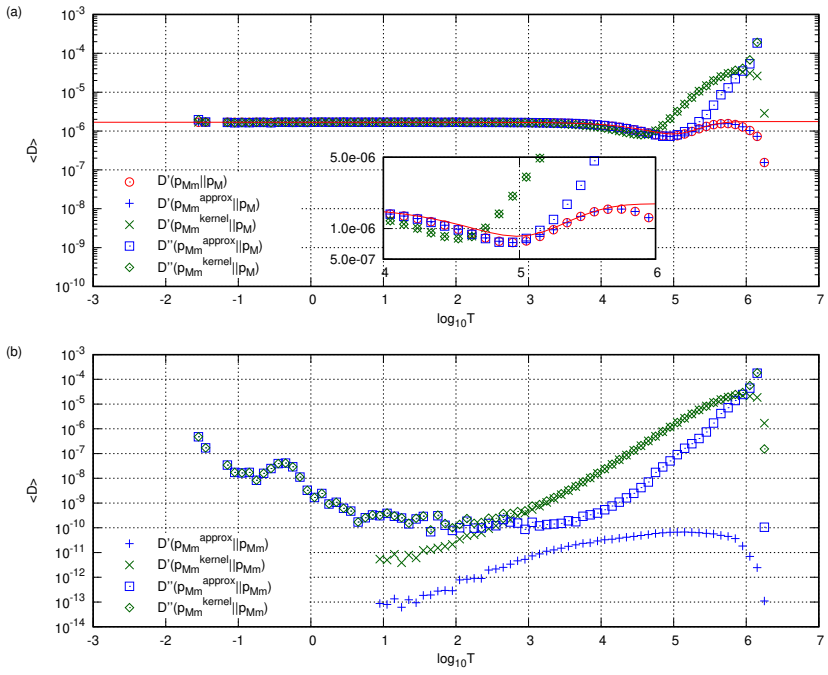

Figure 3.6: Average distances for each elapsed time (T) from the previous large event with a magnitude above M. (a) Distances between the inter-event time distribution and other function. $D(p_{Mm}||p_M)$ (Eq. (I.2) in Appendix I) is shown by the red curve, and the symbols are numerical results for Eq. (3.46). (b) Distances between the inverse probability density function and other function numerically calculated by Eq. (3.46).

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), $k^{\max,\text{approx}}$ was set to be 80 or 320.

Figure 3.7 shows the joint probability mass function (p.m.f.) of $(\hat{k}^{\max}, k^{\max,\text{approx}})$ for $\mathcal{N} = 10^3$ and $\Delta\tau_M = 0.1$. Those for $\mathcal{N} = 10^5$ are presented in Fig. K.7. Here, the population is all the pairs of $(\hat{k}^{\max}, k^{\max,\text{approx}})$ 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) $\tau_M > \max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$, and (b) $\tau_M > T$. In the former case, the p.m.f. was bimodal; the higher peak existed around $\hat{k}^{\max} = k^{\max,\text{approx}}$, 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. K.7). 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.

3.6 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. (3.36)) and its kernel part (Eq. (3.37)) using the numerical Bayesian updating method. The maximum peak time of

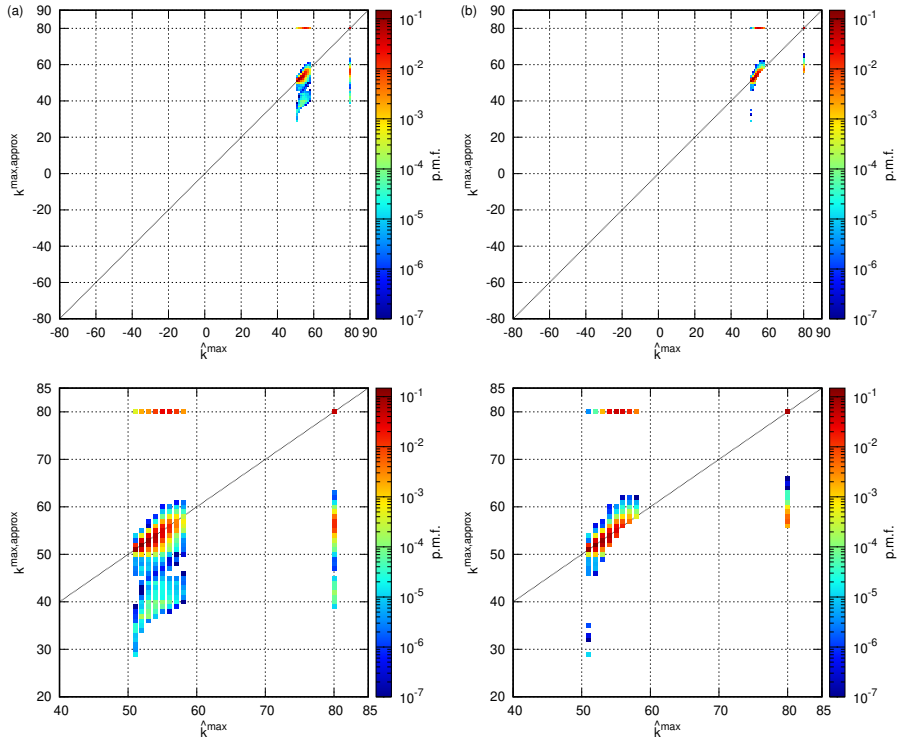

Figure 3.7: Joint probability mass function for $(\hat{k}^{\max}, k^{\max,\mathrm{approx}})$. Numerical search of the maximum peak is conducted for (a) $\tau_M > \max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ and (b) $\tau_M > T$. The horizontal line at $k^{\max,\mathrm{approx}} = 80$ and the vertical line at $\hat{k}^{\max} = 80$ correspond to the cases when the peak is not detected by the peak search. The lower panels show the enlarged versions of the upper panels.

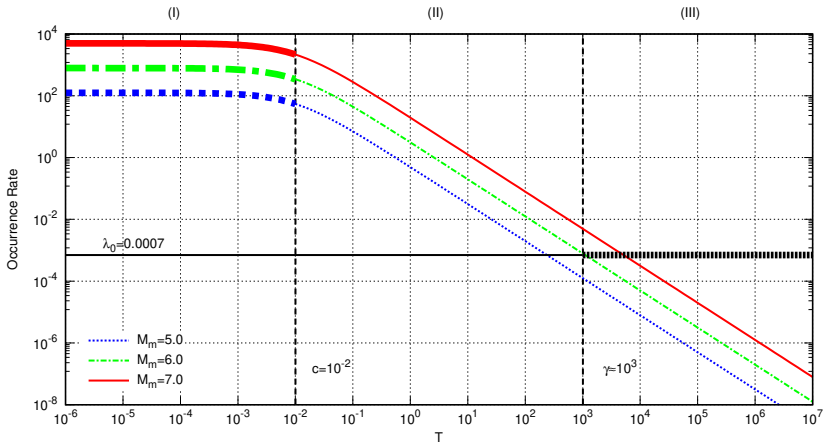

Figure 3.8: The OU law for the parameter values in the text with a different mainshock magnitude M_m (after Ref. [93]). The number of aftershocks per unit day against the elapsed time (T) from the mainshock obeys $\lambda(T) = K10^{\alpha(M_m - M_0)}/(T+c)^{\theta+1}$. The background rate ($\lambda_0 = 0.0007$) is also shown.

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.

3.6.1 Time Series Generation and Bayesian Updating Methods

We applied the numerical Bayesian updating method in Section 3.5.1 to the time series generated by Eq. (1.3) [52, 59] 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 [78] so that the branching ratio is less than 1 [58]. The magnitude thresholds were $(M, m) = (5, 3)$. As in Section 3.5.1, 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 in the sense that the branching ratio ($n_{\rm br} \approx 0.785$) is less than 1 (in the word of this thesis, it is a mixture time series), it is locally non-stationary obeying nearly the OU law after a large event, as shown in Fig. 3.8 (note that the local OU law in Eq. (1.3) is different from the global aftershock decay [58], and thus, the actual decaying must be slightly different from shown in Fig. 3.8). The activity can be categorized into three regimes with respect to the elapsed time (T) from the mainshock, as summarized in Table 3.2.

Table 3.2: 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 Section 3.5.1 (Figs. K.8 – K.12). Bayesian updating (Eqs. (3.43) and (3.44)) 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 $\max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\} < \tau_M$. 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.

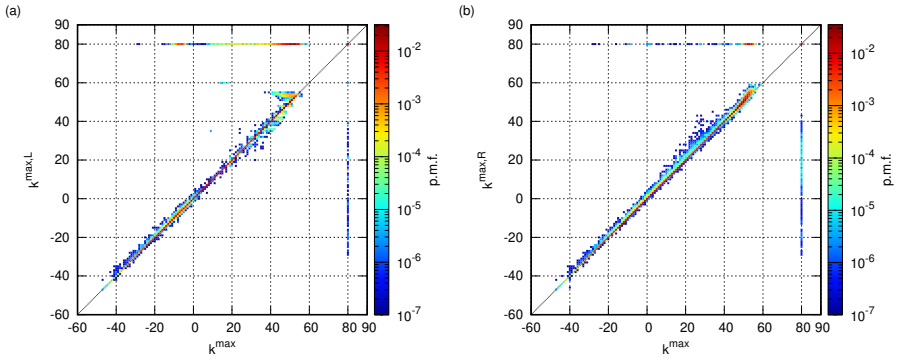

Figure 3.9: Joint probability mass functions for (a) $\left(k^{\max}, k^{\max,\mathrm{L}}\right)$ and (b) $\left(k^{\max}, k^{\max,\mathrm{R}}\right)$. The horizontal lines at $k^{\max,\mathrm{L}} = 80$ and $k^{\max,\mathrm{R}} = 80$ and the vertical line $k^{\max} = 80$ are the cases where the peak is not detected.

The following normalizations were performed in the calculations of Bayesian updating. The result of the calculation in Eq. (3.43) can be very large. In order to compute the approximation functions together with Eq. (3.44), it is necessary to use the function value of p^{kernel}_{Mm} as it is, though it can exceed p_{sup}. Therefore, to avoid such enlargement, we normalized the result of Eq. (3.43) for each update by subtracting the following numerical integration from Eq. (3.43).

$$\ln\left(\sum_{\substack{k;\tau_{M,k}>T \\ p_{\mathrm{sup}}>\ln p^{\mathrm{kernel}}_{Mm}>p_{\mathrm{inf}}}} p^{\mathrm{kernel}}_{Mm}(\tau_{M,k}|\tau^{(1)}_m,\cdots,\tau^{(n)}_m)(\ln 10)\tau_{M,k}\Delta\tau_M\right). \tag{3.47}$$

Further, it is necessary to subtract Eq. (3.47) from the correction term in Eq. (3.44) at the same time (thereby the entire approximation function is multiplied by a constant). Thus, for each update of Eqs. (3.43) and (3.44), the numerical integration (3.47) was computed and subtracted from both.

3.6.2 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. K.10 and K.11). According to the OU 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. (3.44) by either $P^{\mathrm{L}}(\tau_m|\tau_M)$ or $P^{\mathrm{R}}(\tau_m|\tau_M)$. We denote the maximum peak times of these approximation functions by $\tau^{\max,\mathrm{L}}_M$ and $\tau^{\max,\mathrm{R}}_M$, and their corresponding k's in Eq. (3.41) by $k^{\max,\mathrm{L}}$ and $k^{\max,\mathrm{R}}$, respectively. Similarly, they are denoted by $\tau^{\max}_M$ and $k^{\max}$ for the kernel part, hereafter. The numerical search of $k^{\max,\mathrm{L}}$, $k^{\max,\mathrm{R}}$, and $k^{\max}$ was conducted in the same way as indicated in Section 3.5.2 in the range $\tau_M > \max\{\tau^{(1)}_m,\cdots,\tau^{(n)}_m\}$.

Figure 3.9 shows the joint p.m.f. of $(k^{\max}, k^{\max,\mathrm{L}})$ and $(k^{\max}, k^{\max,\mathrm{R}})$, which was calculated in the same way as indicated in Section 3.5.2. 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 $k^{\max,\mathrm{L}} = 80$. 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 ($\tau^{\max}_M$) as the estimator of when the event above M will occur, and we discuss the effectiveness of the forecasting based on the estimates.

3.6.3 Estimation of the Next Large Event Timing and Effectiveness of Forecasting

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

$$\delta_n := \frac{\tau_M^* - \tau_M^{\max,\mathrm{n}}}{\tau_M^*}. \tag{3.48}$$

Equation (3.48) considers that the error $|\tau_M^* - \tau_M^{\max,\mathrm{n}}|$ gets larger as τ_M^* becomes longer. The relative error makes it possible to evaluate the accuracy in a manner that is comparable regardless of τ_M^*.

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

$$-\delta_{\mathrm{th}} \le \delta_n \le \delta_{\mathrm{th}}. \tag{3.49}$$

When Eq. (3.49) is satisfied, the estimation at the n-th update is judged to be plausible for the given threshold value δ_{th} in the present study. This is equivalent for the actual occurrence time to be within the range:

$$\frac{\tau_M^{\max,\mathrm{n}}}{(1+\delta_{\mathrm{th}})} \le \tau_M^* \le \frac{\tau_M^{\max,\mathrm{n}}}{(1-\delta_{\mathrm{th}})}. \tag{3.50}$$

Based on the above accuracy at each update, we further evaluated whether a series of estimations yields effective forecasting. Here, effective forecasting implies that $\tau_M^{\max}$ takes a nearly constant value around τ_M^* continuously from well before the 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. (3.49) is satisfied. Further, we denote the last update as the n_{fin}-th update. When the sequence of updates with a sufficiently long $n_{\le\mathrm{th}}$ exists in the range of $n \in (n_{\mathrm{fin}} - n_{\le\mathrm{th}}, n_{\mathrm{fin}}]$, we consider the forecasting to be effective. We judge the stability of $\tau_M^{\max,\mathrm{n}}$ by Eq. (3.49), and therefore, δ_{th} should not be too large. In the present study, we set $\delta_{\mathrm{th}} = 0.5$ and 0.25.

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

$$R_n := 10/\left(t_{n+9} - t_n\right), \tag{3.51}$$

$$\Delta \log_{10} R_n := \log_{10} R_{n+10} - \log_{10} R_n, \tag{3.52}$$

$$\Delta k_n^{\max} := k^{\max,\mathrm{n}+10} - k^{\max,\mathrm{n}}, \tag{3.53}$$

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

3.6.4 Examples of Bayesian Updating

Figures 3.10 – 3.12 show examples of Bayesian updating and other related amounts for the cases where τ_M^* is included in each regime in Table 3.2. δ_{th} was set to 0.5.

Figure 3.10 shows the first example for $\tau_M^* \in$ (I). Figure 3.10(d) indicates that the occurrence rate is high, and it stays almost constant. The kernel part has a peak as shown in Fig. 3.10(a), and its maximum peak time ($\tau_M^{\max,\mathrm{n}}$) continues to be nearly constant around τ_M^* from well before the large event as shown in Fig. 3.10(b); this is confirmed in Fig. 3.10(c), which indicates that $|\delta_n| \le \delta_{\mathrm{th}}$ is satisfied consecutively for $n \in (n_{\mathrm{fin}} - n_{\le\mathrm{th}}, n_{\mathrm{fin}}]$ with a long $n_{\le\mathrm{th}}$, and in Fig. 3.10(e), that shows that $\Delta k_n^{\max}$ fluctuates around 0. Therefore, in this example, τ_M^* is judged to be effectively forecasted.

Figure 3.11 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. 3.11(d). Figure 3.11(a) indicates that the kernel part has a peak and Figs. 3.11(b) and (c) show that the maximum peak time ($\tau_M^{\max,\mathrm{n}}$) transitions to around τ_M^*, and it consecutively satisfies $|\delta_n| \le \delta_{\mathrm{th}} = 0.5$ from long to immediately before the next large event. This is also confirmed by $\Delta k_n^{\max} \approx 0$ in Fig. 3.11(e). Thus in this case also the forecasting is judged to be effective.

Unlike these two examples, in the third example in Fig. 3.12 for $\tau_M^* \in$ (II), the time series is dominated by the non-stationary activity as shown in Figs. 3.12(d) and (e). Although $|\delta_n| \le \delta_{\mathrm{th}} = 0.5$ is satisfied

only immediately before the large shock, $\tau_M^{\max,\mathrm{n}}$ continues shifting and $|\delta_n| \leq \delta_{\mathrm{th}}$ does not hold as shown in Figs. 3.12(b), (c), and (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.

3.6.5 Statistical Analysis of the Effectiveness of Forecasting

We show the results of the statistical analysis on the effectiveness of forecasting. Only the cases of $n_{\mathrm{fin}} \geq 30$ were used in the analysis to ensure that the temporal information of lower intervals was fully reflected in the estimate. Figure 3.13(a) shows the total number of upper intervals (N) obtained from the test data for each $\tau_M^* \in [10^{0.5l}, 10^{0.5(l+1)})$ with $l \in \mathbb{Z}$. Further, N_{30} represents the total number of upper intervals such that $n_{\mathrm{fin}} \geq 30$, which is shown with the ratio to N. The updates included in these N_{30} upper intervals were analyzed.

Figures 3.13(b – d) show the results of the statistical analysis with $\delta_{\mathrm{th}} = 0.5$. Figure 3.13(b) shows the probability (P_{fin}) of $n_{\leq \mathrm{th}} > 0$ (or $|\delta_{n_{\mathrm{fin}}}| \leq \delta_{\mathrm{th}}$) for each τ_M^*. The average of P_{fin} for the overall τ_M^* is about 0.52, and the P_{fin} for each τ_M^* is about the same, except for $\tau_M^* > \langle \tau_M \rangle$ in which P_{fin} takes a higher probability around 0.67. Of such $n_{\leq \mathrm{th}} > 0$ cases, the proportion ($P_{\geq 30}$) of those with relatively long $n_{\leq \mathrm{th}} \geq 30$ is also shown in Fig. 3.13(b) (the probability distribution of $n_{\leq \mathrm{th}}$ is shown in Fig. K.13(a)). Thus the regions of high $P_{\geq 30}$ are overlapped with regimes (I) and (III), though the former is shifted toward larger τ_M^*. On the other hand, $P_{\geq 30}$ is lower in regime (II); it gradually decreases as τ_M^* gets larger. This is consistent with the average of $n_{\leq \mathrm{th}}$ ($\langle n_{\leq \mathrm{th}} \rangle$, this average is taken for $n_{\leq \mathrm{th}} > 0$), but also with the average of its proportion to n_{fin} ($\langle n_{\leq \mathrm{th}} / n_{\mathrm{fin}} \rangle$) as shown in Fig. 3.13(c). This implies that, as the fraction of non-stationary times in $[0, \tau_M^*)$ increases in regime (II), the domination rate of $n_{\leq \mathrm{th}}$ in the total n_{fin}-updates decreases gradually. These properties are preserved for $\delta_{\mathrm{th}} = 0.25$ (Fig. K.14).

Figure 3.14 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 $P_{\geq 30}$, the distribution is almost symmetrically concentrated near the origin. This implies that, when the time series is dominated by stationarity ($\Delta \log_{10} R \approx 0$), the estimated value is stable ($\Delta k^{\max} \approx 0$). On the other hand, if τ_M^* is in regime (II), the probability density-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 ($|\delta_{\mathrm{n_{fin}}}| \leq \delta_{\mathrm{th}}$) is almost independent of the actual occurrence time (τ_M^*). This suggests that the length of the upper interval can be estimated by the inverse probability density function reflecting the temporal pattern of lower intervals, at the last update when the information of the lower intervals can be utilized fully. Second, the stationarity of the time series is related to the stability of the estimate; if the time series is non-stationary, it causes the estimate $\tau_M^{\max}$ to shift. Third, the domination rate of stationarity in the time series determines the effectiveness of forecasting. Immediately or long after the large event, the stationary time series is dominant. Therefore, based on the second point mentioned above, the estimate becomes stable, which leads to an effective forecasting with a relatively long $n_{\leq \mathrm{th}}$. However, these regions are not identical to regimes (I) and (III). This is attributed to lag until the ratio of the non-stationary region in the time series becomes dominant. On the other hand, in regime (II), the non-stationarity becomes gradually dominant, which leads to the shifting of $\tau_M^{\max}$ and shortening of $n_{\leq \mathrm{th}}$.

Finally, we discuss the effectiveness of forecasting in terms of the duration time ($\tau_{\leq \mathrm{th}}$) during the $n_{\leq \mathrm{th}}$ updates. Figure 3.13(d) shows the average of the duration time ($\langle \tau_{\leq \mathrm{th}} \rangle$) and the average of its ratio to the actual occurrence time ($\langle \tau_{\leq \mathrm{th}} / \tau_M^* \rangle$) for each τ_M^* (the probability density of $\tau_{\leq \mathrm{th}}$ is shown in Fig. K.13(b)). Unlike $\langle n_{\leq \mathrm{th}} \rangle$ in Fig. 3.13(c), $\langle \tau_{\leq \mathrm{th}} \rangle$ increases linearly as τ_M^* gets larger, and it is sufficiently long in regime (III). On the other hand, $\tau_{\leq \mathrm{th}}$ is very short in regime (I); however, the ratio $\langle \tau_{\leq \mathrm{th}} / \tau_M^* \rangle$ is high (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.

3.7 Discussion and Conclusions

This chapter considered Bayes' theorem and Bayesian updating on the inter-event times at different magnitude thresholds in marked point processes. The analytical results for the stationary marked Poisson

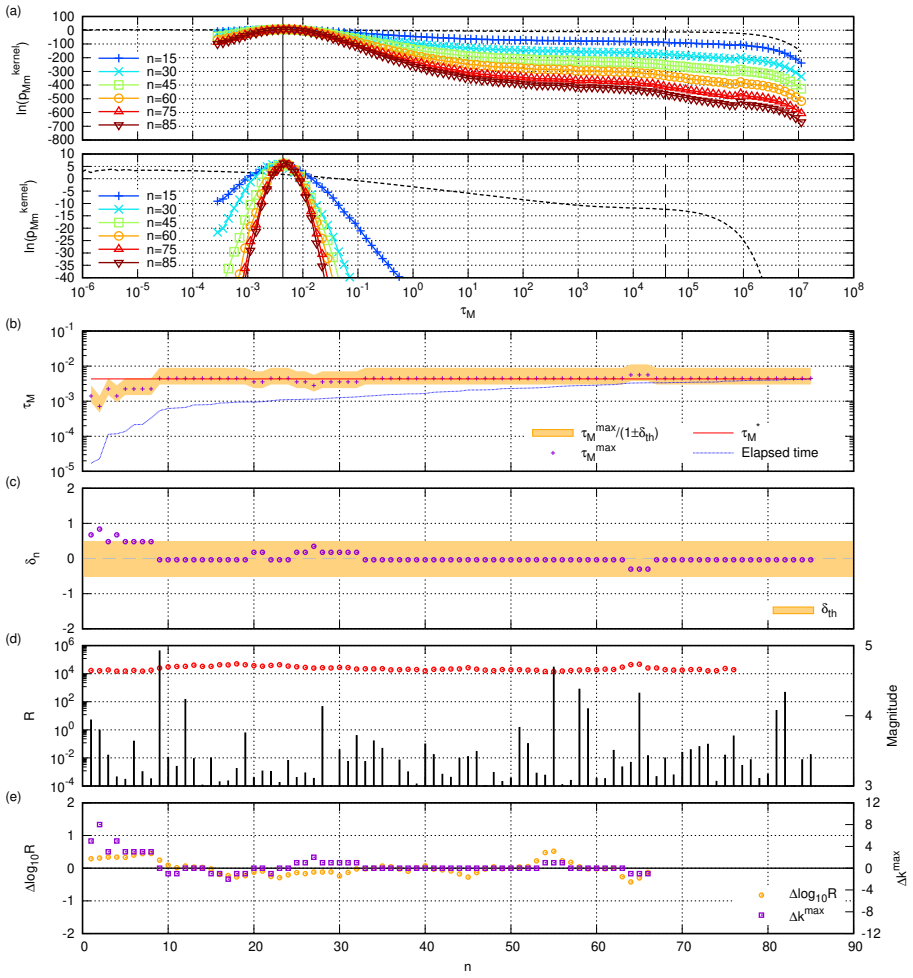

Figure 3.10: The first example of Bayesian updating and related amounts for the case where τ_M^* is in regime (I). (a) The kernel part for each update (n). The lower panel shows the enlarged view of the upper panel near the peak. The dotted curve indicates $p_M(\tau_M)$. The vertical solid line indicates τ_M^* and vertical dotted line the average $\langle\tau_M\rangle$. (b) Evolutions of the estimate ($\tau_M^{\max,\mathrm{n}}$) and the tolerance of error $\left[\frac{\tau_M^{\max,\mathrm{n}}}{1+\delta_{\mathrm{th}}}, \frac{\tau_M^{\max,\mathrm{n}}}{1-\delta_{\mathrm{th}}}\right]$ in Eq. (3.50) with $\delta_{\mathrm{th}} = 0.5$. 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 $[-\delta_{\mathrm{th}}, \delta_{\mathrm{th}}]$. (d) Evolution of the occurrence rate (R_n defined by Eq. (3.51)). The magnitude of the event at each update is indicated by black bars. (e) Evolutions of the variation of the log-occurrence rate ($\Delta \log_{10} R_n$ defined by Eq. (3.52)) and the variation of the log-estimate ($\Delta k_n^{\max}$ defined by Eq. (3.53)).

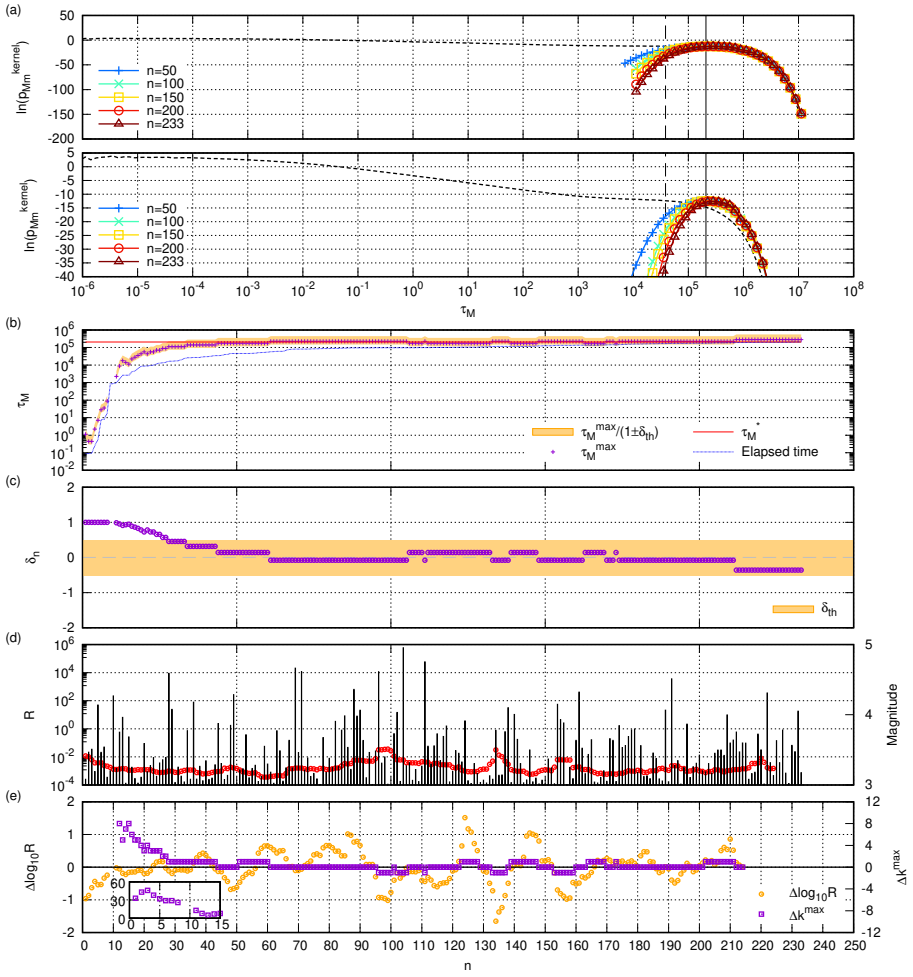

Figure 3.11: The second example for the case that τ_M^* is in regime (Ⅲ). At some updates, the kernel part does not have the maximum peak and the estimate is not determined, which causes some jumps in the time series. The inset in (e) shows $\Delta k_n^{\max}$ at small update counts, indicating a rapid variation of $k^{\max,\mathrm{n}}$ at small n. Other descriptions of the figure are the same as in Fig. 3.10.

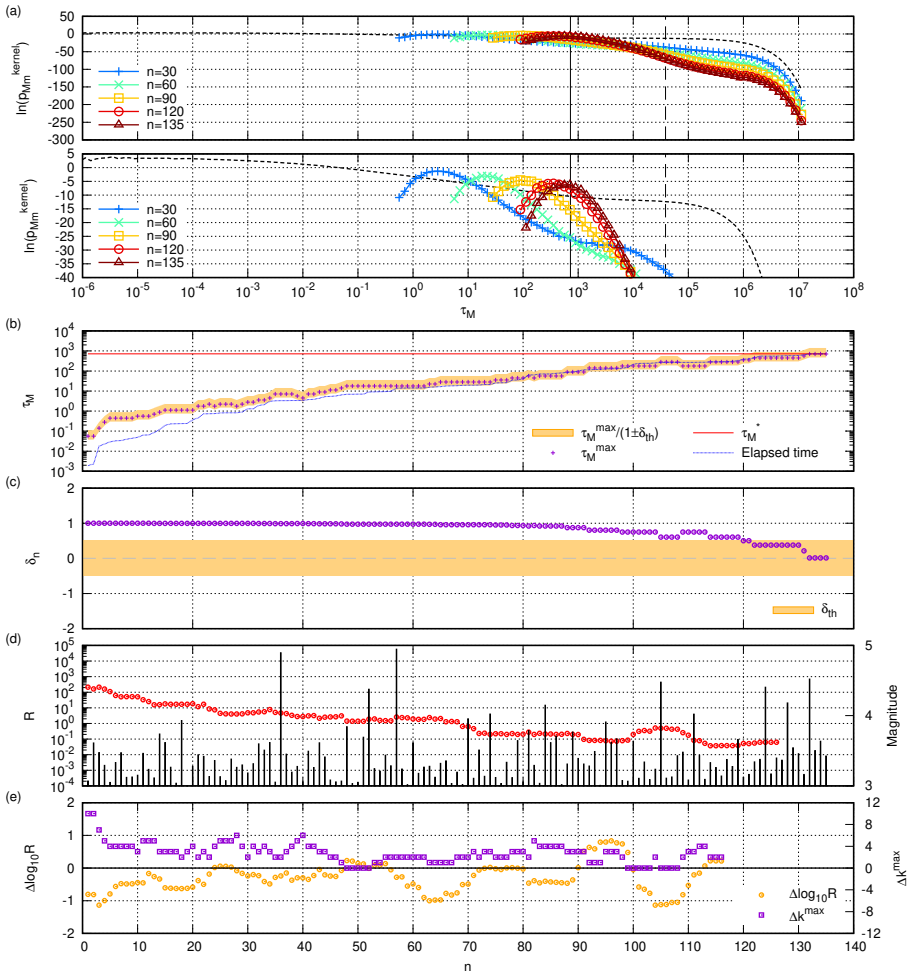

Figure 3.12: The third example for the case that τ_M^* is in regime (II). The descriptions of the figure are the same as in Fig. 3.10.

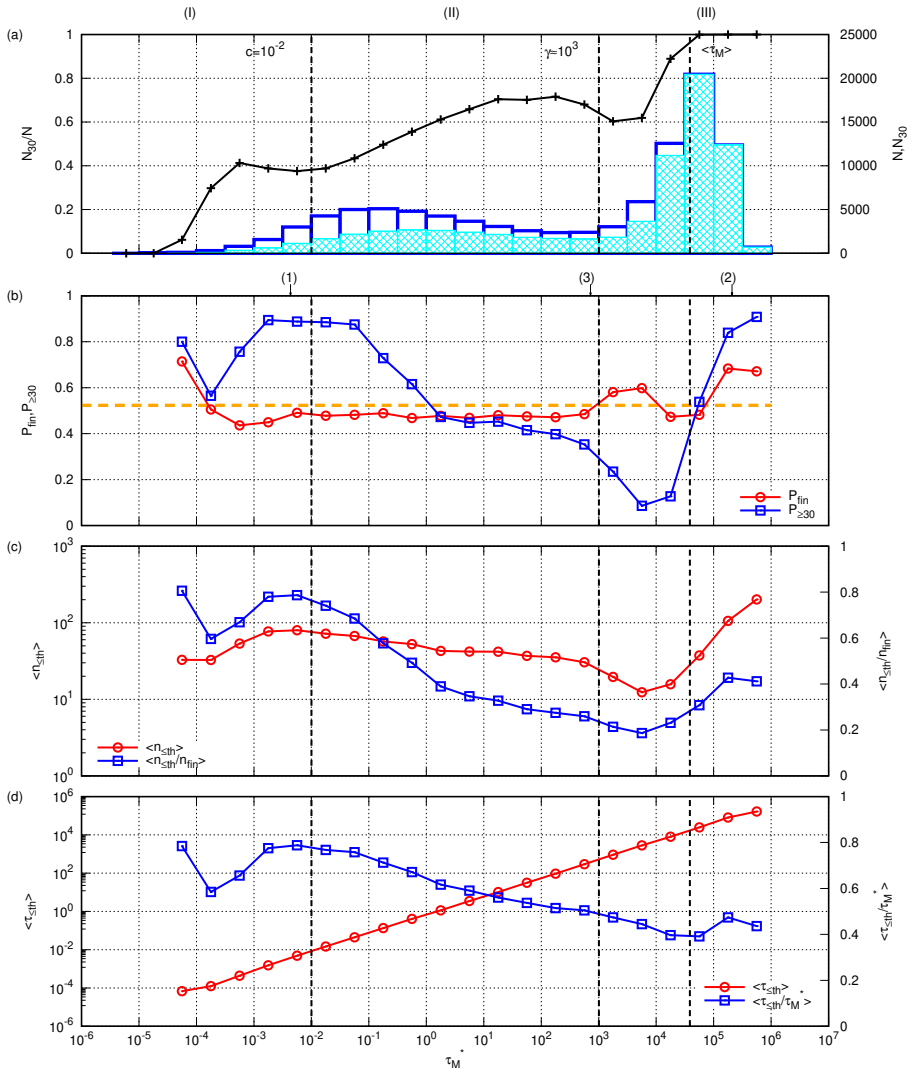

Figure 3.13: (a) (Blue bar) Number of upper intervals N, (Cyan shaded bar) number of upper intervals that include at least 30 updates N_{30}, and (Black points) their ratio N_{30}/N, for each τ_M^* in the test data. (b – d) Statistical results with $\delta_{\rm th} = 0.5$ for each τ_M^*. (1) – (3) indicates the τ_M^*'s of the examples in Figs. 3.10 – 3.12. (b) (Red circle) Probability $P_{\rm fin}$ that $n_{\leq {\rm th}} > 0$ holds (or the probability that Eq. (3.49) is satisfied at the last ($n_{\rm fin}$-th) update), and (Orange dotted line) the average of $P_{\rm fin} \approx 0.52$ for the overall τ_M^*. (Blue square) Proportion $P_{\geq 30}$ of such cases among them where $n_{\leq {\rm th}} \geq 30$. (c) (Red circle) Average of $n_{\leq {\rm th}}$, $\langle n_{\leq {\rm th}} \rangle$, and (Blue square) the average of its proportion to the total number of updates, $\langle n_{\leq {\rm th}}/n_{\rm fin} \rangle$. (d) (Red circle) Average of the duration time $\tau_{\leq {\rm th}}$, $\langle \tau_{\leq {\rm th}} \rangle$ and (Blue aquare) the average of its proportion to the actual occurrence time, $\langle \tau_{\leq {\rm th}}/\tau_M^* \rangle$.

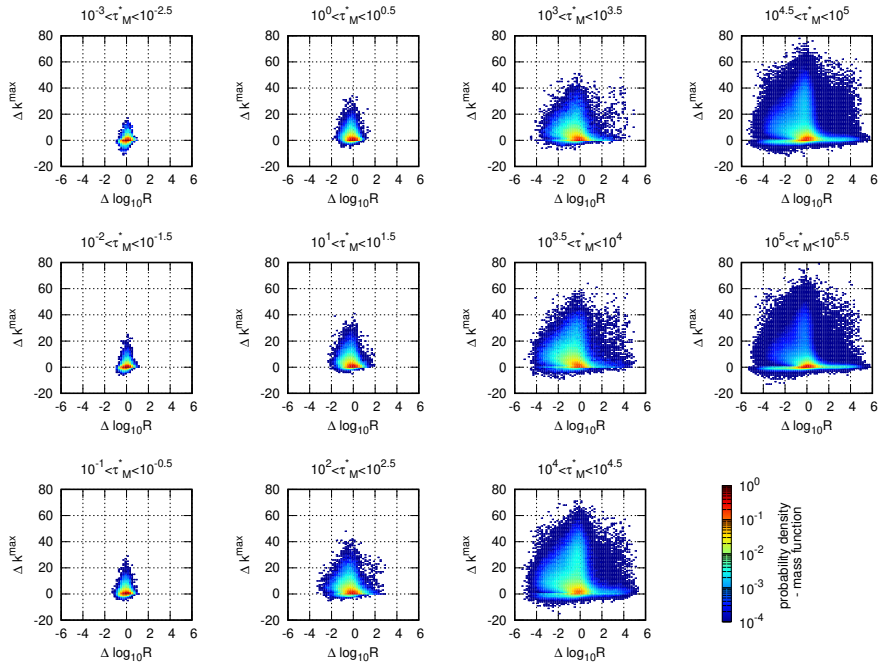

Figure 3.14: Joint probability density-mass functions of $\Delta \log_{10} R$ and $\Delta k^{\max}$ for each τ_M^*.

process 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. This inverse probability density function was derived analytically for the stationary marked Poisson process, 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 stationary marked Poisson process, 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 OU 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 stationary marked Poisson process 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 qualitatively demonstrated the superiority of the Bayesian approach to probabilistic forecasting by the renewal process approach. 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. In Chapter 5, we consider the inverse probability density function in the simplest Bayes' theorem for the time series with weak inter-event correlation.

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 [78], forecasting is considered to be less effective because in such time series, different mainshock-aftershocks sequences overlap [78] 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 [78], and the correlation is easily reflected in the conditional probability. Forecasting is also considered to be improved for the time series with a large branching ratio ($n_{\rm br}$); a larger branching ratio boosts aftershocks for a mainshock [60], which increases the number of updates in the 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 stationary marked Poisson process; finding the corresponding condition for the time series of the ETAS model is 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 [65], 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. [68, 40]) 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. In this way, the Bayesian approach developed in this chapter possibly improves the probabilistic forecasting by point process approach. 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 will be discussed in Chapter 4 [3].

Chapter 4

A Bayesian Inference Method for a Large Magnitude Event in a Spatiotemporal Marked Point Process Representing Seismic Activity

This chapter is based on Ref. [3], which is paper 3 in the list of author's papers.

4.1 Introduction

Chapter 3 presented a Bayesian method to forecast the timing of a large event in marked point processes in a way that improves the renewal process approach in Section 1.2.2 by incorporating the information on small-scale events [2]. This Bayesian method also possibly provides a way to improve the point process approach in Section 1.2.1 by taking into account the effect of temporal patterns on the subsequent major event [2].

This chapter extends the scope of the Bayesian approach to the spatiotemporal marked point process, as shown in Fig. 4.1, to forecast not only the occurrence time but the spatial position of a significant future event incorporating the spatial pattern of epicenters. The main objective of this chapter is to propose a mathematical framework for the Bayesian approach in spatiotemporal marked point processes, which yields an alternative Bayesian method to the preceding study using the Bayesian network [97]. Based on the theory, we discuss the possible contributions of spatiotemporal interactions for forecasting.

Section 4.2 discusses the extensions of the conditional probability and Bayes' theorem in the spatiotemporal marked point process. Section 4.3 considers Bayesian updating in the spatiotemporal marked point process, and Section 4.4 discusses how space-time interactions contribute to Bayesian inference. Finally, Section 4.5 is for discussion and conclusions.

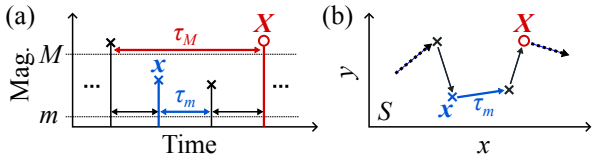

Figure 4.1: Schematic of a spatiotemporal seismic activity represented by (a) a marked point process with a magnitude as a mark, and (b) corresponding jumps of events in the spatial area S. The upper and lower spatiotemporal pairs $\{\boldsymbol{X}, \tau_M\}$ and $\{\boldsymbol{x}, \tau_m\}$ are shown, and these form the combination $\{\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M\}$.

4.2 Bayes' Theorem in Spatiotemporal Marked Point Process

First, the conditional probability for the spatiotemporal marked point process is defined (Fig. 4.1). Let τ_M (τ_m) represent the length of an inter-event time interval at the upper magnitude threshold M (the lower magnitude threshold $m(< M, M := m + \Delta m)$). Further, let $\boldsymbol{X}(\in S)$ ($\boldsymbol{x}(\in S)$) represent the spatial position of the event at the right (left) end of an upper (lower) inter-event time interval (Fig. 4.1). Here S is a closed region on the earth's surface. Although there is no restriction on its size, we assume that S encompasses seismogenic zones as centrally as possible; if a seismogenic zone lies on the edge of S, only a limited portion of aftershocks following a mainshock at X can be considered, potentially impeding precise probabilistic evaluation within this theory.

In the present chapter, the pair of the above-defined spatial position and time interval is referred to as the (spatiotemporal) pair; $\{\boldsymbol{X}, \tau_M\}$ is referred to as the upper (spatiotemporal) pair, and $\{\boldsymbol{x}, \tau_m\}$ the lower (spatiotemporal) pair. When the upper pair $\{\boldsymbol{X}, \tau_M\}$ includes the lower pair $\{\boldsymbol{x}, \tau_m\}$ as shown in Fig. 4.1(a), these constitute the combination of the upper and lower spatiotemporal pairs $\{\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M\}$. The magnitude of the event at $\boldsymbol{x}$ is exceptionally greater than M when the lower pair is located at the leftmost in an upper pair, though otherwise, it is always less than or equal to M. Thus, $p_{mM}(\boldsymbol{x}, \tau_m|\boldsymbol{X}, \tau_M)$ represents the spatiotemporal conditional probability density that a lower spatiotemporal pair is $\{\boldsymbol{x}, \tau_m\}$ under the condition that it is included in the upper spatiotemporal pair $\{\boldsymbol{X}, \tau_M\}$.

We consider Bayes' theorem for the above-defined spatiotemporal conditional probability.
Let $dN_{mM}(\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M)$ represent the number of the combinations of the upper and lower spatiotemporal pairs such that the upper pair falls within $[\boldsymbol{X}, \boldsymbol{X} + d\boldsymbol{X})$ and $[\tau_M, \tau_M + d\tau_M)$, and the lower pair included in it falls within $[\boldsymbol{x}, \boldsymbol{x} + d\boldsymbol{x})$ and $[\tau_m, \tau_m + d\tau_m)$, in the time series. For simplicity, $dN_{mM}(\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M)$ is referred to as the number of the combination $\{\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M\}$ without mentioning the infinitesimal intervals, and hereafter, other numbers of spatiotemporal pairs or combinations denoted by dN are referred to in the same way. $dN_{mM}(\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M)$ can be expressed in two ways:

$$\begin{aligned} &dN_{mM}(\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M)/d\boldsymbol{x}d\tau_m d\boldsymbol{X}d\tau_M \\ &\quad = N_M p_M(\boldsymbol{X}, \tau_M)\frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\boldsymbol{X},\tau_M}}p_{mM}(\boldsymbol{x}, \tau_m|\boldsymbol{X}, \tau_M) \\ &\quad = N_m p_m(\boldsymbol{x}, \tau_m)p_{Mm}(\boldsymbol{X}, \tau_M|\boldsymbol{x}, \tau_m). \end{aligned} \tag{4.1}$$

Here, N_M (N_m) represents the total number of upper (lower) spatiotemporal pairs in the time series. $p_M(\boldsymbol{X}, \tau_M)$ ($p_m(\boldsymbol{x}, \tau_m)$) yields the joint probability density that the upper (lower) pair is $\{\boldsymbol{X}, \tau_M\}$ ($\{\boldsymbol{x}, \tau_m\}$). $p_{Mm}(\boldsymbol{X}, \tau_M|\boldsymbol{x}, \tau_m)$ yields the inverse probability density that the upper pair is $\{\boldsymbol{X}, \tau_M\}$ under the condition that the lower pair $\{\boldsymbol{x}, \tau_m\}$ is found within it. $\langle\langle\tau_m\rangle\rangle_{\boldsymbol{X},\tau_M}$ represents the average time interval of the spatiotemporal conditional probability (extension of Eq. (2.2)):

$$\langle\langle\tau_m\rangle\rangle_{\boldsymbol{X},\tau_M} := \int_S d\boldsymbol{x}\int_0^\infty d\tau_m \tau_m p_{mM}(\boldsymbol{x}, \tau_m|\boldsymbol{X}, \tau_M).$$

From Eq. (4.1), Bayes' theorem is derived as (extension of Eq. (3.4)):

$$p_{Mm}(\boldsymbol{X}, \tau_M|\boldsymbol{x}, \tau_m) = 10^{-b\Delta m}\frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\boldsymbol{X},\tau_M}}\frac{p_{mM}(\boldsymbol{x}, \tau_m|\boldsymbol{X}, \tau_M)}{p_m(\boldsymbol{x}, \tau_m)}p_M(\boldsymbol{X}, \tau_M), \tag{4.2}$$

where the GR law ($N_M/N_m = 10^{-b\Delta m}$) is used.

The normalization condition of the inverse probability in Eq. (4.2) can be checked using the integral equation with the spatiotemporal conditional probability in its kernel, which is derived below. Let $dN_m(\boldsymbol{x}, \tau_m)$ represent the number of the spatiotemporal pairs $\{\boldsymbol{x}, \tau_m\}$ in the time series, then $dN_m(\boldsymbol{x}, \tau_m)$ can be expressed in two ways:

$$\begin{aligned} dN_m(\boldsymbol{x}, \tau_m)/d\boldsymbol{x}d\tau_m &= N_m p_m(\boldsymbol{x}, \tau_m) \\ &= N_M\int_S d\boldsymbol{X}\int_0^\infty d\tau_M\frac{\tau_M p_{mM}(\boldsymbol{x}, \tau_m|\boldsymbol{X}, \tau_M)}{\langle\langle\tau_m\rangle\rangle_{\boldsymbol{X},\tau_M}}p_M(\boldsymbol{X}, \tau_M). \end{aligned}$$

Thus, the integral equation that connects the probability density functions of the upper and lower pair is derived as (extension of Eq. (2.3)):

$$p_m(\boldsymbol{x}, \tau_m) = 10^{-b\Delta m}\int_S d\boldsymbol{X}\int_0^\infty d\tau_M\frac{\tau_M p_{mM}(\boldsymbol{x}, \tau_m|\boldsymbol{X}, \tau_M)}{\langle\langle\tau_m\rangle\rangle_{\boldsymbol{X},\tau_M}}p_M(\boldsymbol{X}, \tau_M). \tag{4.3}$$

The inverse probability density function for only upper and lower time intervals or only upper and lower spatial positions can be derived from the joint probability for the combination of an upper and a lower spatiotemporal pair. Let $p_{mM}(\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M)$ represent the joint probability density of the combination $\{\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M\}$. The total number of the combinations of an upper and a lower spatiotemporal pair in the time series is N_m, and therefore, $p_{mM}(\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M)$ can be expressed in two ways using Eq. (4.1) as follows:

$$\begin{aligned} p_{mM}(\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M) &= \frac{dN_{mM}(\boldsymbol{x}, \tau_m; \boldsymbol{X}, \tau_M)}{N_m d\boldsymbol{x} d\tau_m d\boldsymbol{X} d\tau_M} \\ &= 10^{-b\Delta m} \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\boldsymbol{X}, \tau_M}} p_{mM}(\boldsymbol{x}, \tau_m | \boldsymbol{X}, \tau_M) p_M(\boldsymbol{X}, \tau_M) \\ &= p_m(\boldsymbol{x}, \tau_m) p_{Mm}(\boldsymbol{X}, \tau_M | \boldsymbol{x}, \tau_m). \end{aligned} \tag{4.4}$$

Thus, Bayes' theorem for time intervals already obtained in Chapter 3 (Eq. (3.4)) can be derived by marginalizing the joint probability in Eq. (4.4) for $\boldsymbol{x}$ and $\boldsymbol{X}$ as:

$$p_{Mm}(\tau_M | \tau_m) = 10^{-b\Delta m} \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} \frac{p_{mM}(\tau_m | \tau_M)}{p_m(\tau_m)} p_M(\tau_M).$$

Further, the integral equation for time intervals introduced in Chapter 3 (Eq. (2.3)) can also be derived by marginalizing Eq. (4.3) for $\boldsymbol{x}$ and $\boldsymbol{X}$, noting that the integrand of the r.h.s. of Eq. (4.3) is the joint probability in Eq. (4.4), as:

$$p_m(\tau_m) = 10^{-b\Delta m} \int_0^\infty d\tau_M \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m | \tau_M) p_M(\tau_M).$$

Bayes' theorem for spatial positions can also be derived. First, we define the following quantity:

$$\langle n(\boldsymbol{x}, \tau_m) \rangle_{\boldsymbol{X}, \tau_M} := \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\boldsymbol{X}, \tau_M}} p_{mM}(\boldsymbol{x}, \tau_m | \boldsymbol{X}, \tau_M). \tag{4.5}$$

This is the average number of specific lower spatiotemporal pairs $\{\boldsymbol{x}, \tau_m\}$ included in the upper pair $\{\boldsymbol{X}, \tau_M\}$. Bayes' theorem for the spatial position is obtained by marginalizing the joint probability in Eq. (4.4) for τ_m and τ_M while using Eq. (4.5):

$$p_{Mm}(\boldsymbol{X} | \boldsymbol{x}) = 10^{-b\Delta m} \frac{\langle n(\boldsymbol{x}) \rangle_{\boldsymbol{X}}}{p_m(\boldsymbol{x})} p_M(\boldsymbol{X}). \tag{4.6}$$

Here, $\langle n(\boldsymbol{x}) \rangle_{\boldsymbol{X}}$ is the average number of such events with magnitude $\in (m, M]$ that occur at $\boldsymbol{x}$ and are in between two consecutive large events with magnitudes $> M$, the latter of which occurs at $\boldsymbol{X}$. The integral equation for the spatial position can be derived by marginalizing Eq. (4.3) for τ_m and τ_M using Eq. (4.5) and noting again that the integrand in Eq. (4.3) is the joint probability in Eq. (4.4), as follows:

$$p_m(\boldsymbol{x}) = 10^{-b\Delta m} \int_S d\boldsymbol{X} \langle n(\boldsymbol{x}) \rangle_{\boldsymbol{X}} p_M(\boldsymbol{X}). \tag{4.7}$$

The normalization condition for the inverse probability in Eq. (4.6) can be checked using Eq. (4.7). In particular, consider the case where spatial positions and time intervals are independent. In this case, the average number of lower intervals in an upper interval is $10^{b\Delta m}$ according to the GR law, and therefore, $\langle n(\boldsymbol{x}) \rangle_{\boldsymbol{X}} = 10^{b\Delta m} p_{mM}(\boldsymbol{x} | \boldsymbol{X})$. Thus, Eqs. (4.6) and (4.7) are simplified as:

$$\begin{aligned} p_{Mm}(\boldsymbol{X} | \boldsymbol{x}) &= \frac{p_{mM}(\boldsymbol{x} | \boldsymbol{X})}{p_m(\boldsymbol{x})} p_M(\boldsymbol{X}), \\ p_m(\boldsymbol{x}) &= \int_S d\boldsymbol{X} p_{mM}(\boldsymbol{x} | \boldsymbol{X}) p_M(\boldsymbol{X}). \end{aligned}$$

4.3 Bayesian Updating in Spatiotemporal Marked Point Process

We extend Bayes' theorem to Bayesian updating. In this chapter, for simplicity, the sequence of consecutive lower spatiotemporal pairs $\{\boldsymbol{x}_1, \tau_m^{(1)}, \cdots, \boldsymbol{x}_n, \tau_m^{(n)}\}$, such that all the pairs are included in the

same upper pair, is denoted by $\{\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}\}$ (Fig. 4.2). When this sequence is included in the upper pair $\{\boldsymbol{X}, \tau_M\}$, these constitute the combination of the upper and lower pairs $\{\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M\}$. Let $dN_{mM}(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M)$ represent the total number of such a combination in the time series. Thus, $dN_{mM}(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M)$ can be expressed in two ways:

$$dN_{mM}(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M)/d\boldsymbol{X}d\tau_M d\boldsymbol{x}^n d\tau_m^n = N_M p_M(\boldsymbol{X}, \tau_M)\langle n(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M}$$
$$= N_m R_n p_m(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) p_{Mm}(\boldsymbol{X}, \tau_M|\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}). \quad (4.8)$$

Here, R_n represents the proportion of the number of sequences of consecutive n lower pairs such that all these pairs are included in the same upper pair in the time series to the total number of sequences of consecutive n lower pairs in the time series. Thus, $N_m R_n$ represents the total number of consecutive n lower pairs belonging to the same upper pair and the total number of combinations of an upper and consecutive n lower pairs in a time series. In the stationary marked Poisson process generated with a constant occurrence rate in the ETAS model, $R_n = (1 - \langle\tau_m\rangle/\langle\tau_M\rangle)^{n-1}$ by the GR law (see Section 3.4.1).

Further, $p_m(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})$ represents the joint probability density that a sequence of consecutive n lower pairs belonging to the same upper interval takes $\{\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}\}$. $p_{Mm}(\boldsymbol{X}, \tau_M|\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})$ represents the inverse probability density that the upper pair is $\{\boldsymbol{X}, \tau_M\}$ when $\{\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}\}$ is found within it. $\langle n(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M}$ represents the average number of the sequence $\{\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}\}$ in the upper pair $\{\boldsymbol{X}, \tau_M\}$. From Eq. (4.8), the inverse probability density is:

$$p_{Mm}(\boldsymbol{X}, \tau_M|\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) = 10^{-b\Delta m} R_n^{-1} \frac{\langle n(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M}}{p_m(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})} p_M(\boldsymbol{X}, \tau_M). \quad (4.9)$$

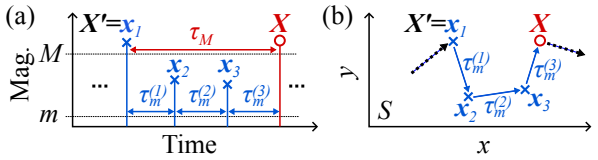

Figure 4.2: Schematic of spatiotemporal seismic activity showing time intervals and spatial positions of events considered in the Bayesian updating; the upper pair $\{\boldsymbol{X}, \tau_M\}$ and the consecutive lower pairs $\{\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}\}$ ($n = 3$ in the figure) included in it are indicated. These form the combination of upper and lower pairs $\{\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M\}$.

We derive the inverse probability density for only time intervals or only spatial positions. Let $p_{mM}(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M)$ represent the joint probability density that the combination of an upper and consecutive lower spatiotemporal pairs is $\{\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M\}$, which can be expressed as:

$$p_{mM}(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M) = \frac{dN_{mM}(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M)}{N_m R_n d\boldsymbol{X} d\tau_M d\boldsymbol{x}^n d\tau_m^n}$$
$$= 10^{-b\Delta m} R_n^{-1} \langle n(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M} p_M(\boldsymbol{X}, \tau_M)$$
$$= p_m(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) p_{Mm}(\boldsymbol{X}, \tau_M|\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}). \quad (4.10)$$

Thus, the inverse probability density function for time intervals is derived by marginalizing Eq. (4.10) for $\boldsymbol{X}, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$,

$$p_{Mm}(\tau_M|\tau_m^{(1:n)}) = 10^{-b\Delta m} R_n^{-1} \frac{\langle n(\tau_m^{(1:n)})\rangle_{\tau_M}}{p_m(\tau_m^{(1:n)})} p_M(\tau_M). \quad (4.11)$$

This is consistent with Eq. (3.23) in Chapter 3. On the other hand, the inverse probability density for spatial positions can be derived by marginalizing Eq. (4.10) for $\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}$,

$$p_{Mm}(\boldsymbol{X}|\boldsymbol{x}_{1:n}) = 10^{-b\Delta m} R_n^{-1} \frac{\langle n(\boldsymbol{x}_{1:n})\rangle_{\boldsymbol{X}}}{p_m(\boldsymbol{x}_{1:n})} p_M(\boldsymbol{X}). \quad (4.12)$$

4.4 Spatiotemporal Intercorrelation in Bayesian Inference

We examine how spatiotemporal correlations appear in Bayesian updating by comparing two cases of correlations among spatial positions and time intervals as shown in Fig. 4.3. Variables connected by an arrow in Fig. 4.3 are assumed to be correlated (meaning neither independent nor conditionally independent given other variables), and variables not connected are assumed to be not only independent, but conditionally independent given other variables. However, the conditional independence between the lower time intervals ($\tau_m^{(i)}$'s) in the same upper time interval is not assumed given the length of their upper time interval τ_M. In the first case, space-time correlations are not assumed (Fig. 4.3(a)), whereas in the second case, correlations between spatial position and time interval indicated with the green arrows in Fig. 4.3(b) are added.

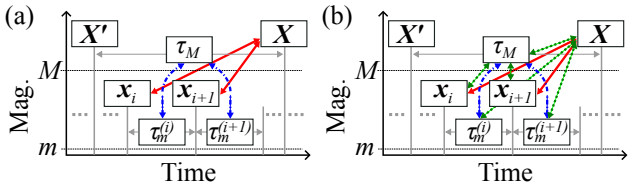

Figure 4.3: Two cases of correlations assumed between time intervals and spatial positions. Colored arrows represent correlations between (red, solid) spatial positions, (blue, dot-dash) time intervals, and (green, dotted) a time interval and a spatial position, respectively.

For the first case, Eq. (4.8) is rewritten as (see Appendix L.1):

$$\begin{aligned} &dN_{mM}(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M)/d\boldsymbol{X}d\tau_M d\boldsymbol{x}^n d\tau_m^n \\ &= N_M p_M(\boldsymbol{X}) p_M(\tau_M) \langle n(\tau_m^{(1:n)})\rangle_{\tau_M} \prod_{i=1}^n p_{mM}(\boldsymbol{x}_i|\boldsymbol{X}) \\ &= N_m R_n p_{Mm}(\boldsymbol{X}|\boldsymbol{x}_{1:n}) p_{Mm}(\tau_M|\tau_m^{(1:n)}) \prod_{i=1}^n p_m(\boldsymbol{x}_i) p_m(\tau_m^{(i)}). \end{aligned} \tag{4.13}$$

Therefore, the Bayesian updating is expressed as the product of the following two respective updates for the time interval and the spatial position.

$$p_{Mm}(\tau_M|\tau_m^{(1:n)}) = 10^{-b\Delta m} R_n^{-1} \frac{\langle n(\tau_m^{(1:n)})\rangle_{\tau_M}}{\prod_{i=1}^n p_m(\tau_m^{(i)})} p_M(\tau_M), \tag{4.14}$$

$$p_{Mm}(\boldsymbol{X}|\boldsymbol{x}_{1:n}) = \frac{\prod_{i=1}^n p_{mM}(\boldsymbol{x}_i|\boldsymbol{X})}{\prod_{i=1}^n p_m(\boldsymbol{x}_i)} p_M(\boldsymbol{X}). \tag{4.15}$$

For the second case, Eq. (4.8) is rewritten as (see Appendix L.2):

$$\begin{aligned} dN_{mM}(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M)/d\boldsymbol{X}d\tau_M d\boldsymbol{x}^n d\tau_m^n &= N_M p_M(\boldsymbol{X}, \tau_M) \langle n(\tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M} \prod_{i=1}^n p_{mM}(\boldsymbol{x}_i|\boldsymbol{X}, \tau_M) \\ &= N_m R_n p_{Mm}(\boldsymbol{X}, \tau_M|\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) \prod_{i=1}^n p_m(\boldsymbol{x}_i) p_m(\tau_m^{(i)}). \end{aligned} \tag{4.16}$$

Thus, the inverse probability is:

$$p_{Mm}(\boldsymbol{X}, \tau_M|\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) = 10^{-b\Delta m} R_n^{-1} \frac{\langle n(\tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M}}{\prod_{i=1}^n p_m(\tau_m^{(i)})} \frac{\prod_{i=1}^n p_{mM}(\boldsymbol{x}_i|\boldsymbol{X}, \tau_M)}{\prod_{i=1}^n p_m(\boldsymbol{x}_i)} p_M(\boldsymbol{X}, \tau_M).$$

The joint probability is obtained by taking the ratio of Eq. (4.16) to $N_m R_n$, in the same way as Eq.

(4.10).

$$p_{mM}(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}; \boldsymbol{X}, \tau_M) = 10^{-b\Delta m} R_n^{-1} \langle n(\tau_m^{(1:n)}) \rangle_{\boldsymbol{X}, \tau_M} p_M(\boldsymbol{X}, \tau_M) \prod_{i=1}^{n} p_{mM}(\boldsymbol{x}_i | \boldsymbol{X}, \tau_M)$$
$$= p_{Mm}(\boldsymbol{X}, \tau_M | \boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) \prod_{i=1}^{n} p_m(\boldsymbol{x}_i) p_m(\tau_m^{(i)}). \tag{4.17}$$

Marginalizing Eq. (4.17) with respect to $\boldsymbol{X}$ or τ_M, the inverse probability density functions for τ_M and $\boldsymbol{X}$ are obtained as follows.

$$p_{Mm}(\tau_M | \boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) = 10^{-b\Delta m} R_n^{-1} \frac{\langle n(\tau_m^{(1:n)}) \rangle_{\tau_M}}{\prod_{i=1}^{n} p_m(\tau_m^{(i)})} \frac{\prod_{i=1}^{n} p_{mM}(\boldsymbol{x}_i | \tau_M)}{\prod_{i=1}^{n} p_m(\boldsymbol{x}_i)} p_M(\tau_M), \tag{4.18}$$

$$p_{Mm}(\boldsymbol{X} | \boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) = 10^{-b\Delta m} R_n^{-1} \frac{\langle n(\tau_m^{(1:n)}) \rangle_{\boldsymbol{X}}}{\prod_{i=1}^{n} p_m(\tau_m^{(i)})} \frac{\prod_{i=1}^{n} p_{mM}(\boldsymbol{x}_i | \boldsymbol{X})}{\prod_{i=1}^{n} p_m(\boldsymbol{x}_i)} p_M(\boldsymbol{X}). \tag{4.19}$$

Equations (4.18) and (4.19) show that the terms representing the correlations between spatial position and time interval are multiplied explicitly to the closed updates for time and space in Eqs. (4.14) and (4.15), respectively.

4.5 Discusson and Conclusions

This chapter examined a Bayesian inference method of the occurrence time and location of the next significant event in a spatiotemporal marked point process using occurrence patterns of smaller events. As described in Section 1.1 in seismic activity, temporal quiescence, spatial gap, or activation (foreshocks) is sometimes recognized to precede major earthquakes [6, 27, 5]. The question of whether the conditional probability can quantitatively treat such qualitative spatiotemporal characteristics of seismic activity and whether the Bayesian updating method can be used for better probabilistic forecasting in actual seismic activity is one for the future.

The study in this chapter only presents a theoretical framework. To verify the effectiveness of this framework for forecasting, numerical examination, as performed in Chapter 3, with a sufficient number of synthetic seismic data that stochastic models (e.g., the HIST-ETAS model [53, 55]) or physical models (e.g., the Olami-Feder-Christensen model [47]) can generate is necessary. Furthermore, for forecasting actual earthquakes, it is required to examine the framework with seismic catalog data. The preliminary analysis (see Appendix M) using a seismic catalog in Southern California [18, 19] does not show apparent improvement in forecasting by adding spatial information ($\boldsymbol{x}$) to the inverse probability. The cause of this seems to be in setting the spatial domain S and the way to subdivide it. Improving this point is important in future detailed analysis.

In this study, the spatiotemporal pair is defined in a way that builds upon Chapter 3, and Bayesian updating is considered based on it. However, this definition includes exceptional events whose magnitude exceeds M in the lower pair. Further, the information of the lower pair immediately before the next large event is not used in real-time Bayesian updating. Therefore, in the practical use of the Bayesian approach, it may be better to define the lower pair as a lower time interval and the spatial position of its subsequent event whose magnitude $\leq M$, in the same way as the upper pair. Bayesian updating based on this definition should be considered excluding the rightmost lower pair in an upper pair.

Chapter 5

Statistical Properties of Inter-event Times in Seismic Time Series Transformed by Occurrence Rate: An Analysis from the Viewpoint of Hierarchy in the Temporal Nature of Seismicity

This chapter is based on Ref. [4], which is paper 4 in the list of author's papers.

5.1 Introduction

Chapters 3 and 4 have shown a theoretical framework of the Bayesian approach and tested it using synthetic time series of seismicity. However, the inverse probability and its approximation function have not been derived for time series with inter-event correlations; the Bayesian approach has yet to achieve probabilistic forecasting in actual seismic time series, and thus, examining the Bayesian framework using seismic catalogs is necessary for its practical forecasting. Therefore, this chapter aims to examine the properties of the conditional and inverse probabilities in the Bayesian approach using seismic catalogs. As a first step in catalog analysis, the analysis is simplified in two aspects.

First, only Bayes' theorem (Eq. (3.7)) is examined; it is interesting in relation to forecasting whether the inverse probability shows seismicity-dependent unimodality even in this simplest case of Bayesian updating, and it is essential to examine the conditional probability for considering general Bayesian updating.

Second, the subject of analysis is restricted to the time series with weak inter-event correlation; such time series can be easy to work on by perturbation from the analytical results for the stationary marked Poisson process. As such, in this study, two kinds of time series inspired by the studies on the scaling universality in the inter-event time distributions (described in Section 1.3.1) are analyzed: the stationary time series and the aftershock sequence transformed by the occurrence rate. We regard the scaling of the inter-event time distribution in Eq. (1.8) as a temporal transformation of these time series by smooth functions of the event rate [27, 98, 99, 100] (Section 5.2 describes details), interpreting the universality as the nature of the temporal fluctuations in the unfolding-transformed time series. This viewpoint can avoid the problem pointed out in Section 1.3.1 and enables us to examine the scaling universality following the original view. Such transformed time series are suitable for analyzing the Bayesian approach in that the inter-event correlations are weak and enable us to consider the scaling universality in light of the hierarchical property of seismic time series while including aftershock sequences. This way, the properties of all three probability distributions associated with Bayes' theorem (Eq. (3.7)) are analyzed,

including their interrelationships.

Thus, this chapter examines seismic time series after temporal transformation to explore the characteristics of the probability densities (the conditional and inverse probabilities and the inter-event time distribution) in Bayes' theorem (Eq. (3.7)); along the way, the universality of the inter-event time distribution is discussed via the integral equation (2.3), based on the characteristics of the temporal hierarchy of the seismic time series. Section 5.2 describes the temporal transformation procedure of seismic time series, information on seismic catalogs and how to set the analysis range, the method to extract time series from these catalogs, their temporal transformation, scaling property of inter-event time distributions of after-transformed time series, and methods for analyzing the conditional probability of these time series. Section 5.3 shows the results of catalog analysis relating to the conditional probability. Section 5.4, based on the results in Section 5.3, derives the conditional probability, the scaling function of the inter-event time distribution, and the inverse probability in this order and compares them with the results of catalog analysis. Finally, Section 5.5 summarizes and concludes the study and presents some discussions.

5.2 Methods

5.2.1 Transformation of Time Series

This subsection describes the time series transformation applied in this chapter. Let t_k and $\tau_k (:= t_{k+1} - t_k)$ be the k-th event's occurrence time and the k-th inter-event time in a time series, respectively. In the scaling of Eq. (1.8) for an aftershock sequence, time intervals are transformed by instantaneous occurrence rate as follows [80]:

$$\tau_k \mapsto R(t_{k+1})\tau_k. \tag{5.1}$$

However, this transformation has ambiguity in terms of the choice of time to take the occurrence rate (any time between t_k and t_{k+1} can be the argument of $R(\cdot)$ [80]). Instead, we apply the following transformation that is a modification of Eq. (5.1):

$$\tau_k \mapsto R(t_{k+1})t_{k+1} - R(t_k)t_k. \tag{5.2}$$

This is equivalent to converting the occurrence times:

$$t_k \mapsto R(t_k)t_k. \tag{5.3}$$

In particular, for stationary time series ($R(t) \approx R$ (const.)) and aftershock sequences ($R(t) \simeq Kt^{-p}$), the transformation of Eq. (5.3) is:

$$t_k \mapsto \begin{cases} Rt_k & \text{(stationary)}, \\ Kt_k^{1-p} & \text{(aftershock)}. \end{cases} \tag{5.4}$$

For these time series, the transformation in Eq. (5.4) is equivalent to transforming the time to the following transformed time (the same transformation for aftershock sequences was already considered in Ref. [27, 101]), except for the coefficient:

$$t_k \mapsto z_k := \int_{t_{\min}}^{t_k} R(s)ds, \tag{5.5}$$

where $t_{\min}$ is its starting time ($t_{\min} = 0$) for stationary time series and the lower bound of the time range in which $R(t) \simeq Kt^{-p}$ holds for aftershock sequences. Actually, Eqs. (5.5) for the stationary time series and the aftershock sequence ($p \neq 1$) are as follows:

$$t_k \mapsto z_k = \begin{cases} Rt_k & \text{(stationary)}, \\ \frac{Kt_k^{1-p}}{1-p} - \frac{Kt_{\min}^{1-p}}{1-p} & \text{(aftershock)}. \end{cases} \tag{5.6}$$

The constant ($Kt_{\min}^{1-p}/(1-p)$) in Eq. (5.6) does not affect the inter-event time statistics in $\{z_k\}$.

After these transformations using temporal variation trends ($R(t)$), the time series are aligned to stationary time series with an average occurrence rate of 1. Thus, the temporal fluctuations in different

time series become comparable with each other. The statistics of intervals for $\{z_k\}$ suggest the nature of such fluctuations, and Eq. (1.8) can be regarded as implying the universality of such fluctuations. This type of transformation has been used in the field of quantum chaos and is known as the unfolding transformation [98, 99, 100]. In the following, the transformation described above is referred to as the unfolding transformation or unfolding procedure.

Thus, this study performs the following transformation for the stationary and aftershock time series in seismic catalogs:

$$t_k \mapsto w_k = \begin{cases} t_k & \text{(stationary)}, \\ t_k^{1-p} & \text{(aftershock, } p < 1), \\ -t_k^{1-p} & \text{(aftershock, } p > 1). \end{cases} \tag{5.7}$$

The unfolding procedure is completed by further scaling the transformed time series with the average occurrence rate over the entire time series ($w_k \mapsto z_k := w_k/\langle w \rangle$), which can be done by applying the rescaling of the inter-event time distribution for these transformed time series. These time series ($\{w_k\}$) are the objects of catalog analysis. Note that, hereafter in this chapter, for the inter-event times of the transformed time series, the same symbol of τ is used as was used for the inter-event times before the transformation.

5.2.2 Catalog Information and Procedure of Time Series Transformation

In the approach of this study, three seismic catalogs are used; one is the global Centroid Moment Tensor (CMT) catalog [21, 22, 23], and the two are local catalogs in Southern California (SCEDC) [18, 19] and Japan (JMA) [16, 17]. For each earthquake catalog, the author determined the space-time windows to extract stationary and aftershock time series and the magnitude ranges for analysis while considering the catalog completeness in the following way.

CMT Catalog

The global CMT catalog covering large earthquakes worldwide [21, 22, 23] was used. The moment magnitude is unaffected by saturation at the large scale [Section 5.2 in Ref. [5]][13, 14, 15] and thus is appropriate for this study, where the magnitude values are significant to examine the dependence of statistics on cut-off magnitude. The CMT catalog starts from 01/01/1976, though, as shown in Fig. 5.1(a), the number of recorded events in 1976 is less than in other years, and thus the author chose the term to analyze 01/01/1977 – 12/31/2022. No spatial restriction was imposed. Figure 5.2(a) shows $p(M)$ and $P(M)$ for the CMT catalog in the above period. The completeness magnitude M_c calculated by the MBASS method [32] was 5.4. The figure also shows the b-value ($\hat{b}$ hereafter) calculated for each neighboring two points (at magnitude M and $M + 0.1$, where 0.1 is the increment of magnitude in this chapter) of $P(M)$, i.e., $\hat{b} = -(\log_{10} P(M + 0.1)/P(M))/0.1$. Referring to these results (the range larger than or equal to M_c, and where $\hat{b}$ fluctuation is as small as possible, in particular, avoid the large magnitudes where $\hat{b}$ fluctuation becomes very large) and the apparent shape of the graph ($\log_{10} p(M)$ versus M being close to the straight line the GR law indicates), the author determined the magnitude range to set the cut-off magnitude in the following analysis to be $5.4 \leq M \leq 7.5$. The b-value calculated in this magnitude range using the maximum likelihood estimate [33, 34] was about 0.990. The cumulative number of earthquakes $N(t)$ since 01/01/1977 at 00:00:00 in the above-determined space-time window with magnitudes ≥ 5.4 is shown in Fig. 5.1(b); the whole time series can be judged to be almost stationary by the nearly linear form of $N(t)$, $N(t) \propto t$, as in the preceding studies [79, 80] (though, precisely speaking, there appears to be gradual increasing trend even for $M \geq 5.4$ as observed in Fig. 5.1(a), which may due to the improvements in analysis method [22]).

Southern California Catalog

The Southern California catalog [18, 19] contains the earthquakes in the spatial range $112°W$ – $123°W$ and $29°N$ – $38°N$ and the time domain 01/01/1981 – 03/31/2022. The process to determine the space-time and magnitude ranges to analyze from this catalog was as follows. First, the author divided the spatial area into smaller cells of $0.5° \times 0.5°$. Then, M_c was calculated by the MBASS method for each cell

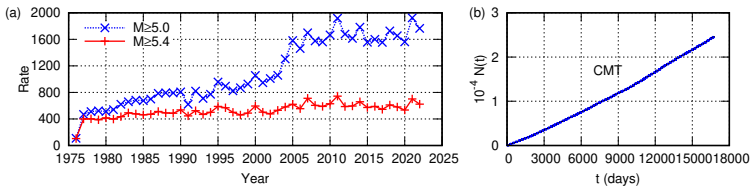

Figure 5.1: (a) Annual number of earthquakes recorded in the global CMT catalog (blue × symbol with dotted line for only events with magnitude ≥ 5 and red + symbol with bold line for ≥ 5.4) and (b) the cumulative number of earthquakes ($N(t)$) recorded in the CMT catalog with magnitude ≥ 5.4 against the time (t) from January 1, 1977, at 00:00:00.

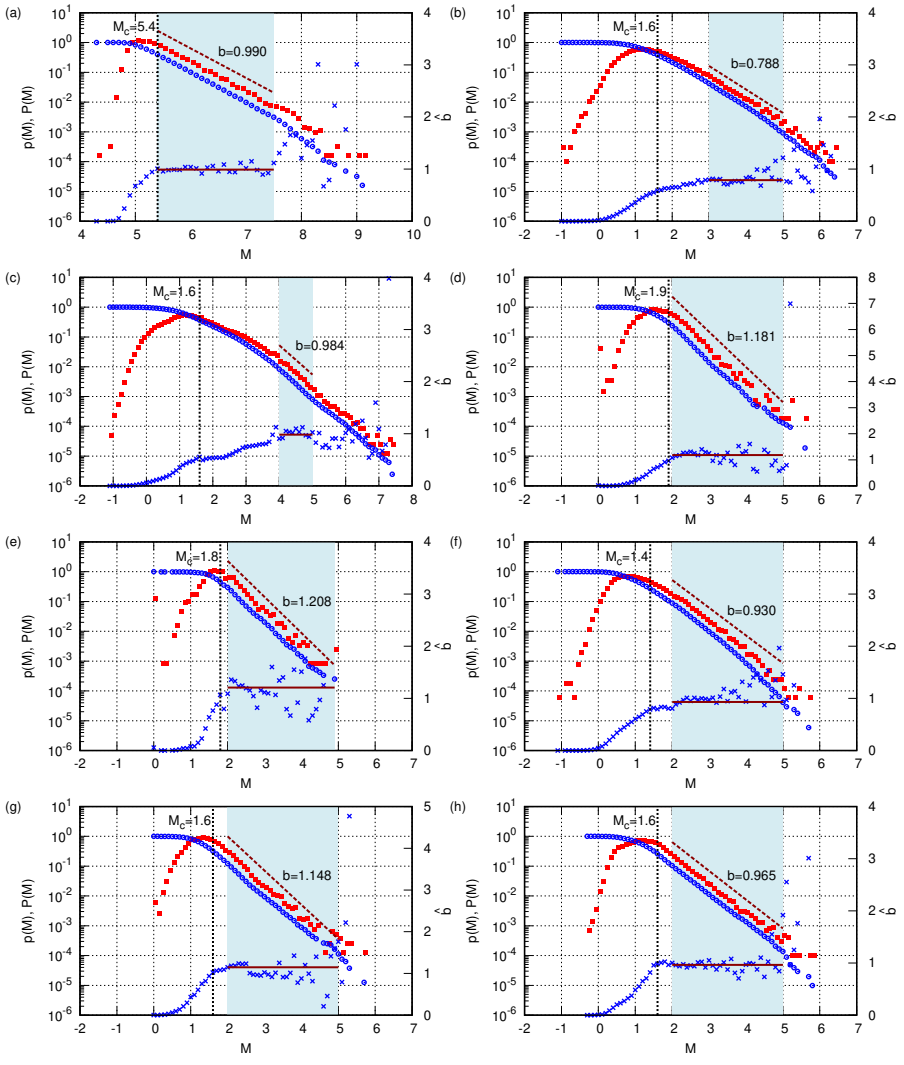

Figure 5.2: (Red filled square) $p(M)$, (Blue circle) $P(M)$, and (Blue ×) $\hat{b}$ for (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. The vertical dotted line shows the completeness magnitude (M_c) calculated by the MBASS method [32]. The magnitude range to set the cut-off magnitude is shaded by light blue, the dotted line shows the GR law with the b-values determined using the maximum likelihood estimate [33, 34]. The horizontal line indicates the b-value for comparison with $\hat{b}$.

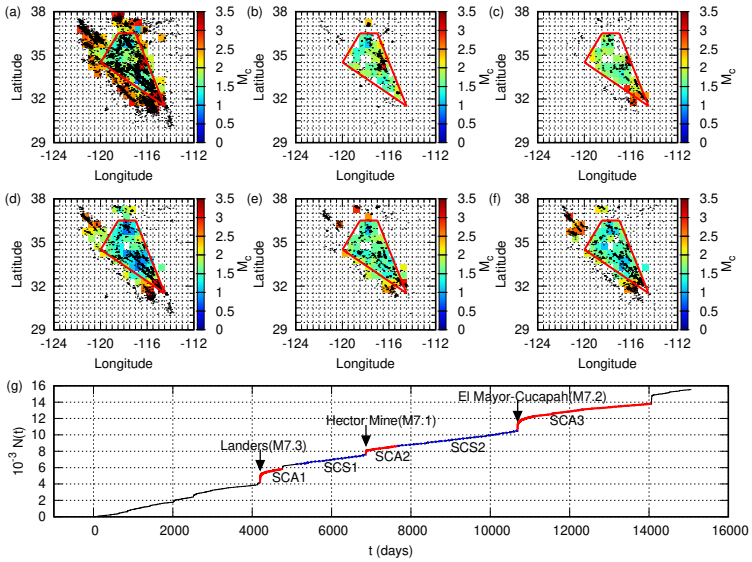

Figure 5.3: Spatial distribution of M_c calculated by the MBASS method for the Southern California catalog of the time domain (a) 01/01/1981 – 03/31/2022, (b) SCA1, (c) SCA2, (d) SCA3, (e) SCS1, and (f) SCS2. The black dots show the epicenters of the earthquake with magnitude ≥ 3. The red quadrilateral represents the spatial domain to be analyzed. Red $\times$ symbols in (b-d) indicate the mainshock epicenter in each time domain. (g) $N(t)$ for earthquakes with magnitude ≥ 3, from January 1, 1981, at 00:00:00; the stationary domains (SCS1 and SCS2) are indicated by blue dotted lines and mainshock-aftershocks (SCA1, SCA2, and SCA3) are by bold red curves.

with at least 100 events to check its spatial distribution; this process refers to the method used in Ref. [40] (the same as described in Section 1.1). Figure 5.3(a) shows the distribution map of M_c. This result suggests that the M_c at the periphery tends to take higher values (around 3 or higher) than the central part, as preceding studies have already suggested [102, 37, 40]. Figure 5.3(g) shows $N(t)$ for earthquakes with magnitude ≥ 3; based on this figure, the author selected two stationary time series (SCS1 and SCS2) and three mainshock-aftershock sequences (SCA1, SCA2, and SCA3; the aftershocks of Landers, Hector Mine, and El Mayor-Cucapah earthquakes [103]) (see Table 5.1 for details on time domains). The distributions of M_c were re-calculated for the earthquakes in each time window; Figures 5.3(b-f) show the M_c map for each time domain. From Figs. 5.3(b-f), the author determined by sight the quadrilateral area with the vertices at $(120°W, 34.5°N)$, $(118.5°W, 36.5°N)$, $(117°W, 36.5°N)$, and $(114.5°W, 31.5°N)$, which is the spatial window excluding the periphery with high M_c as much as possible and including the three mainshock epicenters; the earthquakes included in this area were the target of the analysis below.

After setting the space-time window by the above-mentioned process, the author determined the magnitude ranges to set the cut-off magnitude for the two stationary time series. Referring to the M_c values by the MBASS method and $\hat{b}$ values, the author visually determined the magnitude range the GR exponential decay holds (see Figs. 5.2(g, h) and Table 5.1). The maximum likelihood estimate yielded the b-values in these ranges as summarized in Table 5.1. On the other hand, for the three mainshock-aftershock sequences, the author further narrowed down the time domain in which $R(t) \propto t^{-p}$ holds. Figure 5.5 shows the number of aftershocks with magnitude ≥ 2 per day versus time (t) from each mainshock. As shown in the figure, $R(t) \propto t^{-p}$ holds for each aftershock sequence after sufficient time passed from the mainshock. Based on the figure, the author chose the time domain where $R(t) \propto t^{-p}$ holds by sight and computed the p-values by fitting within these regimes (see Table 5.2 for details). This procedure can exclude the time domain immediately after the mainshock, where M_c tends to become large [38, 39]. In this way, the space-time windows for the aftershock regimes were set. For the events within these space-time windows, the author determined the magnitude ranges and the b-values in the same manner as for the stationary time series; see Figs. 5.2(d-f) for the magnitude frequency and Table 5.1 for the b-values.

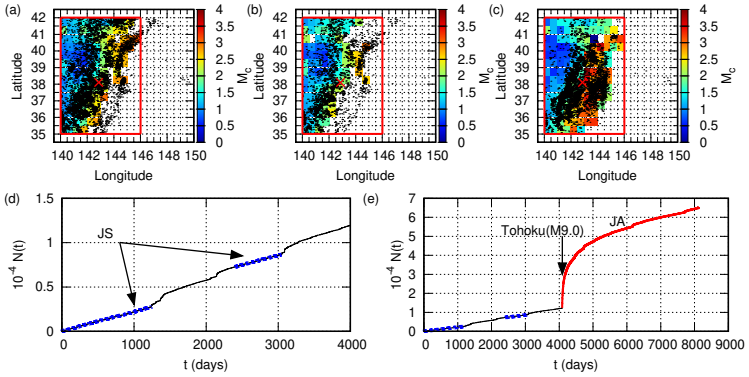

Figure 5.4: Spatial distribution of M_c calculated by the MBASS method for the JMA catalog around the epicenter of Tohoku mainshock (shown by red $\times$ symbol) in the time domain (a) 01/01/2000 – immediately before the mainshock time, (b) JS, and (c) JA. The black dots show the epicenters of earthquakes with magnitude (a, b) ≥ 3 and (c) ≥ 4 in $140°E$ – $150°E$ and $35°N$ – $42°N$ (note that the range shown is slightly wider than this) for each time domain. Inside the red rectangle is the spatial domain to analyze. (d, e) $N(t)$ for earthquakes with magnitude ≥ 3 in the area $140°E$ – $150°E$, $35°N$ – $42°N$ from January 1, 2000, at 00:00:00 (JST); the bold red curve indicates the aftershock sequence (JA) and the dotted blue lines stationary regimes (JS).

JMA Catalog

The JMA catalog [16, 17] covers earthquakes throughout Japan, and in this study, the author focused on the spatial area around the Tohoku mainshock (M9) in 2011. First, the author provisionally set the spatial domain $140°E$ – $150°E$ and $35°N$ – $42°N$ and the time domain 01/01/2000 – 03/31/2022. Then, the author examined the spatial distribution of M_c in the same way as the Southern California catalog, before (from 01/01/2000 to immediately before the Tohoku mainshock time, Fig. 5.4(a)) and after (from the Tohoku mainshock time to 03/31/2022, Fig. 5.4(c)) the Tohoku mainshock. Figures 5.4(a) and (c) suggest that M_c are partially as high as 3 to 4. Although the cut-off magnitude 3 is somewhat low compared to these results, the author examined $N(t)$ for earthquakes with magnitude ≥ 3 as shown in Figs. 5.4(d, e) and selected the stationary time domains (JS) and the aftershock sequence of the Tohoku mainshock (JA) (see Table 5.1 for the respective time domains). Figure 5.4(b) shows the M_c distribution in JS calculated by the MBASS method. Taking the results in Figs. 5.4(b, c) into account, the author determined the spatial domain to analyze to the inside of the rectangle $140°E$ – $146°E$ and $35°N$ – $42°N$.

Figure 5.2(b) shows the magnitude frequency in the above-set space-time window of JS. As in the case of Southern California, the author determined the magnitude range by visual inspection from the linear part of Fig. 5.2(b) such that larger than M_c calculated by the MBASS method. On the other hand, for JA, the author narrowed down the time domain to the regime obeying $R(t) \propto t^{-p}$; From the occurrence rate graph in Fig. 5.5, the author visually determined such time domain to be $10 < t$. After setting the space-time window for JA, the author finally determined the magnitude range as other time series based on Fig. 5.2(c). The magnitude ranges and the b-values for JS and JA are summarized in Table 5.1.

Time Series Transformation

Using the above-mentioned process, the author obtained four stationary time series (CMT, JS, SCS1, and SCS2) and aftershock sequences (JA, SCA1, SCA2, and SCA3). The author applied the transformation in Eq. (5.7) for these time series. Figure 5.6 shows the cumulative number of aftershocks ($\hat{N}(w)$) against the transformed time $w(= \pm t^{1-p})$. Note that the transformation is different for each aftershock sequence depending on the p-value. After the transformation, one can see that $\hat{N}(w)$ increases linearly for the transformed time w in the time ranges shown in Table 5.2. In the following, the author analyzes these transformed aftershock time series (JA, SCA1, SCA2, SCA3) as well as the stationary time series (CMT, JS, SCS1, SCS2). Hereafter, these unfolding-transformed time series are referred to by the names in Table 5.1. The author performed the catalog analysis for these time series by setting the cut-off magnitude in

Table 5.1: Information on spatial domain, time domain, magnitude range set by the author, and the b-value determined by the maximum likelihood estimate in the magnitude range, the total number of earthquakes in the spatiotemporal domain with magnitude $\geq$ the lower limit of the magnitude range, of seismic catalogs used in this study. For the mainshock-aftershock sequences (JA, SCA1, SCA2, SCA3), the number of events is bracketed because the events to be analyzed are more limited (see Table 5.2). The first part of JS includes events up to just before the M7.1 earthquake on May 26, 2003, at 18:24:33.42 (JST). Also, the second part of JS includes events up to April 30, 2008; the term is set to avoid activity that could be considered as foreshocks prior to the M7.0 event on May 8, 2008, at 1:45:18.77 (JST). SCA1 includes events up to just before the Northridge earthquake (M6.7) on January 17, 1994, SCS1 just before the Hector Mine earthquake (M7.1), SCS2 just before the El Mayor-Cucapah earthquake (M7.2), and SCA3 just before the foreshock (M6.4) of the 2019 Ridgecrest M7.1 earthquake [103].

Catalog	Name	Spatial domain	Time domain	Magnitude range	b-value	Number
CMT	CMT	Whole world	01/01/1977 – 12/31/2022	$5.4 \leq M \leq 7.5$	0.990	24648
JMA	JS	$140°E$ – $146°E$	01/01/2000 – 05/26/2003	$3 \leq M \leq 5$	0.788	2737
		$35°N$ – $42°N$	08/01/2006 – 04/30/2008	$3 \leq M \leq 5$		1377
	JA		03/11/2011 – 03/31/2022	$4 \leq M \leq 5$	0.984	(9162)
SCEDC	SCA1	The area with vertices:	06/28/1992 – 01/17/1994	$2 \leq M \leq 5$	1.181	(17308)
	SCS1	$(120°W, 34.5°N)$,	01/01/1995 – 10/16/1999	$2 \leq M \leq 5$	1.148	9665
	SCA2	$(118.5°W, 36.5°N)$,	10/16/1999 – 12/31/2001	$2 \leq M \leq 4.9$	1.208	(9890)
	SCS2	$(117°W, 36.5°N)$,	01/01/2002 – 04/04/2010	$2 \leq M \leq 5$	0.965	10589
	SCA3	$(114.5°W, 31.5°N)$	04/04/2010 – 07/04/2019	$2 \leq M \leq 5$	0.930	(18852)

Table 5.2: Time ranges of aftershock sequences to be analyzed. The p-value and the number of events within the time range are also shown. p-values are calculated for events with magnitude $\geq$ minimum magnitude in Table 5.1.

Name	$R(t) \propto t^{-p}$ range	p-value	Number
JA	$10 < t$	0.895	6706
SCA1	$10 < t$	1.061	12856
SCA2	$10^{0.5} < t < 10^2$	1.005	3531
SCA3	$10 < t$	0.820	14978

increments of 0.1.

5.2.3 Scaling of Inter-event Time Distributions

Figure 5.7 shows the inter-event time distribution after rescaled by the average ($R_m = 1/\langle \tau_m \rangle$) as in Eq. (1.8), for each unfolding-transformed time series. This figure is drawn using the following method, which is equivalent to changing the bin width according to the inter-event time length [77]. First, the inter-event times at cut-off magnitude m are transformed as $\tau_m \mapsto \sigma_m := \log_{10} \tau_m$. Let $\tilde{N}_m(\sigma_m)$ denote the number of such transformed intervals of length within $[\sigma_m, \sigma_m + \Delta\sigma)$ in the time series, where $\Delta\sigma$ is fixed to 0.1 in this chapter. Then the probability density function of this transformed intervals ($\tilde{p}_m(\sigma_m)$) is calculated as:

$$\tilde{p}_m(\sigma_m) = \frac{\tilde{N}_m(\sigma_m)}{N_m \Delta\sigma}.$$

$p_m(\tau_m)$ is obtained by re-transforming $\tilde{p}_m(\sigma_m)$ by $\sigma_m \mapsto \tau_m$. By using the class mark $(\sigma_m + 0.5\Delta\sigma)$ for $d\sigma_m/d\tau_m = 1/(10^{\sigma_m} \ln 10)$,

$$p_m(\tau_m) = \frac{1}{10^{\sigma_m + 0.5\Delta\sigma} \ln 10} \frac{\tilde{N}_m(\sigma_m)}{N_m \Delta\sigma}. \tag{5.8}$$

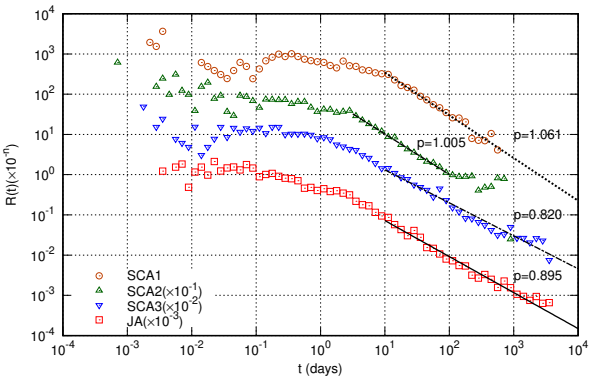

Figure 5.5: Number of aftershocks with magnitude $\geq$ minimum magnitude in Table 5.1 per day against time from the mainshock. SCA2, SCA3, and JA results are multiplied by 10^{-1}, 10^{-2}, and 10^{-3}, respectively. Black (dotted) lines are the fitting curves by $R(t) \propto t^{-p}$. Fitting ranges and the p-values are summarized in Table 5.2.

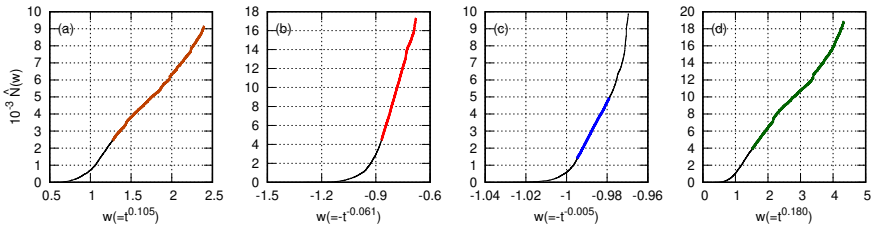

Figure 5.6: Cumulative number of earthquakes with magnitude $\geq$ minimum magnitude in Table 5.1 ($\hat{N}(w)$) against transformed time from the mainshock $w = \pm t^{1-p}$ for (a) JA, (b) SCA1, (c) SCA2, and (d) SCA3. Bold lines indicate the stationary regimes in the transformed aftershock sequences.

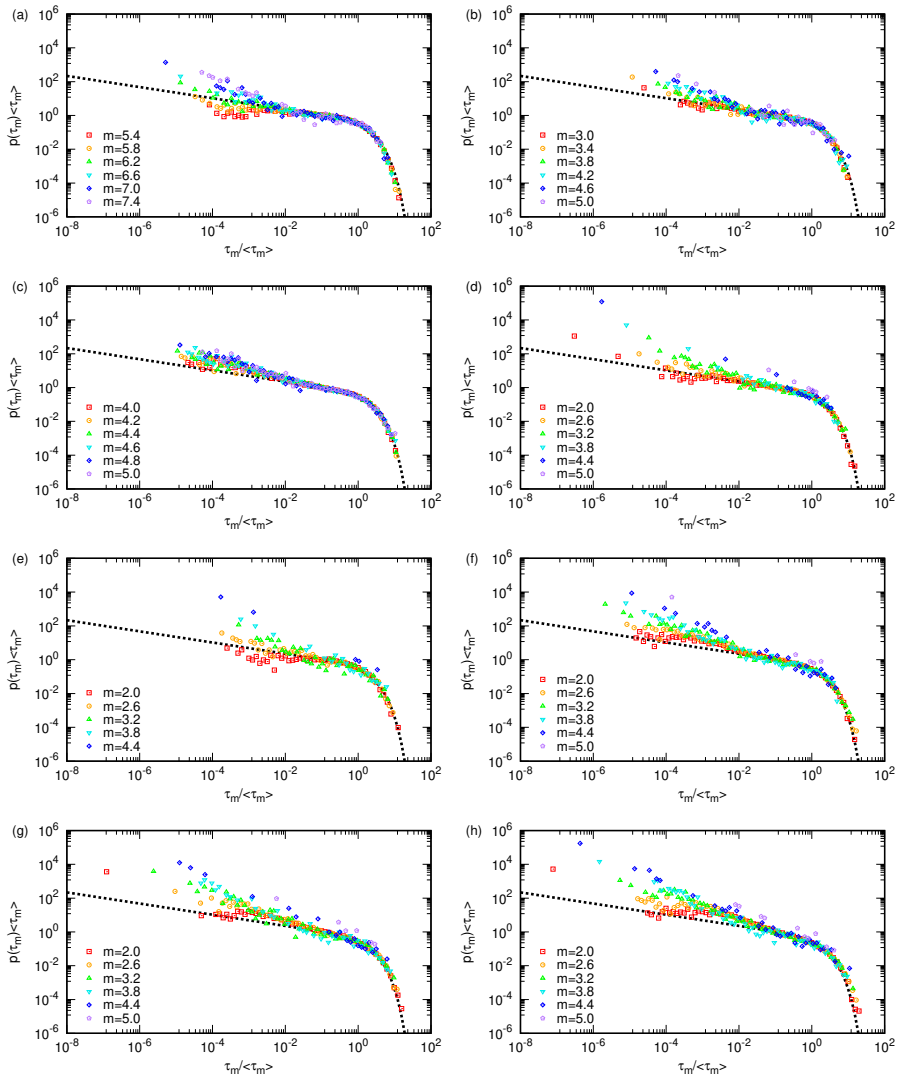

Figure 5.7: Rescaled inter-event time distributions by the average interval $\langle \tau_m \rangle$ for the unfolding-transformed time series of (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. For each time series, the results of several cut-off magnitudes are shown. The dotted curve is the gamma distribution in Eq. (1.9).

In this way, the inter-event time distribution is drawn by plotting Eq. (5.8) against $10^{\sigma_m + 0.5\Delta\sigma}$. Finally, the results shown in Fig. 5.7 are obtained through further rescaling by $\langle \tau_m \rangle$.

One can see that the rescaled distributions are around the same curve that is expressed in Eq. (1.9), except for very short intervals ($\tau_m / \langle \tau_m \rangle \lesssim 10^{-2}$). For such short intervals, the rescaled distributions, in many cases, tend to take larger values than those used in Eq. (1.9); this may be due in part to the limitation in terms of accuracy, and such limitation is often removed by setting a minimum interval (around 10 seconds [87] to 2 minutes [77]). Also, preceding studies provided theoretical explanations for this trend using the ETAS model [83, 84, 85].

These characteristics observed in Fig. 5.7 are consistent with the results of preceding catalog analyses [72, 73, 77]. Thus, the scaling property of the inter-event time distribution was also confirmed for the unfolding-transformed time series.

5.2.4 Method to Examine the Conditional Probability

The properties of the conditional probability are examined as the change in each component of the representation Eq. (3.10) of the conditional probability from that for the stationary marked Poisson Process. Correlations between events in seismic time series change these components from Eqs. (3.11) and (3.14). The conditional probability in time series with weak inter-event correlations can be studied by examining such changes. The subsequent section presents the findings from analyzing the unfolding-transformed time series extracted in the preceding subsection using this approach.

5.3 Results of Catalog Analysis

5.3.1 Results for the Component $\Psi_{mM}(i|\tau_M)$

Scaling Property for the Average of the Conditional Probability

First, the author shows the scaling property for the average of the conditional probability. By definition, the length of an inter-event time interval at the upper cut-off magnitude (τ_M) divided by this average ($\langle\langle\tau_m\rangle\rangle_{\tau_M}$) is equal to the average number of intervals at the lower cut-off magnitude included in that upper interval, i.e.,

$$\frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} = \sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M). \tag{5.9}$$

This demonstrates that the scaling property associated with Eq. (5.9) as well as the average of the conditional probability is linked only to the component $\Psi_{mM}(i|\tau_M)$.

Subtracting 1 from Eq. (5.9) yields the average number of magnitude $\geq m$ events included in the inter-event time of length τ_M at the upper cut-off magnitude M. Figure 5.8 shows the results after scaling this average using the factor $A_{\Delta m}$ (in Eq. (3.12)) for magnitude and the average interval $\langle\tau_M\rangle$ for time, across each time series. These results were obtained in the following way.

Define the transformed upper interval $\sigma_M := \log_{10}\tau_M$. For a set of cut-off magnitudes (m, M), let $\tilde{N}_M(\sigma_M)$ be the number of inter-event intervals at the upper cut-off magnitude, whose length is within $[\sigma_M, \sigma_M + \Delta\sigma)$. Also, let $\tilde{N}_m(\sigma_M)$ be the total number of inter-event intervals at the lower cut-off magnitude included in these intervals at the upper cut-off magnitude. Then the result shown in Fig. 5.8 for a specific pair (m, M) is obtained by plotting the following against $10^{\sigma_M+0.5\Delta\sigma}/\langle\tau_M\rangle$:

$$\frac{1}{A_{\Delta m}}\left(\frac{\tilde{N}_m(\sigma_M)}{\tilde{N}_M(\sigma_M)} - 1\right). \tag{5.10}$$

Figure 5.8 shows that the average numbers $(\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M} - 1)$ computed by the above way collapse on a curve $(\zeta(y))$ independent of the cut-off magnitudes (m, M), after rescaling by $\langle\tau_M\rangle$ and $A_{\Delta m}$, as follows:

$$\frac{\frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} - 1}{A_{\Delta m}} = \zeta\left(\frac{\tau_M}{\langle\tau_M\rangle}\right). \tag{5.11}$$

Hereafter, the variable $y := \tau_M/\langle\tau_M\rangle$ is used.

Seismic catalogs have a limitation in the number of events. Therefore, in particular for the combination of (m, M) with large Δm, the number of sample data to calculate Eq. (5.10) becomes insufficient. This leads to increased fluctuations in the results and makes it difficult to identify a trend. To obtain the scaling function $(\zeta(y))$ utilizing as many sample data as possible, the author used all sample data obtained from the combinations of (m, M) with the same Δm together as the population, based on the above result in the following way.

Define the rescaled and transformed upper interval $\sigma := \log_{10} y$. For all (m, M) with the same Δm (i.e., $(m, M) = (m_0, m_0 + \Delta m)$, $(m_1, m_1 + \Delta m)$, $\cdots$, $(m_n, m_n + \Delta m)$, where $m_j = m_{\min} + 0.1j$ ($m_{\min}$ is the minimum magnitude in Table 5.1) and n is the upper limit of j for the given magnitude range and Δm), let $\hat{N}_{m_j+\Delta m}(\sigma)$ be the total number of rescaled and transformed inter-event time intervals of length within $[\sigma, \sigma + \Delta\sigma)$ at the cut-off magnitude $m_j + \Delta m$. Also, let $\hat{N}_{m_j}(\sigma)$ be the total number of

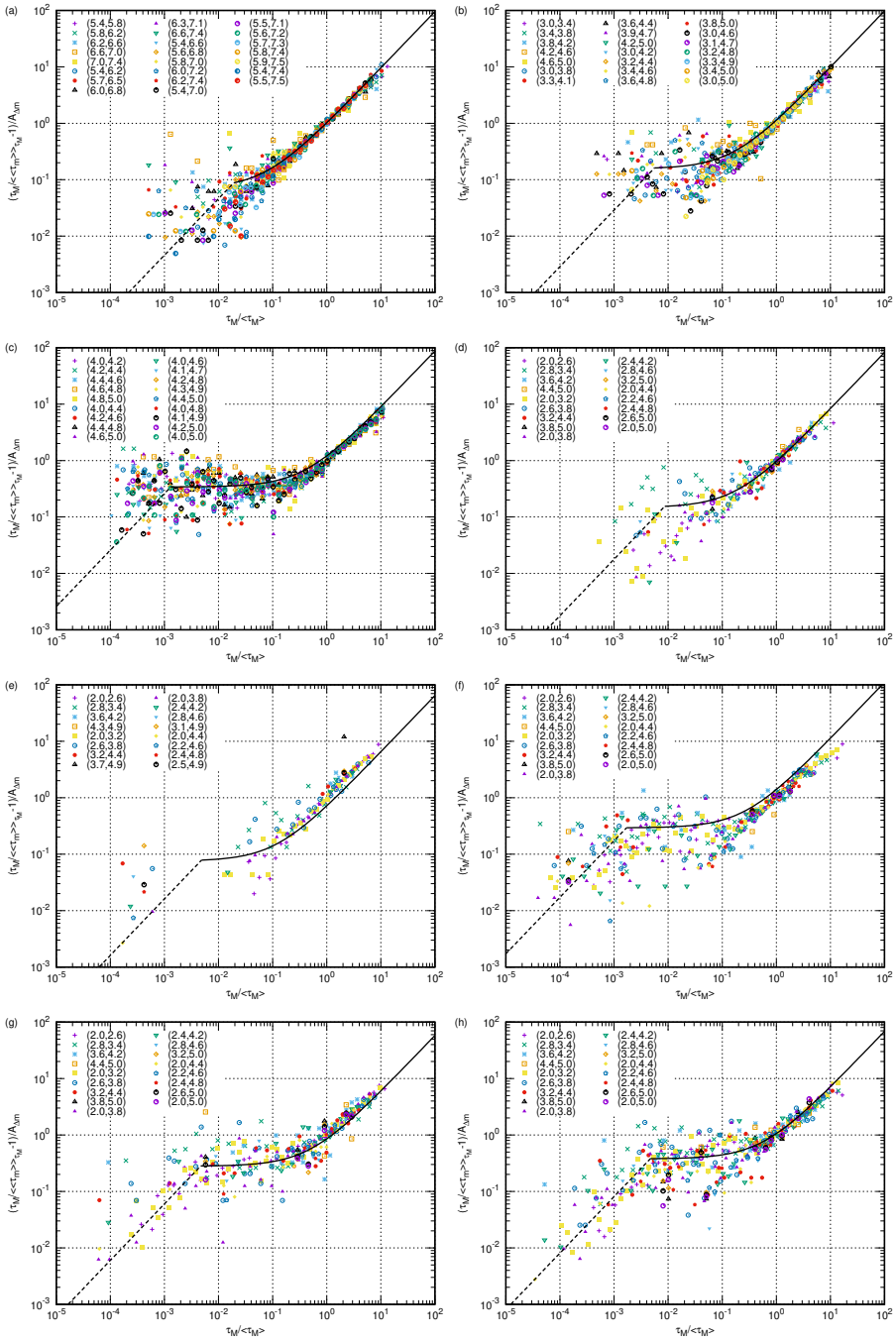

Figure 5.8: Results of rescaled $[\tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M} - 1]$ for some pairs of specific cut-off magnitudes (m, M) for (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. Equation (5.13) is also indicated by a black (dotted) line for reference.

inter-event intervals at the cut-off magnitude m_j included in these time intervals at the cut-off magnitude $m_j + \Delta m$. Then the result for Δm is obtained by plotting the following against $10^{\sigma+0.5\Delta\sigma}$:

$$\frac{1}{A_{\Delta m}} \left(\frac{\sum_{j=0}^{n} \hat{N}_{m_j}(\sigma)}{\sum_{j=0}^{n} \hat{N}_{m_j+\Delta m}(\sigma)} - 1 \right). \tag{5.12}$$

Such results for all possible Δm set within the magnitude range in Table 5.1 are shown in Fig. 5.9 by gray + symbols. In particular, the results for several Δm values are indicated by colored symbols. These results are nearly the same as those for the specific pairs of (m, M) in Fig.5.8, indicating that the scaling property in Eq. (5.11) holds. The results also suggest that the method using the dependence on the magnitude difference as in Eq. (5.12) can effectively extract the scaling function from the limited number of seismic data, as long as the GR law holds.

The following linear functions are used as the fitting function of $\zeta(y)$:

$$\zeta(y) \approx \begin{cases} \alpha y + \beta & (\text{if } y \geq y_0), \\ \gamma y & (\text{if } y < y_0). \end{cases} \tag{5.13}$$

The parameter values $(\alpha, \beta, \gamma, y_0)$ in Eq. (5.13) are different for each time series and can be determined as follows. First, β was determined by fitting in the range $y \in [10^{-2}, 10^{-1}]$. Next, as $\zeta(y) \sim \alpha y$ for large y, the author determined α by fitting in log-scale $\log_{10} \zeta(y) \sim \log_{10} \alpha + \log_{10} y$, in the range $y \geq 1$. In the same way, γ was determined by fitting in log-scale $\log_{10} \zeta(y) \sim \log_{10} \gamma + \log_{10} y$, in the range $y \leq 10^{-2}$. Lastly, y_0 was determined as the intersection of the above two lines. The resulting parameter values are summarized in Table 5.3, and the fitting curves with these parameter values are shown in Fig. 5.9 (and also in Fig. 5.8 for reference).

Table 5.3: Parameter values (to six decimal places) in Eq. (5.13) for each time series.

Name	α	β	γ	y_0
CMT	0.953828	0.069634	4.603897	0.019077
JS	0.979298	0.156119	28.882991	0.005595
JA	0.869569	0.340167	258.752423	0.001319
SCA1	0.822337	0.147506	17.838524	0.008669
SCA2	0.644104	0.075412	16.157159	0.004861
SCA3	1.106579	0.292309	172.511206	0.001705
SCS1	0.575125	0.279903	60.274019	0.004689
SCS2	0.681038	0.377327	81.000735	0.004698

Functional Form of $\Psi_{mM}(i|\tau_M)$

Figure 5.10 shows the component $\Psi_{mM}(i|\tau_M)$ for $i \leq 5$ and $\Delta m = 0.5$, 1.0, and 1.5 for the case of CMT. The cases of other time series are shown in Figs. S1-(1) – S1-(7). In the figure, the variable τ_M is scaled by the average $\langle \tau_M \rangle$. Only the small Δm results are shown because, when Δm becomes large, the number of samples for calculating the distribution is reduced, and it seemed insufficient to obtain clear outcomes. The results in Fig. 5.10 were obtained in the following way.

Let $\tilde{N}_M(i|\sigma_M)$ be the number of transformed time intervals at the upper cut-off magnitude ($\sigma_M = \log_{10} \tau_M$), whose length is within $[\sigma_M, \sigma_M + \Delta\sigma)$, and each of them includes i pieces of inter-event intervals at the lower cut-off magnitude. $\Psi_{mM}(i|\tau_M)$ can be drawn by plotting the following against $10^{\sigma_M+0.5\Delta\sigma}$:

$$\frac{\tilde{N}_M(i|\sigma_M)}{\tilde{N}_M(\sigma_M)}. \tag{5.14}$$

The results shown in Figs. 5.10 and S1-(1) – S1-(7) were obtained by further scaling the x-axis value by the average, that is, plotting Eq. (5.14) against $10^{\sigma_M+0.5\Delta\sigma}/\langle \tau_M \rangle$. In the figures, the results for several

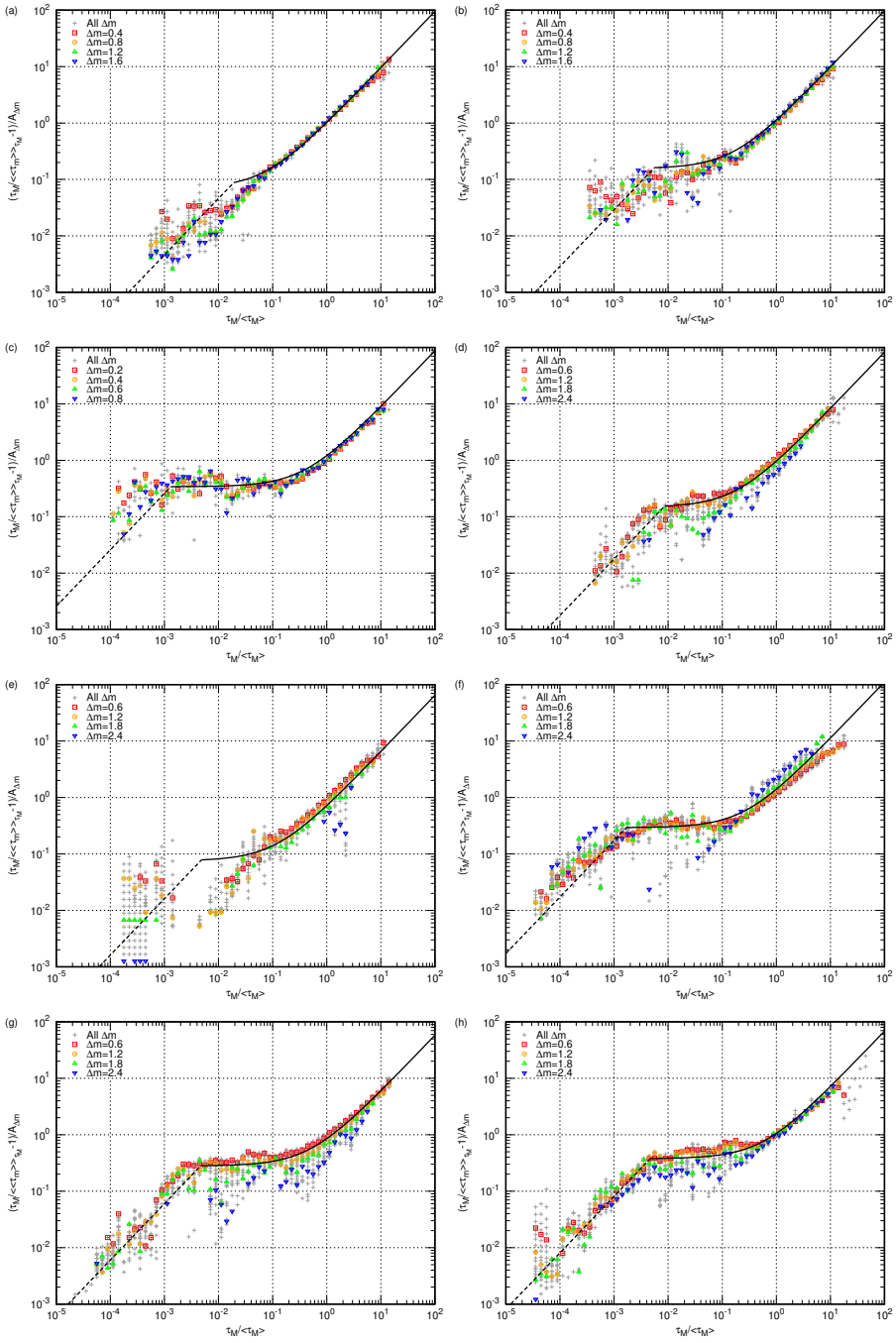

Figure 5.9: Results of $[\tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M} - 1]$ calculated by Eq. (5.12) for the mixed populations of (m, M) with the same Δm. Results of all Δm are shown by gray + symbols. The black solid line shows $\alpha y + \beta$ for $y \geq y_0$ and the black dotted line γy for $y < y_0$ with the parameter values in Table 5.3.

pairs of (m, M) are shown. The figures suggest that the probabilities $\Psi_{mM}(i|\tau_M)$ of the same Δm and i collapse on nearly the same function $(\hat{\Psi}_{\Delta m}(i|y))$ after rescaling by $\langle\tau_M\rangle$:

$$\Psi_{mM}(i|\tau_M) = \hat{\Psi}_{\Delta m}(i|y)\,. \tag{5.15}$$

Based on this scaling property, the author calculated $\hat{\Psi}_{\Delta m}(i|y)$ by taking into consideration all the sample data with the same Δm and y, likewise Section 5.3.1. Let $\hat{N}_{m_j+\Delta m}(i|\sigma)$ be the number of rescaled and transformed time intervals ($\sigma = \log_{10} y$) of length within $[\sigma, \sigma + \Delta\sigma)$ at the upper cut-off magnitude $(m_j + \Delta m)$, each of which includes i pieces of inter-event intervals at the lower cut-off magnitude m_j. $\hat{\Psi}_{\Delta m}(i|y)$ was also drawn in Figs. 5.10 and S1-(1) – S1-(7) by plotting the following against $10^{\sigma+0.5\Delta\sigma}$ by + symbols:

$$\frac{\sum_{j=1}^{n} \hat{N}_{m_j+\Delta m}(i|\sigma)}{\sum_{j=1}^{n} \hat{N}_{m_j+\Delta m}(\sigma)}. \tag{5.16}$$

The results of Eq. (5.16) resemble those for particular combinations of (m, M) and thus support the scaling property in Eq. (5.15).

The figures also indicate that $\Psi_{mM}(i|\tau_M)(=\hat{\Psi}_{\Delta m}(i|y))$ can be described empirically by the following negative binomial distribution with the scaling function $\zeta(y)$:

$$\Psi_{mM}(i|\tau_M) = \frac{\Gamma(i-1+B_{\Delta m}\zeta(y))}{(i-1)!\Gamma(B_{\Delta m}\zeta(y))}\left(\frac{B_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{i-1}, \tag{5.17}$$

where,

$$B_{\Delta m} = \frac{A_{\Delta m}}{b\Delta m\,(b\Delta m + 2)}.$$

In the figures, Eq. (5.17) was drawn with the b-values listed in Table 5.1 and the parameter values listed in Table 5.3. One can see that Eq. (5.17) describes the numerical results of $\Psi_{mM}(i|\tau_M)$, though it tends to deviate at the tail.

Note that Eq. (5.17) becomes the following Poisson distribution as $\Delta m \to 0$:

$$\Psi_{mM}(i|\tau_M) \xrightarrow[\Delta m\to 0]{} \frac{(A_{\Delta m}\zeta(y))^{(i-1)}}{(i-1)!}e^{-A_{\Delta m}\zeta(y)}. \tag{5.18}$$

In particular, when $\zeta(y) = y$ is identical to Eq. (3.11). However, such consistency does not hold for $\Delta m \gneqq 0$, and there is room for improvement in this respect.

Thus, the weak inter-event correlations cause two modifications in $\Psi_{mM}(i|\tau_M)$ compared to the stationary Poisson process. First, the distribution form of $\Psi_{mM}(i|\tau_M)$ changes to a negative binomial distribution for $\Delta m > 0$. Second, the scaling function $\zeta(y)$ deviates from $\zeta(y) = y$. In particular, when $\Delta m \to 0$, $\zeta(y)$ represents a perturbation from the stationary Poisson process caused by the weak correlations.

5.3.2 Results for the Component $\rho_{mM}(\tau_m|2, \tau_M)$ and the Assumption Regarding $\rho_{mM}(\tau_m|i, \tau_M)$ with $i \geq 3$

As for the component $\rho_{mM}(\tau_m|i, \tau_M)$, the author quantifies its change from Eq. (3.14) by using $dh^{(i)}_{mM}(\tau_m, \tau_M)$ defined as in Eq. (5.19).

$$\rho_{mM}(\tau_m|i, \tau_M) = \frac{(i-1)}{\tau_M}\left(1 - \frac{\tau_m}{\tau_M}\right)^{i-2}\left(1 + dh^{(i)}_{mM}(\tau_m, \tau_M)\right)\theta(\tau_M - \tau_m). \tag{5.19}$$

Note that, as pointed out in Section 5.3.1, the average of the conditional probability is determined only by the component $\Psi_{mM}(i|\tau_M)$. Therefore, the changes $dh^{(i)}_{mM}(\tau_m|\tau_M)$ affect the higher order moments than 1 in the conditional probability.

In this study, the author examined only the case of $i = 2$. By the normalization condition and the symmetry of $\rho_{mM}(\tau_m|2, \tau_M)$, $dh^{(2)}_{mM}(\tau_m, \tau_M)$ must satisfy the followings:

$$\int_0^{\tau_M} dh^{(2)}_{mM}(\tau_m, \tau_M)d\tau_m = 0,$$

$$dh^{(2)}_{mM}(\tau_m, \tau_M) = dh^{(2)}_{mM}(\tau_M - \tau_m, \tau_M).$$

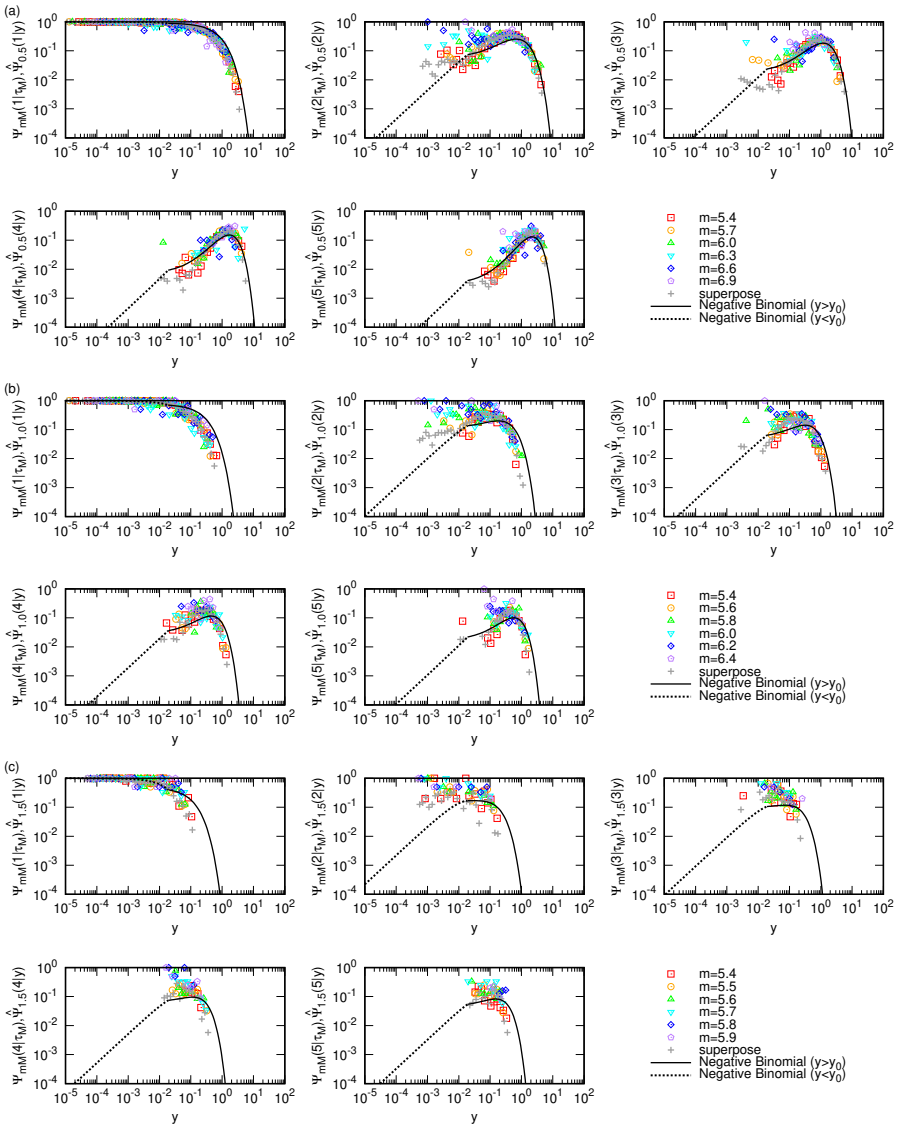

Figure 5.10: $\Psi_{mM}(i|\tau_M)$ and $\hat{\Psi}_{\Delta m}(i|y)$ with $i \leq 5$ for CMT, plotted against $y(:= \tau_M/\langle\tau_M\rangle)$. The cases of (a) $\Delta m = 0.5$, (b) $\Delta m = 1.0$, and (c) $\Delta m = 1.5$ are shown. The results of $\Psi_{mM}(i|\tau_M)$ for several specific pairs of cut-off magnitude (m, M) are shown by open-colored symbols. The gray + symbols show $\hat{\Psi}_{\Delta m}(i|y)$ for mixed populations of (m, M) with the same Δm. (Dotted) curves show Eq. (5.17) with the parameter values listed in Tables 5.1 and 5.3.

The author also examined the distribution function of $\rho_{mM}(\tau_m|2,\tau_M)$:

$$P_{mM}(\tau_m|2,\tau_M) := \int_0^{\tau_m} \rho_{mM}(s|2,\tau_M)ds,$$

which is τ_m/τ_M for the stationary marked Poisson process.

Figure 5.11 shows the numerical results of $\rho_{mM}(\tau_m|2,\tau_M)$ and $P_{mM}(\tau_m|2,\tau_M)$ for CMT. The latter was scaled by τ_M values in the x-axis direction. The results for other time series are shown in Figs. S2-(1) – S2-(7). The results were obtained in the following way.

Define the transformed lower interval $\sigma_m := \log_{10}\tau_m$. Let $\tilde{N}_{mM}(\sigma_m;\sigma_M|i=2)$ be the number of the transformed inter-event time intervals at the lower cut-off magnitude, whose length falls within $[\sigma_m, \sigma_m+\Delta\sigma)$ and each of them is included in the transformed inter-event time interval at the upper cut-off magnitude ($\sigma_M = \log_{10}\tau_M$) such that its length is within $[\sigma_M, \sigma_M+\Delta\sigma)$ and it includes only two ($i=2$) inter-event intervals at the lower cut-off magnitude. Also, let $\tilde{N}_m(\sigma_M|i=2)$ be the total number of inter-event intervals at the lower cut-off magnitude, included in the inter-event intervals at the upper cut-off magnitude such that its transformed length is within $[\sigma_M, \sigma_M+\Delta\sigma)$ and includes two intervals at the lower cut-off magnitude. Finally, I denote the density $\rho_{mM}(\tau_m|2,\tau_M)$ after applying the above variable transformations by $\tilde{\rho}_{mM}(\sigma_m|2,\sigma_M)$, which is calculated as follows:

$$\tilde{\rho}_{mM}(\sigma_m|2,\sigma_M) = \frac{\tilde{N}_{mM}(\sigma_m;\sigma_M|i=2)}{\tilde{N}_m(\sigma_M|i=2)\Delta s}. \tag{5.20}$$

$\rho_{mM}(\tau_m|2,\tau_M)$ is obtained by re-transforming this by $\sigma_m \mapsto \tau_m$ and $\sigma_M \mapsto \tau_M$,

$$\rho_{mM}\left(\tau_m|2,\tau_M\right) = \frac{1}{10^{\sigma_m+0.5\Delta s}\ln 10}\frac{\tilde{N}_{mM}(\sigma_m;\sigma_M|i=2)}{\tilde{N}_m(\sigma_M|i=2)\Delta s}. \tag{5.21}$$

$\rho_{mM}(\tau_m|2,\tau_M)$ in Figs 5.11 and S2-(1) – S2-(7) were drawn by plotting Eq. (5.21) against $10^{\sigma_m+0.5\Delta s}$ and $10^{\sigma_M+0.5\Delta s}$. Also, $P_{mM}(\tau_m|2,\tau_M)$ can be calculated by adding $\tilde{\rho}_{mM}\left(\sigma_m|2,\sigma_M\right)\Delta s$ and was plotted against $10^{\sigma_m+\Delta s}/10^{\sigma_M+0.5\Delta s}$ and $10^{\sigma_M+0.5\Delta s}$.

As shown in these figures, the densities $\rho_{mM}\left(\tau_m|2,\tau_M\right)$ are nearly the same regardless of (m,M). In addition, $\rho_{mM}\left(\tau_m|2,\tau_M\right) \approx 1/\tau_M$, suggesting that $dh^{(2)}_{mM}(\tau_m,\tau_M) \approx 0$. However, as shown in the lower panels of the figures, $dh^{(2)}_{mM}(\tau_m,\tau_M)$ tends to take a positive value for small τ_m. This tendency seems to reflect the temporal clustering of moderate aftershocks in stationary time series and secondary aftershocks in transformed aftershock sequences.

$dh^{(2)}(\tau_m,\tau_M) \approx 0$ implied by the figures reflects the weak inter-event correlations in the time series; the situation seems to be the same for $i \geq 3$, that is, $dh^{(i)}_{mM}(\tau_m,\tau_M) \approx 0$ for $i \geq 3$. Thus, in the following, the author assumes that inter-event correlations in these time series are sufficiently weak so as not to affect the components $\rho_{mM}(\tau_m|i,\tau_M)$ and there is no change in $\rho_{mM}(\tau_m|i,\tau_M)$ (i.e., $dh^{(i)}_{mM}(\tau_m,\tau_M) \approx 0$ for $i \geq 2$) from the stationary marked Poisson process. Note that this assumption should be re-considered when dealing with time series that include prominent (secondary) aftershocks.

5.4 Derivation of the Probability Density Functions Related to Inter-event Times

The analysis in the previous section yielded the following results and suggestions for unfolding-transformed seismic time series: the component $\Psi_{mM}(i|\tau_M)$ can be empirically described by the negative binomial distribution (Eq. (5.17)) with the scaling function $\zeta(y)$ (which can be fit by the linear functions in Eq. (5.13)) in the newly found scaling property (Eq. (5.11)), and the other component $\rho_{mM}(\tau_m|i,\tau_M)$ is almost unaffected by weak inter-event correlations and assumed to be the same as in the stationary Poisson process. Based on them, this section derives the functional forms of probability densities in Bayes' theorem (Eq. (3.7)) and checks whether the derived functions can describe the probability densities.

5.4.1 Conditional Probability

First, the conditional probability is derived. The conditional probability rescaled by the transformations $\tau_m \mapsto x := \tau_m/\langle\tau_M\rangle$ and $\tau_M \mapsto y$ ($\hat{p}_{mM}(x|y)$) is derived as in Eq. (5.22) (see Appendix N). Hereafter,

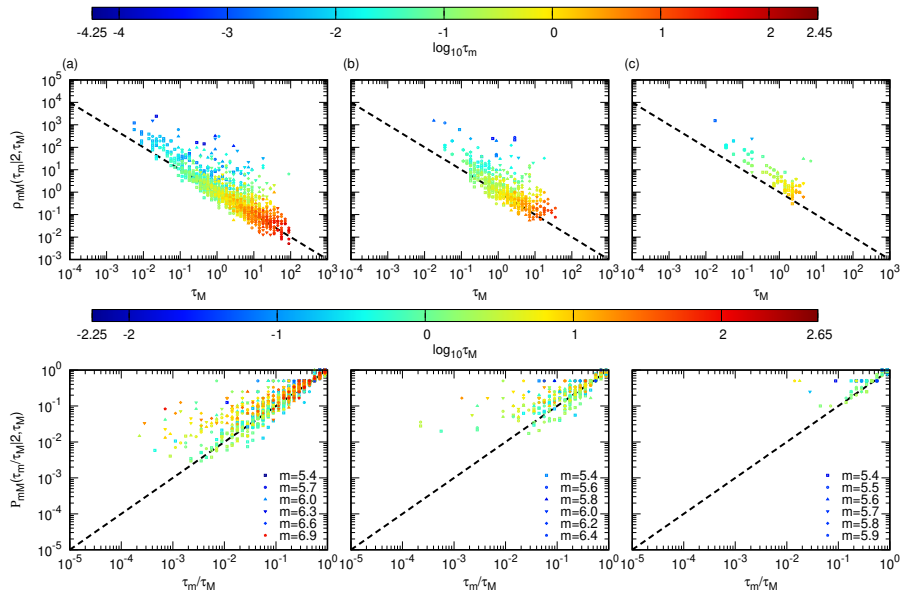

Figure 5.11: Numerical results of (upper panels) $\rho_{mM}(\tau_m|2,\tau_M)$ and (lower panels) $P_{mM}(\tau_m|2,\tau_M)$ for several pairs of (m,M) with (a) $\Delta m = 0.5$, (b) $\Delta m = 1.0$, and (c) $\Delta m = 1.5$, for CMT. The upper (lower) panels show $\rho_{mM}(\tau_m|2,\tau_M)$ ($P_{mM}(\tau_m|2,\tau_M)$) in a cross-section parallel to the τ_M (τ_m/τ_M) axis. The colors in the upper (lower) panels indicate τ_m (τ_M) values. The dotted black line in the upper (lower) panels indicates $\rho_{mM}(\tau_m|2,\tau_M) = 1/\tau_M$ ($P_{mM}(\tau_m|2,\tau_M) = \tau_m/\tau_M$), which corresponds to $dh^{(2)}_{mM}(\tau_m,\tau_M) = 0$ in Eq. (5.19).

variable x is used with this definition.

$$\hat{p}_{mM}(x|y) = (A_{\Delta m}\zeta(y)+1)^{-1}\left\{\delta(y-x)\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\right.$$
$$+\theta(y-x)\frac{[A_{\Delta m}B_{\Delta m}(y-x)\zeta(y)+A_{\Delta m}(x+y)+2B_{\Delta m}y]}{(A_{\Delta m}x+B_{\Delta m}y)^2}$$
$$\left.\times A_{\Delta m}B_{\Delta m}\zeta(y)\left(\frac{B_{\Delta m}y}{A_{\Delta m}x+B_{\Delta m}y}\right)^{B_{\Delta m}\zeta(y)}\right\}. \tag{5.22}$$

As Eq. (5.22) indicates, the rescaled conditional probability depends on Δm.

To compare with Eq. (5.22), the author numerically obtained the conditional probability similarly to Section 5.3.2, though the condition of $i = 2$ is removed. Let $\tilde{N}_{mM}(\sigma_m;\sigma_M)$ be the number of the transformed inter-event time intervals at the lower cut-off magnitude ($\sigma_m = \log_{10}\tau_m$), whose length is within $[\sigma_m, \sigma_m+\Delta\sigma)$ and each of them is included in the transformed inter-event time interval at the upper cut-off magnitude ($\sigma_M = \log_{10}\tau_M$) of length within $[\sigma_M, \sigma_M+\Delta\sigma)$. This represents the total number of pairs of σ_m and σ_M in the time series. Also, let $\tilde{p}_{mM}(\sigma_m|\sigma_M)$ be the probability density of σ_m under the condition that it is included in the inter-event time interval of transformed length σ_M at the upper cut-off magnitude. Thus, $\tilde{p}_{mM}(\sigma_m|\sigma_M)$ is calculated as follows:

$$\tilde{p}_{mM}(\sigma_m|\sigma_M) = \frac{\tilde{N}_{mM}(\sigma_m;\sigma_M)}{\tilde{N}_m(\sigma_M)\Delta\sigma},$$

and $p_{mM}(\tau_m|\tau_M)$ is:

$$p_{mM}(\tau_m|\tau_M) = \frac{1}{10^{\sigma_m+0.5\Delta\sigma}\ln 10}\frac{\tilde{N}_{mM}(\sigma_m;\sigma_M)}{\tilde{N}_m(\sigma_M)\Delta\sigma}. \tag{5.23}$$

$\hat{p}_{mM}(x|y)$ can be drawn by plotting the following against (x,y):

$$\hat{p}_{mM}(x|y) = \langle\tau_M\rangle p_{mM}(\tau_m|\tau_M). \tag{5.24}$$

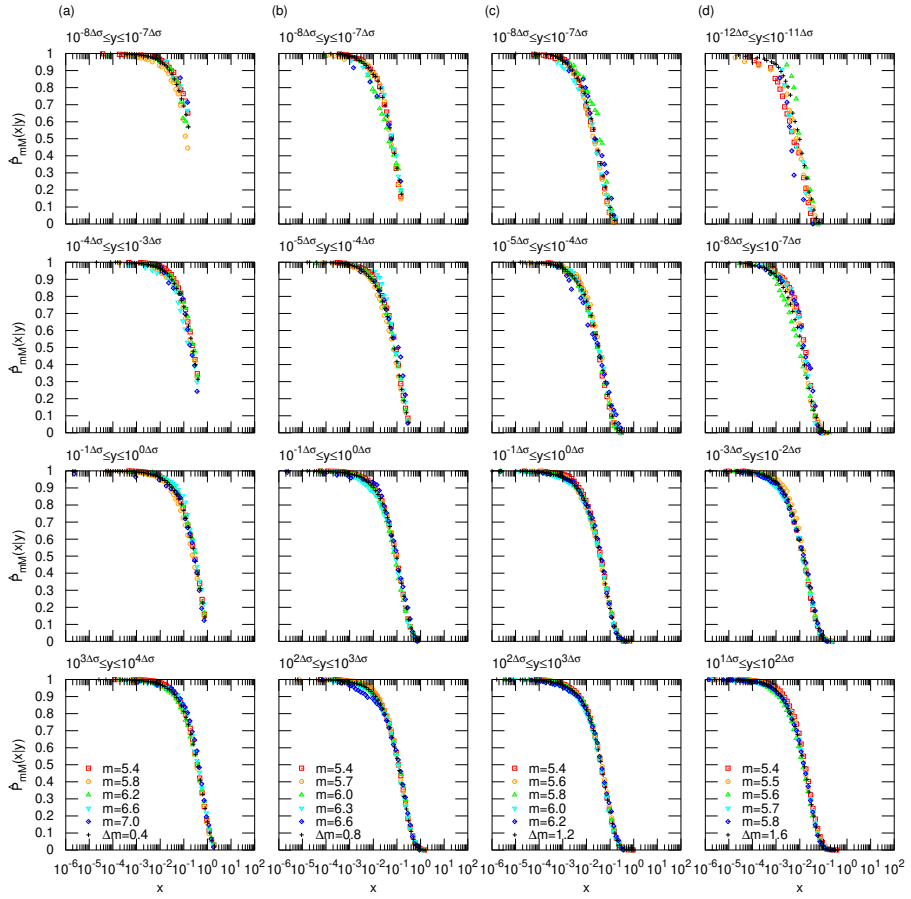

Figure 5.12: $\hat{P}_{mM}(x|y)$ for the time series of CMT. Results for certain pairs of cut-off magnitudes (m, M) with (a) $\Delta m = 0.4$, (b) $\Delta m = 0.8$, (c) $\Delta m = 1.2$, and (d) $\Delta m = 1.6$, are drawn in cross-sections for the ranges of y shown in the figure. Results for the pairs of (m, M) are represented by colored open symbols. Black + symbols show $\overline{P}_{\Delta m}(x|y)$ where $y = 10^{(j+0.5)\Delta\sigma}$, $j \in \mathbb{Z}$. Among the j values for which $\overline{P}_{\Delta m}(x|y)$ is obtained, the values are denoted such that all results for the pairs of (m, M) displayed in the legend are included within $(10^{j\Delta\sigma}, 10^{(j+1)\Delta\sigma})$ by $j = j_0, j_1, \cdots, j_K$. The ranges of y shown in the figure are $(10^{j\Delta\sigma}, 10^{(j+1)\Delta\sigma})$ for j values nearest to the four points which are dividing $[j_0, j_K]$ into five equal parts.

Figure 5.12 shows $\hat{P}_{mM}(x|y)$ defined below:

$$\hat{P}_{mM}(x|y) := \int_x^\infty \hat{p}_{mM}(x'|y)dx',$$

numerically calculated for the time series of CMT. In the figure, $\hat{P}_{mM}(x|y)$ is shown in cross-section views for various y ranges. The figure suggests that the conditional probability for the same Δm collapse on the same function, that is,

$$\hat{p}_{mM}(x|y) = \hat{p}_{\Delta m}(x|y). \tag{5.25}$$

This is consistent with the implication by Eq. (5.22). The same scaling property was also found for other time series, as shown in Figs. S3-(1) – S3-(7).

Based on the scaling property in Eq. (5.25), the author calculated the conditional probability using all the pairs of cut-off magnitudes with the same Δm. Define the rescaled and transformed interval $\varsigma := \log_{10} x$, and $\sigma = \log_{10} y$. Let $\hat{N}_{m_j, m_j+\Delta m}(\varsigma; \sigma)$ be the total number of the inter-event time intervals of rescaled and transformed length within $[\varsigma, \varsigma + \Delta\sigma)$ at the cut-off magnitude m_j, that is included in the inter-event time intervals of rescaled and transformed length within $[\sigma, \sigma + \Delta\sigma)$ at the cut-off magnitude

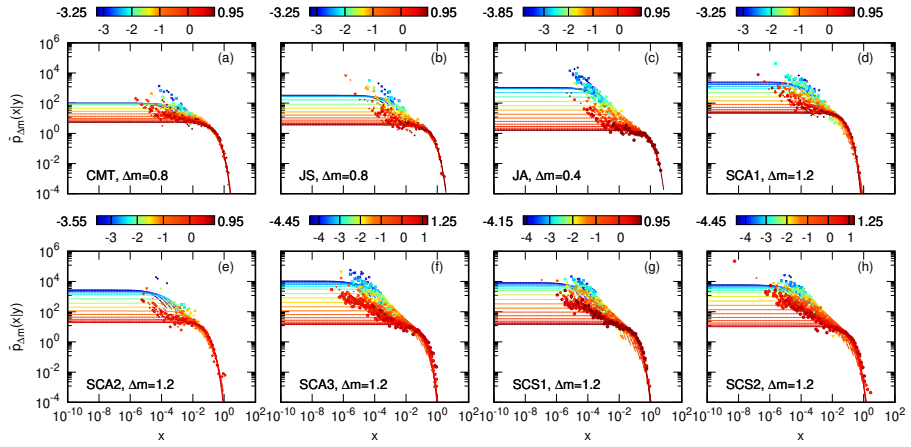

Figure 5.13: Equation (5.22) calculated with the parameter values from Table 5.3 (curves) and numerical results of $\overline{p}_{\Delta m}(x|y)$ by catalog analysis (symbols) for time series of (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. The value of Δm is shown in the figure. Results are drawn in cross-sections for $y = 10^{(3j+0.5)\Delta\sigma}$, $j \in \mathbb{Z}$. The color corresponds to the $\log_{10} y$ value, and its range is shown in the color bar. Note that when the result of catalog analysis is not obtained for a j, only the corresponding Eq. (5.22) is shown.

$m_j + \Delta m$. The conditional probability for Δm ($\overline{p}_{\Delta m}(x|y)$) is calculated as follows:

$$\overline{p}_{\Delta m}(x|y) = \frac{1}{10^{\varsigma+0.5\Delta\sigma}\ln 10}\frac{\sum_{j=0}^{n}\hat{N}_{m_j,m_j+\Delta m}(\varsigma;\sigma)}{\sum_{j=0}^{n}\hat{N}_{m_j}(\sigma)\Delta\sigma}. \tag{5.26}$$

In Fig. 5.12, also $\overline{P}_{\Delta m}(x|y)$ defined as follows is shown.

$$\overline{P}_{\Delta m}(x|y) := \int_x^\infty \overline{p}_{\Delta m}(x'|y)dx'.$$

$\overline{P}_{\Delta m}(x|y)$ is almost consistent with $\hat{P}_{mM}(x|y)$, supporting the scaling property in Eq. (5.25) and the superposition method in Eq. (5.26).

Figure 5.13 shows the results of Eq. (5.26) plotted against $10^{\varsigma+0.5\Delta\sigma}$ and Eq. (5.22) with the parameter values listed in Table 5.3 at a given value of Δm. The results for other Δm values are summarized in Figs. S4-(1) – S4-(8). The figures show that the results of catalog analyses are described by Eq. (5.22).

5.4.2 Inter-event Time Distribution

Second, the scaling function ($f(y)$ in Eq. (1.8)) of the inter-event time distribution is derived. Under the assumption that $dh^{(i)}_{mM}(\tau_m, \tau_M) = 0$ for $i \geq 2$, the scaling functions $f(y)$ and $\zeta(y)$ are shown to be equivalent, and using Eq. (5.13) (see Appendix O),

$$f(y) = \begin{cases} \left(\dfrac{y_0^\beta e^{\alpha y_0}}{\frac{e^{\gamma y_0}-1}{\gamma} + \frac{y_0^\beta e^{\alpha y_0}}{\alpha^{1-\beta}}\Gamma(1-\beta, \alpha y_0)}\right) y^{-\beta}e^{-\alpha y} & (\text{for } y \geq y_0), \\ \left(\dfrac{e^{\gamma y_0}}{\frac{e^{\gamma y_0}-1}{\gamma} + \frac{y_0^\beta e^{\alpha y_0}}{\alpha^{1-\beta}}\Gamma(1-\beta, \alpha y_0)}\right) e^{-\gamma y} & (\text{for } y < y_0). \end{cases} \tag{5.27}$$

Here $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function.

Figure 5.14 compares Eq. (5.27) with the rescaled inter-event time distributions for each time series. Equation (5.27) was drawn using the parameter values in Table 5.3.

Equation (5.27) appears to be relatively complicated. However, Fig. 5.14 also shows that the overall shape, except for very short intervals, of the scaled distribution can be described by the gamma

distribution in Eq. (5.28) that is the extension of $f(y)$ in Eq. (5.27) for $y \geq y_0$ to $y \in [0, \infty)$ with re-normalization:

$$f(y) = \frac{\alpha^{1-\beta}}{\Gamma(1-\beta)} y^{-\beta} e^{-\alpha y}. \tag{5.28}$$

5.4.3 Inverse Probability

Third, the inverse probability is derived by substituting the results in Sections. 5.4.1 and 5.4.2 into Bayes' theorem (Eq. (3.7)). The rescaled inverse probability ($\hat{p}_{Mm}(y|x)$) was derived as follows (see Appendix P):

$$\begin{aligned}\hat{p}_{Mm}(y|x) = \frac{\left(\frac{y}{x}\right)^{-\beta} e^{-\alpha(y-(A_{\Delta m}+1)x)}}{(A_{\Delta m}+1)^{2-\beta}} \Bigg\{ \delta(y-x) \left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)} \\ +\theta(y-x) \frac{[A_{\Delta m}B_{\Delta m}(y-x)\zeta(y) + A_{\Delta m}(x+y) + 2B_{\Delta m}y]}{(A_{\Delta m}x + B_{\Delta m}y)^2} \\ \times A_{\Delta m}B_{\Delta m}\zeta(y) \left(\frac{B_{\Delta m}y}{A_{\Delta m}x + B_{\Delta m}y}\right)^{B_{\Delta m}\zeta(y)} \Bigg\}.\end{aligned} \tag{5.29}$$

As with the conditional probability, $\hat{p}_{Mm}(y(>x)|x)$ depends on Δm.

As in Section 5.4.1, the author numerically computed the inverse probability from catalog data as follows. Let $\tilde{p}_{Mm}(\sigma_M|\sigma_m)$ be the probability density of the transformed inter-event time interval at the upper cut-off magnitude ($\sigma_M = \log_{10}\tau_M$) under the condition that inside of which the transformed time interval of length $\sigma_m (= \log_{10}\tau_m)$ is found. Thus, $\tilde{p}_{Mm}(\sigma_M|\sigma_m)$ is calculated as follows:

$$\tilde{p}_{Mm}(\sigma_M|\sigma_m) = \frac{\tilde{N}_m(\sigma_m;\sigma_M)}{\tilde{N}_m(\sigma_m)\Delta\sigma},$$

and $p_{Mm}(\tau_M|\tau_m)$ is:

$$p_{Mm}(\tau_M|\tau_m) = \frac{1}{10^{\sigma_M+0.5\Delta\sigma}\ln 10} \frac{\tilde{N}_m(\sigma_m;\sigma_M)}{\tilde{N}_m(\sigma_m)\Delta\sigma}. \tag{5.30}$$

Thus the scaling of Eq. (5.30) yields $\hat{p}_{Mm}(y|x)$:

$$\hat{p}_{Mm}(y|x) = \langle\tau_M\rangle p_{Mm}(\tau_M|\tau_m), \tag{5.31}$$

and,

$$\hat{P}_{Mm}(y|x) := \int_y^\infty \hat{p}_{Mm}(y'|x)dy',$$

is shown for the case of CMT in Fig. 5.15 as sectional diagrams for several ranges of x. The results for other time series are shown in Figs. S5-(1) – S5-(7). The results suggest the property:

$$\hat{p}_{Mm}(y|x) = \hat{p}_{\Delta m}(y|x), \tag{5.32}$$

as implied by Eq. (5.29).

As the Δm dependence was confirmed, I calculated $\hat{p}_{\Delta m}(y|x)$ taking into consideration all the pair of (m, M) with equal Δm as follows:

$$\overline{p}_{\Delta m}(y|x) = \frac{1}{10^{\sigma+0.5\Delta\sigma}\ln 10} \frac{\sum_{j=0}^n \hat{N}_{m_j,m_j+\Delta m}(\varsigma;\sigma)}{\sum_{j=0}^n \hat{N}_{m_j}(\varsigma)\Delta\sigma}. \tag{5.33}$$

Figure 5.15 shows $\overline{P}_{\Delta m}(y|x) := \int_y^\infty \overline{p}_{\Delta m}(y'|x)dy'$ and its consistency with $\hat{P}_{Mm}(y|x)$ supports Eqs. (5.32) and (5.33).

The results of Eq. (5.33) are shown in Fig. 5.16 against $10^{\sigma+0.5\Delta\sigma}$, along with Eq. (5.29) substituting the parameter values in Table 5.3. Figure 5.16 shows only the results for a particular Δm; see Figs. S6-(1) – S6-(8) for the results at other Δm values. It can be seen that Eq. (5.29) illustrates the result of catalog analysis for individual time series.

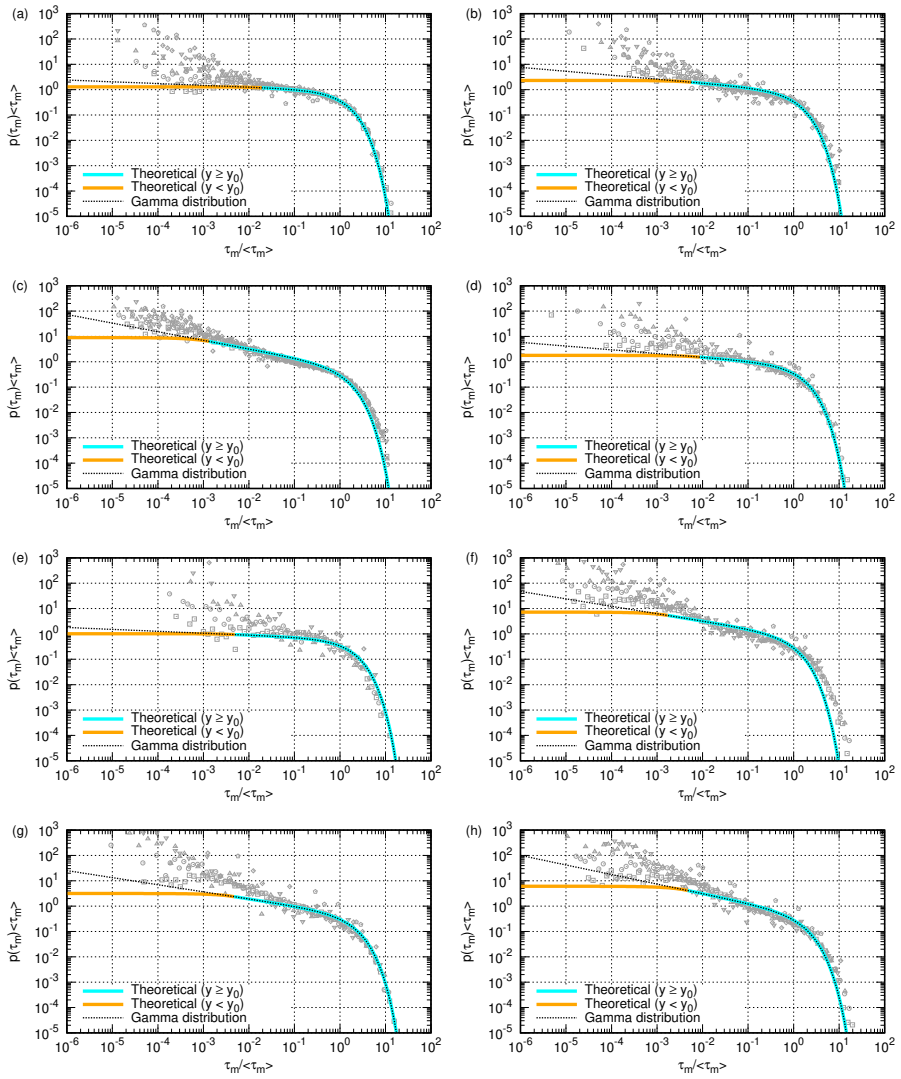

Figure 5.14: Equations (5.27) (orange curve for $y < y_0$ and cyan curve for $y \geq y_0$) and (5.28) (black dotted curve) as well as numerical results of the rescaled inter-event time distributions by catalog analysis (gray symbols, each symbol represents a result at the same cut-off magnitude as in Fig. 5.7) for time series of (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. Equations (5.27) and (5.28) are calculated using the parameter values listed in Table 5.3.

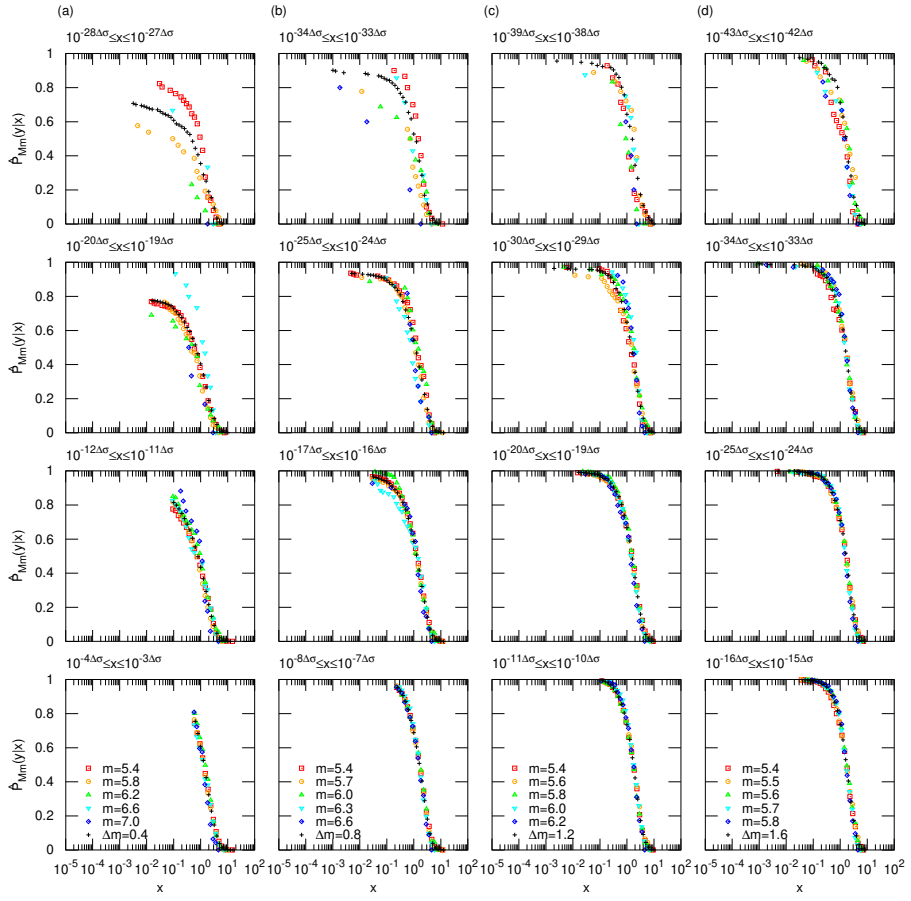

Figure 5.15: $\hat{P}_{Mm}(y|x)$ for the time series of CMT. Results for some pairs of (m, M) with (a) $\Delta m = 0.4$, (b) $\Delta m = 0.8$, (c) $\Delta m = 1.2$, and (d) $\Delta m = 1.6$ are shown by colored symbols in cross-sections in the range of x indicated in the figure. Also, $\overline{P}_{\Delta m}(y|x) := \int_y^\infty \overline{p}_{\Delta m}(y'|x)dy'$, where $x = 10^{(j+0.5)\Delta\sigma}$, $j \in \mathbb{Z}$ is shown by black + symbols. The x ranges indicated in the figure are determined by utilizing the same method as outlined in the caption of Fig. 5.12.

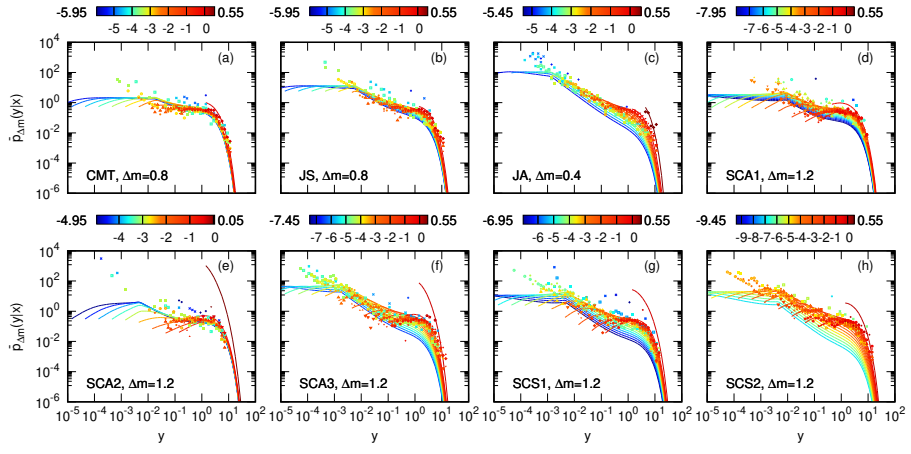

Figure 5.16: Equation (5.29) (curves) with the parameter values in Table 5.3 and numerical result of $\overline{p}_{Mm}(y|x)$ by catalog analysis (symbols) for time series of (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. The values of Δm are shown in the figure. The results are drawn in cross-section for $x = 10^{(5j+0.5)\Delta\sigma}$, where $j \in \mathbb{Z}$. The color corresponds to the value of $\log_{10} x$ and its range is shown in the color bar. Note the same possible absence of symbols as in Fig. 5.13.

5.5 Discussion and Conclusions

This chapter examined the Bayesian forecasting framework of earthquakes' timing using three seismic catalogs toward its practical use. Following Corral's method, the author extracted stationary (almost constant occurrence rate) and aftershock (obeying OU occurrence rate) time series from three seismic catalogs and transformed them into weak inter-event correlation ones. For these unfolding-transformed time series, the author examined the simplest form of the Bayesian approach, Bayes' theorem, composed of three distribution functions: the conditional probability to quantify temporal hierarchy in time series, the inverse probability directly related to forecasting, and the inter-event time distribution for which scaling universality was re-examined.

A new scaling property on the temporal hierarchy of seismic time series was found by analyzing the average of the conditional probability; the linear functions can fit its scaling function ($\zeta(y)$), though its fitting parameter values differ, reflecting seismic characteristics. The two functional components of the conditional probability were analyzed; one can be empirically described by the negative binomial distribution with the above scaling function ($\zeta(y)$), and the other is almost unchanged from that of the stationary marked Poisson process, which may be attributed to the weakness of inter-event correlation of the unfolding-transformed time series.

Based on them, functional forms of the conditional probability and the scaling function ($f(y)$) of the inter-event time distribution were derived. The theoretical derivation based on the present catalog analysis suggested that the scaling functions for these probability distributions ($\zeta(y)$ and $f(y)$) are essentially equivalent; the difference in the parameter values in the fitting linear functions of $\zeta(y)$ appears as that in the scaling function $f(y)$ of the inter-event time distribution, indicating the negative view of universality. Finally, the inverse probability was derived from Bayes' theorem. These derived distribution functions could describe the characteristics of the results of catalog analysis, such as the functional form and peak emergence in the inverse probability.

5.5.1 On the Scaling Universality in the Inter-event Time Distribution in Seismicity

The derived scaling function of the inter-event time distribution was described as an exponential distribution for small intervals ($y < y_0$) and a gamma distribution for large intervals ($y \geq y_0$). However, the overall scaling function could be described just by the re-normalized latter gamma distribution. The parameter values to characterize distributions are similar but different for each time series. These results are consistent with the conclusions of the preceding studies in the following respects.

First, in the preceding studies [72, 73, 77], the (generalized) gamma distribution was used to fit the results of catalog analysis, except for short intervals. Theoretical studies under some assumptions on correlations between events derived the distribution near the generalized gamma distribution [83, 84, 85, 104]. Equation (5.28) provides another theoretical basis for the fitting instead of such approaches with seismic model time series.

Second, theoretical studies using the ETAS model showed that the universality of the scaling function $f(y)$ is only approximate [83, 84, 85]. Our result is also consistent with them; the parameter values in Eq. (5.27) or (5.28) were similar but slightly different, reflecting the characteristics of seismic activity aggregated as the difference in $\zeta(y)$.

However, the difference in $f(y)$ for each time series was not as pronounced as in the preceding numerical study [78]. This is because the author restricted the analysis only to the stationary time series and transformed aftershock sequences, in accordance with the original Corral's method. The bimodality of the inter-event time distribution pointed out in that study [78] seems to be manifested in the switching of the distribution in Eq. (5.27), though it could not satisfactorily describe the catalog analysis at short intervals ($y < y_0$). A detailed study of $dh^{(i)}_{mM}(\tau_m, \tau_M)$, set to 0 in this study, may clarify this point.

These results on the scaling universality in the inter-event time distribution from the viewpoint of the unfolding transformation indicate the following: such temporally unfolded time series are indistinguishable regardless of original seismic activity to the extent that their temporal property can be described in the same gamma distribution with slightly different parameter values aggregating seismic character. This conclusion about the temporal nature of the renewal process at a certain cut-off magnitude was derived in the present study from that of the hierarchical relation at two different cut-off magnitudes, i.e., the new scaling property of the conditional probability. Thus, this study extends the above conclusion to

the level of the hierarchical structure of seismic time series; unfolding transformed seismic time series are indistinguishable even at the hierarchical level to the extent that the same functional form of the conditional probability can describe them.

5.5.2 On the Method of Unfolding Transformation of Seismic Time Series

The following problems still need to be solved with the method the author used in this chapter. Applying the unfolding procedure requires a smooth function representing the temporal variation trend of occurrence rate. In studies of quantum chaos, such functions are given theoretically [98]. However, this study does not have such a theoretical basis, and the temporal variation trends of occurrence rate are given empirically.

This remains the ambiguity in judging whether an aftershock sequence is moderate and thus regarded as fluctuations in a stationary time series or a clear aftershock sequence. Preceding studies using stochastic models considered all aftershock sequences to be the former; thus, there is no ambiguity in this respect. However, in our approach, a modest aftershock sequence can be regarded as a sequence such that its occurrence rate fluctuates around a smooth temporal variation described by the OU rate or just a fluctuation in a stationary time series; a deviation from a constant occurrence rate. For example, this study considered the secondary aftershocks in JA as the deviation in the unfolding-transformed time series, though they could be considered as other aftershock activities. The fitting parameter values of $\zeta(y)$ vary depending on which of the two is adopted, though at present, there is no clear way to determine.

In this way, the unfolding procedure applied in this study leaves some ambiguity in determining the smooth function to describe the trend of the occurrence rate; our conclusion on the universality is tentative in this respect. For a more detailed discussion, it is necessary to establish an objective method to determine the smooth function of the occurrence rate trend adaptable to seismic activity, such as using the moving average.

5.5.3 Inter-event Correlations Reflected in the Conditional Probability

The kind of inter-event correlations reflected in the conditional probability requires further examination. The temporal clustering represented by the OU law seems the primary one. However, preceding studies [40, 68] suggested other inter-event correlations; the conditional probability may reflect such correlations. Clarifying this point is important for the theoretical understanding of the Bayesian approach and the properties of the inter-event time distribution.

5.5.4 Property of the Inverse Probability and its Significance in Forecasting Earthquakes

This chapter derived the inverse probability using the GR law and the new scaling property on the hierarchical structure of seismic time series. Although the derivation was based not on analytically rigorous probabilistic calculations but on empirical fitting, assumption, and approximation, this advances the point estimate of the large-earthquake timing in time series with inter-event correlations in the previous study [2] in Chapter 3 to probabilistic forecasting.

The derived inverse probability has a peak when Δm is large. Such a feature enables us to narrow the timing of future large earthquakes better than the inter-event time distribution, the prior probability without such a peak. The inverse probability is shown to depend on the fitting parameters of $\zeta(y)$ and the b-value reflecting differences in seismic characteristics by time series. Such difference led to the negative conclusion on the scaling universality, though, on the other hand, it suggests the possibility of effective forecasting using the inverse probability. That is, the inverse probability can reflect the characteristics of seismic history in the space-time domain, including minor-scale ones, and varies depending on the length of the lower interval. Therefore, the inverse probability can provide better forecasting than the existing inter-event time distribution by reflecting additional information on the seismic history and the real-time occurrence of small events.

Chapter 3 suggested that the forecasting is relatively ineffective when the non-stationary Omori – Utsu relaxation process that causes inhibiting steady estimate is dominant in activity [2]. The unfolding procedure may improve forecasting in such a relaxation process. This study suggested that the stationary time series are indistinguishable from the unfolding-transformed aftershock sequences at the hierarchy

level, and thus, the author expects that the Bayesian inference in the transformed aftershock sequence improves to the same extent as other stationary time domains. Therefore, it is conceivable to apply the unfolding transformation to seismic time series before performing Bayesian inference. In such an approach, it is sufficient to consider a less daunting Bayesian framework for time series with weak inter-event correlations. However, in this approach, the already-mentioned problem of determining the smooth function of the occurrence rate remains; in this respect, it seems important to use the moving average in real time.

In this way, the Bayesian approach is expected to contribute to improving the accuracy of probabilistic forecasting of the earthquake's timing. However, toward its practical use, it is significant to examine the Bayesian method in a way that relaxes the two conditions imposed in this study. First, this study considered Bayes' theorem that yields the inverse probability given only one length of a lower inter-event time as the condition; its extension to Bayesian updating is essential to utilize more information on multiple intervals at the lower cut-off magnitude. Second, this study derived the inverse probability under the condition that the inter-event correlation is weak. The author already discussed the Bayesian approach in combination with the unfolding transformation, in which the theory for the time series with weak inter-event correlation is enough. While this approach is advantageous in the theoretical analysis, there remains difficulty in time series treatment. Another more straightforward approach is extending the framework to a general seismic time series that includes notable aftershock activities, which can avoid such difficulty, is theoretically intriguing, and may improve forecasting because such major aftershocks possibly yield correlations between events that can more effectively work in the Bayesian approach. Thus, further examination of the inverse probability in Bayesian updating using general seismic time series is significant for more practical and effective probabilistic forecasting.

5.5.5 A Method to Use Limited Seismic Activity Data

The catalog analysis found that the conditional and inverse probabilities collapse on Δm-dependent functions after rescaling by the average inter-event times. The author utilized this property to use as much data as possible from the limited seismic data. This method seems helpful for further research into analyzing seismic time series with limited data.

However, the author points out that this approach only applies in the magnitude range the GR law strictly holds. This can be seen, in particular, in the case of SCA3; as shown in Fig. 5.14(f), the tail part of the theoretically derived scaling function ($f(y)$) deviated from the result of catalog analysis (a similar trend is also observed in Figs. 5.14(c) and (h)). The cause of this deviation can be traced back to the poor fitting parameter estimation in Eq. (5.13) due to the incomplete scaling collapse of $\zeta(y)$ (Fig. 5.9(f)), and the fundamental cause is considered to be in the minor discrepancy in the magnitude frequency relative to the exact GR law (which can be seen in the slight trend of $\hat{b}$ in Fig. 5.2(f)). Thus, while the method in this study is effective in compensating for the shortage of seismic data, it should be noted that the results are sensitive to the accurate following of magnitude frequency to the GR law.

Chapter 6

Conclusion

This chapter summarizes the contents of the thesis and discusses the future outlook.

Chapter 2 introduced a method to characterize the temporal hierarchy in marked point processes. This Chapter considered only the magnitude and time of seismic activity (Fig. 1.2(c)). We considered the conditional probability between inter-event times (τ_m and τ_M) of point processes at different cut-off magnitude M and $m(< M)$ (see Fig. 1.2(c)) and the integral equation with the conditional probability at its kernel that is linking the inter-event time distributions at different cut-off magnitudes ($p_m(\tau_m)$ and $p_M(\tau_M)$). Based on the hierarchical nature captured by the conditional probability, the properties of the inter-event time distributions can be discussed by the integral equation. The property of the conditional probability was examined using seismic catalogs, the mixture time series and the aftershock sequence before and after the Tohoku mainshock in the local JMA catalog and the stationary time series in the global PDE catalog. Based on the properties, the multi-fractal relation was derived from the integral equation, and a qualitative explanation of the hyperbolic shape of the multi-fractal relation was given. However, the theoretical approach could not reach a quantitative explanation, and after considering the causes of the discrepancies between the theory and the catalog analysis, we concluded that the detailed understanding of the conditional probability, the new amount introduced to characterize the hierarchy of seismic time series, is necessary. Thus, this Chapter yielded new basic tools throughout this thesis to consider the hierarchical structure of seismic activity, and a qualitative description based on the properties found in the conditional probability was given for the multi-fractal relation for the empirical hierarchical property of the inter-event time statistics. The functional form of the conditional probability is gradually clarified in the subsequent chapters in considering the Bayesian method for probabilistic forecasting in Chapter 3 and the analysis of the time series with weak inter-event correlations in Chapter 5.

Chapter 3 considered the Bayesian method to extend the renewal process approach for probabilistic forecasting earthquake timing in Section 1.2.2 incorporating the information of small-scale earthquakes. This Chapter continues to consider the magnitude and time aspects as in Fig. 1.2(c)). Bayes' theorem based on the conditional probability in Chapter 2 was first considered theoretically. Second, the extension to Bayesian updating, which yields the inverse probability of the inter-event times at the upper threshold magnitude (τ_M), taking into consideration the multiple inter-event times at its lower threshold magnitude, was considered for simple stationary marked Poisson processes, analytically. Toward the application of Bayesian updating to the time series with inter-event correlations, the approximation function of the inverse probability, which involves the product of the conditional probabilities, was also derived. Bayesian updating was applied to the ETAS time series, and its effectiveness for forecasting large event timing was statistically examined. The effectiveness of forecasting was shown to depend on the stationarity of the time series; if the time series is dominated by the stationary activity, the estimate of the next large event timing can be stable around the actual timing; though if the non-stationary relaxation process is dominant, the estimate continues to shift while updating and thus effective forecasting is not achieved. In particular, in two time regimes, immediately or long after a large event, the effectiveness was statistically shown to be relatively good. Both time regimes are important for seismic risk assessment. On the other hand, during the relaxation process of Omori – Utsu aftershocks, the effectiveness worsens.

Although the prior distribution of the Bayesian updating, the inter-event time distribution, was almost a monotonic function, the inverse probability was shown to have a peak, along with its condition, around

the actual event time when the stationarity was dominant. This qualitatively suggests the superiority of the Bayesian approach over the renewal process approach based on the inter-event time distribution. However, the inverse probability itself was not derived for time series with inter-event correlations, which was partially worked on in Chapter 5. Derivation of the inverse probability, or its approximation function, is an important issue both for probabilistic forecasting as well as a quantitative comparison between these two approaches. We have discussed a way to incorporate the correlation between seismic activity and the subsequent major event into the point process approach by replacing the prior distribution with the conditional intensity function of the ETAS model. As this is just a suggestion, a quantitative comparison is necessary as well. The ineffectiveness of forecasting in the non-stationary regime suggested in this Chapter is discussed in Chapter 5.

Chapter 4 extended the Bayesian framework considered in Chapter 3 to the spatiotemporal marked point process as in Figs. 1.2(a, b) as well as (c). We extended Bayes' theorem in Chapter 3 by considering the extension of the conditional probability and the integral equation introduced in Chapter 2 to the spatiotemporal version and further considered the theoretical framework of Bayesian updating. The extension of the Bayesian framework in Chapter 4 is important for forecasting not only timing but the place of future earthquakes, taking into consideration the seismic gaps or quiescence qualitatively captured so far for quantitative forecasting. It was theoretically examined how the interaction between the time intervals and the spatial locations of the different scale events affects the forecasting within the Bayesian framework. As the preliminary analysis of the spatiotemporal Bayesian approach showed unclear results, it is necessary to use a seismicity-like model time series such as the HIST-ETAS model or the OFC model, which can prepare abundant data for statistical evaluation of the effectiveness of the spatiotemporal Bayesian approach.

Chapter 5 examined the Bayesian framework theoretically considered and numerically tested in Chapter 3 using seismic catalogs. The probability densities in Bayes' theorem, the conditional probability, the inter-event time distribution, and the inverse probability were examined using three seismic catalogs: the global CMT catalog and the local JMA and Southern California catalogs. Only the seismic time series temporally transformed by the occurrence rate was analyzed for ease of analysis and for consideration of Corral's scaling universality in the inter-event time distribution in terms of the hierarchical nature of seismicity. The catalog analysis showed a new scaling property related to the conditional probability. Based on that scaling property, the functional forms of the three probability densities were derived. The derived inverse probability reproduces the characteristic of the catalog analysis, such as the peak shape. Although the correlation between events is weak in the target time series in this Chapter, this result is important for probabilistic forecasting in the Bayesian framework. The scaling universality of the inter-event time distribution was also reconsidered, regarding it as a universality in the temporal fluctuation in the unfolding transformed seismic time series.

The catalog analysis showed that the hierarchical structures captured by the conditional probability in the unfolding transformed time series are indistinguishable between originally stationary and aftershock time series. This suggests a possible way to improve the forecasting in the relaxation process in which Bayesian inference was shown to be invalid in Chapter 3. Using the moving average and unfolding-transforming the time series in real-time can be a way to improve the forecasting in the relaxation process.

In this way, the present study examined the Bayesian approach in marked point processes for forecasting future large events. The most important remaining task is to derive the inverse probability or its approximation function in Bayesian updating for time series with (strong or weak) inter-event correlations and to compare it quantitatively with existing approaches. It is also interesting to apply the Bayesian approach to other physical processes than seismic activity that can be represented by marked point processes. In the process of considering the Bayesian framework, it is inevitable to better understand the conditional probability and the integral equation that form the basis of this theory. It should also contribute to a better understanding of the temporal properties of seismicity, including its hierarchical nature, and further the complexity of seismic activity.

Acknowledgement

The author thanks the supervisor, Professor Ken Umeno, for his guidance. He was kind enough to discuss my study, and I learned a lot from his attitude as a professional researcher. The author thanks Professor Satoshi Tsujimoto and Professor Satoshi Taguchi for valuable comments and for reviewing the thesis.

The author thanks Professor Yoji Aizawa for his collaboration and valuable advice throughout the course of the study, from the initial stage of my research activity. The author is grateful to Professor Takahiro Hatano for his research guidance.

Lastly, I would like to thank and honor my parents and grandparents.

Appendix A

Derivation of Eq. (2.24)

This Appendix is based on the Appendix in Ref. [1], paper 1 in the list of the author's papers.

Here we show the process to derive the Eq. (2.24). First, by substituting $\tau_m = \langle \tau_M \rangle = \beta_M \Gamma(1+1/\alpha_M)$, Eq. (2.23) is rewritten in the form:

$$\frac{\left(\frac{\beta_M}{\beta_m}\right)^{\alpha_m - 1} \Gamma\left(1+\frac{1}{\alpha_M}\right)^{\alpha_m - \alpha_M} (1+\overline{\eta}(1)\Delta m)}{\exp\left[\left(\frac{\beta_M}{\beta_m}\right)^{\alpha_m} \Gamma\left(1+\frac{1}{\alpha_M}\right)^{\alpha_m} - \Gamma\left(1+\frac{1}{\alpha_M}\right)^{\alpha_M}\right]} = \frac{\alpha_M \Gamma\left(1+\frac{1}{\alpha_M}\right)}{\alpha_m \Gamma\left(1+\frac{1}{\alpha_m}\right)} \times \frac{1-\overline{a}_{\Delta m}(1)\left\{1-\exp\left[-\left(\frac{\beta_M \Gamma\left(1+\frac{1}{\alpha_M}\right)}{\beta_m}\right)^{\alpha_m}\right]\right\}}{1-\overline{a}_{\Delta m}(1)\left\{1-\frac{\beta_m}{\beta_M}\frac{\gamma\left(\frac{1}{\alpha_m}, \left(\frac{\beta_M \Gamma\left(1+\frac{1}{\alpha_M}\right)}{\beta_m}\right)^{\alpha_m}\right)}{\alpha_m \Gamma\left(1+\frac{1}{\alpha_M}\right)}\right\}}. \tag{A.1}$$

In order to derive the relation between α_m and β_m, the above equation is differentiated with respect to M and taken the limit of $\Delta m \to 0$ (or $M \to m$). The l.h.s. and r.h.s. of the equation become:

$$\begin{aligned}
\lim_{M\to m}\frac{\partial}{\partial M}(\text{l.h.s.}) &= (\alpha_m - 1)\frac{\beta_m'}{\beta_m} - \alpha_m' \log\Gamma\left(1+\frac{1}{\alpha_m}\right) \\
&\quad + \alpha_m \Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}\left(-\frac{\beta_m'}{\beta_m} + \frac{\alpha_m'}{\alpha_m^2}\psi\left(1+\frac{1}{\alpha_m}\right)\right) + \overline{\eta}(1), \\
\lim_{M\to m}\frac{\partial}{\partial M}(\text{r.h.s.}) &= \frac{\alpha_m'}{\alpha_m}\left(1 - \frac{\psi\left(1+1\frac{1}{\alpha_m}\right)}{\alpha_m}\right) + \overline{a}_0'(1)\left[e^{-\Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}} - \frac{\gamma\left(\frac{1}{\alpha_m}, \Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}\right)}{\alpha_m \Gamma\left(1+\frac{1}{\alpha_m}\right)}\right] \\
&\quad - \alpha_m' \Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}\left(\log\Gamma\left(1+\frac{1}{\alpha_m}\right) - \frac{\psi\left(1+\frac{1}{\alpha_m}\right)}{\alpha_m}\right).
\end{aligned} \tag{A.2}$$

Therefore:

$$\begin{aligned}
\frac{\beta_m'}{\beta_m} &= \frac{1}{(\alpha_m - 1) - \alpha_m \Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}}\left\{-\overline{\eta}(1) + \overline{a}_0'(1)\left[e^{-\Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}} - \frac{\gamma\left(\frac{1}{\alpha_m}, \Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}\right)}{\alpha_m \Gamma\left(1+\frac{1}{\alpha_m}\right)}\right]\right. \\
&\quad \left. + \alpha_m'\left[\frac{1}{\alpha_m} - \frac{\psi\left(1+\frac{1}{\alpha_m}\right)}{\alpha_m^2} + \log\Gamma\left(1+\frac{1}{\alpha_m}\right)\left(1 - \Gamma\left(1+\frac{1}{\alpha_m}\right)^{\alpha_m}\right)\right]\right\}.
\end{aligned} \tag{A.3}$$

Finally, Eq. (2.24) is obtained by integrating both sides of Eq. (A.3) from m_0 to m.

Appendix B

Derivation of the Conditional and Inverse Probability Density Functions for the Stationary Marked Poisson Process

This Appendix is based on the Appendix in Ref. [2], paper 2 in the list of the author's papers.

This Appendix derives the conditional and inverse probability density functions analytically for the stationary marked Poisson process.

First, we derive the conditional probability density function (Eq. (3.15)) by substituting Eqs. (3.11) – (3.14) into Eq. (3.10). The denominator of Eq. (3.10) 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 (3.15) is obtained by rearranging the above equations.

We confirm that Eq. (2.3) with this conditional probability in its kernel has the exponential distribution (Eq. (3.9)) as the solution. By dividing both sides of Eq. (2.3) by N_m and rewriting it using $N_M/N_m = \langle\tau_m\rangle/\langle\tau_M\rangle$ as well as Eq. (3.15):

$$\begin{aligned}
p_m(\tau_m) = \frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\int_{\tau_m}^{\infty}\Big[&e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}\delta(\tau_M-\tau_m)\\
&+\frac{A_{\Delta m}}{\langle\tau_M\rangle}e^{-A_{\Delta m}\frac{\tau_m}{\langle\tau_M\rangle}}\left(A_{\Delta m}\frac{\tau_M-\tau_m}{\langle\tau_M\rangle}+2\right)\theta(\tau_M-\tau_m)\Big]p_M(\tau_M),
\end{aligned} \tag{B.1}$$

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. (B.1) is equivalent to the l.h.s., $p_m(\tau_m) = e^{-\frac{\tau_m}{\langle\tau_m\rangle}}/\langle\tau_m\rangle$. Substitute $p_M(\tau_M) = e^{-\frac{\tau_M}{\langle\tau_M\rangle}}/\langle\tau_M\rangle$ into the r.h.s. of Eq. (B.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:

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

and the integral involving the step function (R_2) is:

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

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

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

Second, we derive the inverse probability density function (Eq. (3.16)). From Eq. (3.15), the generalized probability density functions for the stationary marked Poisson process are derived as:

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

Equation (3.16) is obtained by substituting the above equations in Eq. (3.7).

Derivative of Eq. (3.16) by τ_M is:

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

Therefore, the inverse probability density function has a peak at:

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

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

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

Appendix C

Derivation of Eq. (3.24) from Eq. (3.23)

This Appendix is based on the Appendix in Ref. [2], paper 2 in the list of the author's papers.

The summation part in the r.h.s. of Eq. (3.23) is:

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

The first term on the r.h.s. of the above equation is transformed by substituting Eqs. (3.11) – (3.14) as:

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

The second term except the step function is also transformed as:

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

Finally, Eq. (3.24) is obtained by substituting Eqs. (C.1) and (C.2) in Eq. (3.23), with the denominator of the r.h.s. of Eq. (3.23):

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

Appendix D

Another Bayesian Updating Method

This Appendix is based on the Appendix in Ref. [2], paper 2 in the list of the author's papers.

In this Appendix, we consider another method of Bayesian updating from the one introduced in Section 3.4; 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 stationary marked Poisson process.

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

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

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

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

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

Therefore, the first representation is obtained as:

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

This equation is rewritten using Eq. (3.9) in the explicit form as:

$$N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) = N_M \left(1 - \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle}\right)^{n-1} \frac{1}{\langle \tau_m \rangle^n} e^{-\frac{\sum_{i=1}^{n} \tau_m^{(i)}}{\langle \tau_m \rangle}} p^*_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) d\tau_M d\tau_m^n. \quad \text{(D.2)}$$

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

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

Therefore, the second representation is obtained as:

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

This equation is rewritten in the explicit form using Eqs. (3.11) – (3.14) in the same way as in Appendix C.

$$N^*_{mM}(\tau_M, \tau_m^{(1)}, \cdots, \tau_m^{(n)}) = N_M \frac{1}{\langle \tau_M \rangle} e^{-\frac{\tau_M}{\langle \tau_M \rangle}} d\tau_M d\tau_m^n \left(\frac{A_{\Delta m}}{\langle \tau_M \rangle}\right)^{n-1} \\ \times \left[e^{-A_{\Delta m}\frac{\tau_M}{\langle \tau_M \rangle}} \delta\left(\tau_M - \sum_{i=1}^{n} \tau_m^{(i)}\right) + \left(\frac{A_{\Delta m}}{\langle \tau_M \rangle}\right) e^{-A_{\Delta m}\frac{\sum_{i=1}^{n} \tau_m^{(i)}}{\langle \tau_M \rangle}} \theta\left(\tau_M - \sum_{i=1}^{n} \tau_m^{(i)}\right) \right]. \tag{D.3}$$

Finally, $p^*_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)})$ is derived from Eqs. (D.2) and (D.3) as

$$p^*_{Mm}(\tau_M|\tau_m^{(1)}, \cdots, \tau_m^{(n)}) \\ = \frac{\langle \tau_m \rangle}{\langle \tau_M \rangle} \left[\frac{A_{\Delta m}}{\langle \tau_M \rangle} e^{-\frac{\tau_M - \sum_{i=1}^{n} \tau_m^{(i)}}{\langle \tau_M \rangle}} \theta\left(\tau_M - \sum_{i=1}^{n} \tau_m^{(i)}\right) + e^{-\frac{\tau_M - \sum_{i=1}^{n} \tau_m^{(i)}}{\langle \tau_m \rangle}} \delta\left(\tau_M - \sum_{i=1}^{n} \tau_m^{(i)}\right) \right]. \tag{D.4}$$

This is different from Eq. (3.24), which reflects the difference whether the position of lower intervals is specified.

Appendix E

Derivation of Eq. (3.29)

This Appendix is based on the Appendix in Ref. [2], paper 2 in the list of the author's papers.

First, we substitute Eqs. (3.11) – (3.14) into Eq. (3.28):

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

In the above equation, the summation part of the term that includes the step function can be transformed as:

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

Finally, Eq. (3.29) is obtained by rearranging the above equations.

Appendix F

Derivation of Eq. (3.34)

This Appendix is based on the Appendix in Ref. [2], paper 2 in the list of the author's papers.

In this Appendix, $P(\tau_m^{(1)}, \cdots, \tau_m^{(l)}|\tau_M)$ is derived for the stationary marked Poisson process. First, we divide the summation in Eq. (3.33) into two parts:

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

This equation is further rewritten by substituting Eqs. (3.11) – (3.14) in the same way as in Appendix C. 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 $P_i(\tau_m^{(i)}|\tau_M) := \left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right) e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}$.

Therefore:

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

Finally, because $\tau_M \gneqq \sum_{i=1}^{l} \tau_m^{(i)}$ holds for $l < n$ by the condition of Eq. (3.26):

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

Appendix G

Derivation of Eq. (3.39)

This Appendix is based on the Appendix in Ref. [2], paper 2 in the list of the author's papers.

In this Appendix, the approximation function of the inverse probability density function for the stationary marked Poisson process (Eq. (3.39)) is derived. By substituting Eqs. (3.9), (3.15), and (3.34) into Eq. (3.36):

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

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. (G.1) are respectively transformed as:

$$
\begin{aligned}
&\prod_{i=1}^{n}\frac{e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}\frac{A_{\Delta m}}{\langle\tau_M\rangle}\left[A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\left(1-\frac{\tau_m^{(i)}}{\tau_M}\right)+2\right]}{\frac{1}{\langle\tau_m\rangle}e^{-\frac{\tau_m^{(i)}}{\langle\tau_m\rangle}}\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}+1\right)} \\
&\quad=\left(A_{\Delta m}\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^n\left(\prod_{i=1}^{n}e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}+\frac{\tau_m^{(i)}}{\langle\tau_m\rangle}}\right)\left[\prod_{i=1}^{n}\frac{A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}+1-\left(A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}-1\right)}{A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}+1}\right] \\
&\quad=\left(A_{\Delta m}\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^n\left(\prod_{i=1}^{n}e^{\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}\right)\left[\prod_{i=1}^{n}\left(1-\frac{\tau_m^{(i)}-\frac{\langle\tau_M\rangle}{A_{\Delta m}}}{\tau_M+\frac{\langle\tau_M\rangle}{A_{\Delta m}}}\right)\right].
\end{aligned}
\tag{G.2}
$$

$$
\begin{aligned}
\prod_{i=1}^{n}\frac{\left(\frac{A_{\Delta m}}{\langle\tau_M\rangle}\right)e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}}{\frac{1}{\langle\tau_m\rangle}e^{-\frac{\tau_m^{(i)}}{\langle\tau_m\rangle}}} &= \left(A_{\Delta m}\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^n\prod_{i=1}^{n}e^{-A_{\Delta m}\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}+\frac{\tau_m^{(i)}}{\langle\tau_m\rangle}} \\
&= \left(A_{\Delta m}\frac{\langle\tau_m\rangle}{\langle\tau_M\rangle}\right)^n\prod_{i=1}^{n}e^{\frac{\tau_m^{(i)}}{\langle\tau_M\rangle}}.
\end{aligned}
\tag{G.3}
$$

Finally, Eq. (3.39) is derived by substituting Eqs. (G.2) and (G.3) into Eq. (G.1).

Appendix H

Relation Between the Inverse Probability Density Function and its Approximation Function in the Stationary Marked Poisson Process

This Appendix is based on the Appendix in Ref. [2], paper 2 in the list of the author's papers.

In this Appendix, we discuss the relation between the inverse probability density function (Eq. (3.23)) and Eq. (3.36), i.e., the approximations made on Eq. (3.23) that correspond to the assumptions made in Section 3.4.2 to derive Eq. (3.36).

The summation in Eq. (3.23) can be decomposed into:

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

The first term on the r.h.s. of Eq. (H.1) is equivalent to $-(n-1)\prod_{i=1}^{n}P_i$ (Appendix F), and then, this term formally coincides with the correction term in Eq. (3.36). Therefore, the second term corresponds to the kernel part (Eq. (3.37)).

n-consecutive lower intervals must be included in only one upper interval. Under this condition, three constraints are imposed on the lower intervals, which appear on the l.h.s. of Eq. (H.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 $(i-j+1)$, $\rho_{mM}\left(\tau_m^{(j)}|i-j+1,\tau_M-\sum_{k=1}^{j-1}\tau_m^{(k)}\right)$.

These constraints are relaxed in the derivation in Section 3.4.2. In the view in Section 3.4.2, the upper intervals of length τ_M are collected and the new time series is generated as shown in Fig. 3.4. 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. (H.1), i.e., the correction term in Eq. (3.36). (3) The constraints imposed on the condition in ρ_{mM} are removed; because the probability of the length of j-th interval is only not affected by other lower intervals, the temporal part of ρ_{mM} is replaced by τ_M (Eq. (H.2)). In addition, the constraint on the number of division can be eliminated by taking the average (Eq. (H.3)).

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

$$\begin{aligned} &\approx \frac{\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M)\rho_{mM}(\tau_m^{(j)}|i,\tau_M)}{\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M)} \\ &= p_{mM}(\tau_m^{(j)}|\tau_M). \end{aligned} \tag{H.3}$$

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

In this way, the approximate view in Section 3.4.2 implies the following replacement in the exact inverse probability density function.

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

Appendix I

Distance between the Inverse Probability Density Function and the Inter-event Time Distribution

This Appendix is based on the Appendix in Ref. [2], paper 2 in the list of the author's papers.

In this Appendix, we derive the distance between the inverse probability density function (Eq. (3.24)) and the inter-event time distribution ($p_M(\tau_M)$):

$$D(p_{Mm}||p_M) := \int_T^\infty \left|p_\theta(\tau_M, T) - p_M(\tau_M)\right|^2 d\tau_M, \tag{I.1}$$

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. (I.1), the distance is derived as:

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

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 J

On the Cause of the Separation of $\langle D'(p_{Mm}^{\rm approx}||p_M)\rangle$ and $\langle D''(p_{Mm}^{\rm approx}||p_M)\rangle$ at Large T

This Appendix is based on the Appendix in Ref. [2], paper 2 in the list of the author's papers.

In this Appendix, we examine the cause of the separation between $\langle D'(p_{Mm}^{\rm approx}||p_M)\rangle$ and $\langle D''(p_{Mm}^{\rm approx}||p_M)\rangle$ at long-elapsed time T. Let us compare Fig. 3.6 to Figs. K.6(a) and K.6(c) for $\mathcal{N} = 10^5$. The separation is suppressed compared to that shown in Fig. 3.6, 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 (3.44), and to the improvement of the calculation of distance in Eq. (3.46).

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

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

For this $\boldsymbol{P}_1'$, the interpolation and extrapolation procedures are conducted in the same way as in Section 3.5.1, and the numerical updating is executed.

Figures K.6(b) and K.6(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. K.6(a) and K.6(c), the separation is suppressed further. Combined with the results for the kernel part, these results suggest the following; the number of samples to calculate $\boldsymbol{P}_1$ is so small compared to that of the conditional probability (the number of sample is only one for an upper interval for $\boldsymbol{P}_1$ whereas all the lower intervals included in an upper interval are used as a sample to calculate the conditional probability), in particular for a large k, that its fluctuation becomes too large to compute the correction term precisely.

Appendix K

Supplementary Information for Chapter 3

This Appendix is based on the Supplemental Material of Ref. [2], paper 2 in the list of the author's papers.

The following figures are provided in this Appendix. Figures K.1 – K.7 present information regarding the numerically generated stationary marked Poisson process. Figure K.1 shows inter-event time distributions ($p_m(\tau_m)$ and $p_M(\tau_M)$) and their interpolation and extrapolation functions. Figure K.2 shows the conditional probability density function of the length of the lower intervals included in the upper interval of length τ_M ($p_{mM}(\tau_m|\tau_M)$) and its interpolation and extrapolation functions. Figure K.3 shows the conditional probability density function of the leftmost lower interval length in the upper interval of length τ_M ($P_1(\tau_m|\tau_M)$) and its interpolation and extrapolation functions. Figure K.4 shows the average number of lower intervals included in the upper interval of length τ_M ($\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M}$) and its interpolation and extrapolation functions. Figure K.5 shows an example of Bayesian updating. Figure K.6 shows the average distances between the distribution functions and the (part of) approximation functions including the results of numerical Bayesian updating using statistical amounts calculated using the $\mathcal{N} = 10^5$ time series. Figure K.7 shows the joint probability mass function of the logarithm of the positions of the maximum peak of the inverse probability density function and its approximation function calculated by numerical Bayesian updating with $\mathcal{N} = 10^5$. Figures K.8 – K.14 are about the numerically generated time series of the ETAS model. Figure K.8 shows $p_m(\tau_m)$ and $p_M(\tau_M)$, and their interpolation and extrapolation functions. Figure K.9 shows $p_{mM}(\tau_m|\tau_M)$ and its interpolation and extrapolation functions. Figures K.10 and K.11 show the conditional probability density functions of the leftmost and rightmost lower interval length included in the upper interval of length τ_M ($p^{\mathrm{L}}(\tau_m|\tau_M)$ and $p^{\mathrm{R}}(\tau_m|\tau_M)$, respectively) and their interpolation and extrapolation functions. Figure K.12 shows $\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M}$ and its interpolation and extrapolation functions. Figure K.13 shows the probability distribution/density function of $n_{\leq\mathrm{th}}$ and $\tau_{\leq\mathrm{th}}$ for $\delta_{\mathrm{th}} = 0.5$. Figure K.14 shows the statistical results of the effectiveness of forecasting for $\delta_{\mathrm{th}} = 0.25$.

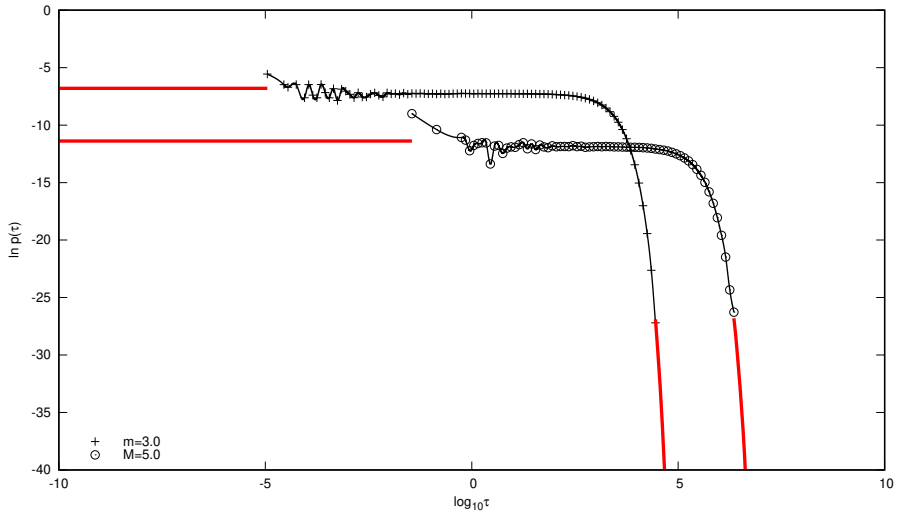

Figure K.1: Inter-event time distributions at magnitude thresholds $m = 3$ and $M = 5$ calculated for the numerically generated stationary marked Poisson process, and their interpolation and extrapolation functions. Symbols ($+$ and $\odot$) show the probability density functions obtained numerically. The intervals between the symbols are interpolated by cubic spline functions represented by the black thin curves. Further, outside of the range covered by the symbols are extrapolated by the fitting functions at the edge represented by the red bold lines; constant function ($\ln p(\tau) \equiv C$) on the small side and the exponential function ($\ln p(\tau) = A\tau + B$) on the large side. The parameter values (A, B, C) are determined by the least squares method using 10 points at each end.

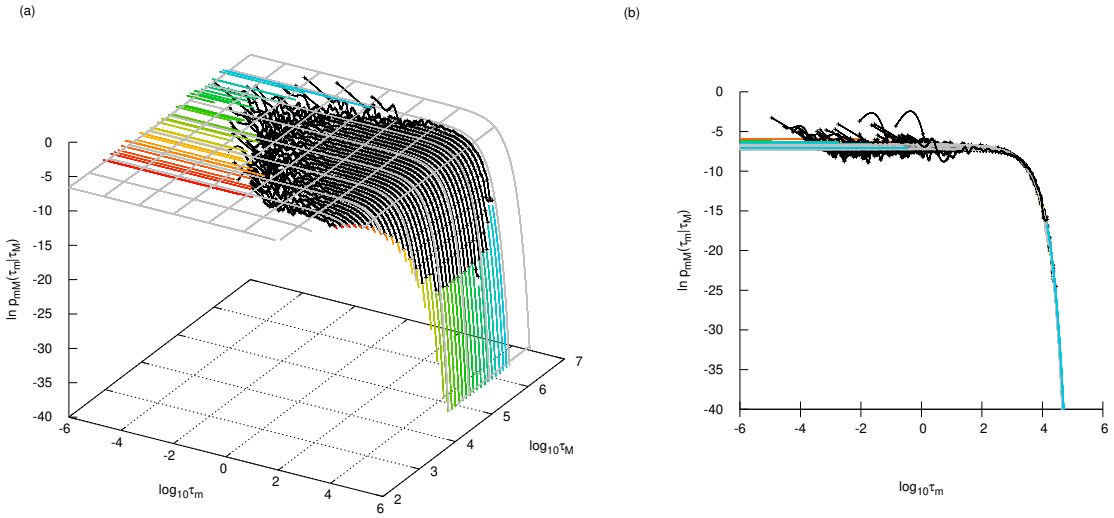

Figure K.2: Conditional probability density function of the lower interval length (τ_m) included in the upper interval of length τ_M for the stationary marked Poisson process, and its interpolation and extrapolation functions in (a) an oblique view and (b) a horizontal view parallel to the $\log_{10}\tau_M$-axis. Gray curved surface is the part of the step function of Eq. (3.15). Symbols ($+$) represent the probability density function obtained numerically. For each τ_M, only data points with 30 or more points in the range of $\tau_M > \tau_m$ are displayed. The intervals between the symbols in the $\log_{10}\tau_m$-axis direction are interpolated by cubic spline functions represented by the black thin curves. Outsides of the range covered by the symbols in the $\log_{10}\tau_m$-axis direction are extrapolated by the fitting functions at the edge represented by the colored bold lines (the color varies with $\log_{10}\tau_M$); constant function ($\ln p(\tau) \equiv C$) on the small τ_m side and the exponential function ($\ln p(\tau) = A\tau + B$) on the large τ_m side for each τ_M. The parameter values (A, B, C) are determined by the least squares method using 10 points at each end for each τ_M.

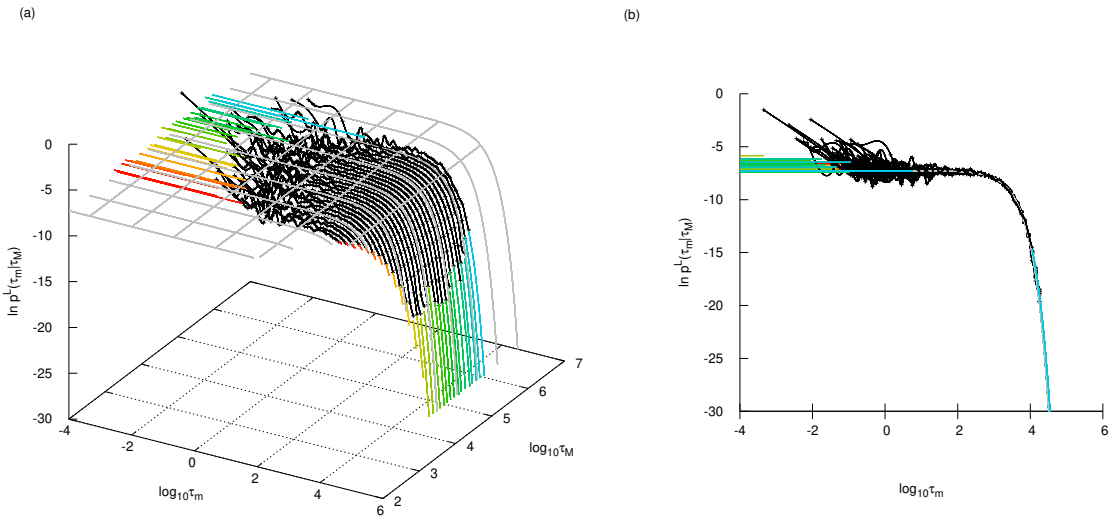

Figure K.3: Conditional probability density function of the leftmost lower interval length included in the upper interval of length τ_M for the stationary marked Poisson process, and its interpolation and extrapolation functions in (a) an oblique view and (b) a horizontal view parallel to the $\log_{10}\tau_M$-axis. Gray curved surface shows the part of the step function of Eq. (3.29). The description of the figure is the same as in Fig. K.2.

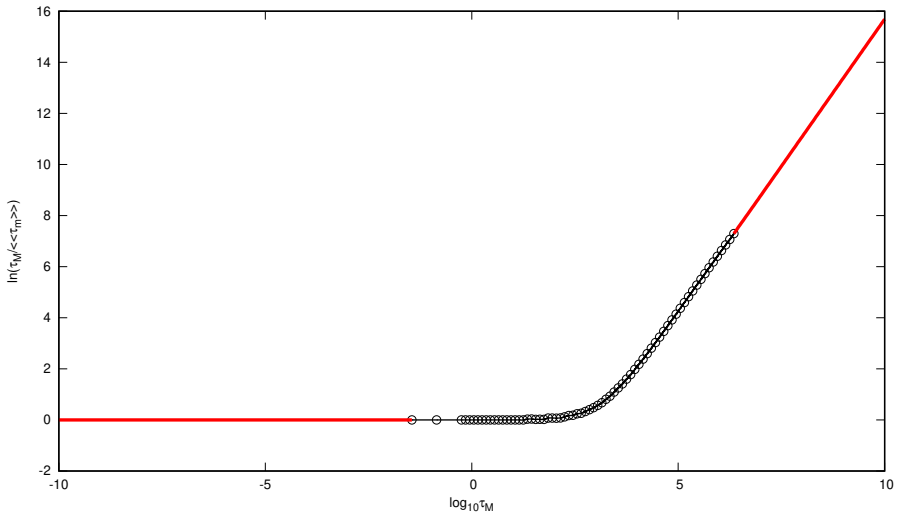

Figure K.4: Average number of lower intervals included in the upper interval of length τ_M ($\tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M}$) for the stationary marked Poisson process, and its interpolation and extrapolation functions. Symbols ($\odot$) indicate the results obtained numerically. The intervals between the symbols are interpolated by cubic spline functions represented by the black thin curves. Outsides of the range covered by the symbols are replaced or extrapolated by following functions represented by red bold lines; $\ln \tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M} \equiv 0$ on the small side, whereas the fitting function $\tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M} = A\tau_M + B$ on the large side with the parameter values determined by the least squares method using 10 points at the end.

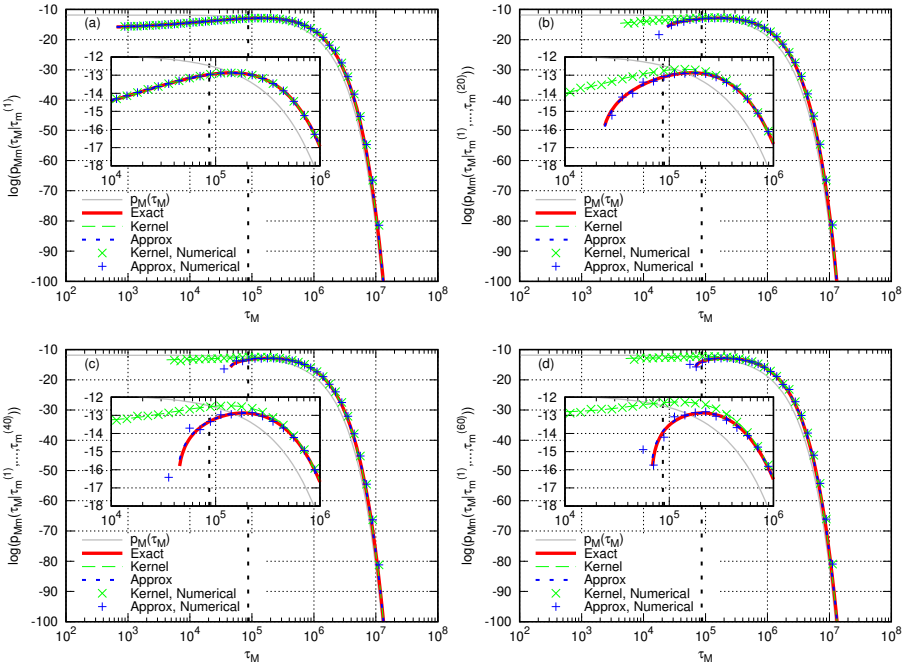

Figure K.5: Examples of the Bayesian updating on a numerically generated stationary marked Poisson process. In this example, the total number of updates between the events with a magnitude above M was 73. The horizontal axis represents the elapsed time from the previous event with a magnitude above M, and the vertical axis is the logarithmic scale of the inverse probability density function and its approximation function for $n =$ (a) 1, (b) 20, (c) 40, and (d) 60, respectively. Gray thin curve is $p_M(\tau_M)$ and black vertical dotted line indicates the actual elapsed time of the next large event with a magnitude above M. At each update, (Exact) the inverse probability density function (Eq. (3.24)), (Approx) the approximation function (Eq. (3.39)), and (Kernel) its kernel part (Eq. (3.40)) are shown. Further, the results with the numerical updating method (Eqs. (3.43) and (3.44)) are shown for the approximation function and its kernel part; (Kernel, Numerical) represents the result of Eq. (3.43), and (Approx, Numerical) is for Eqs. (3.43) and (3.44).

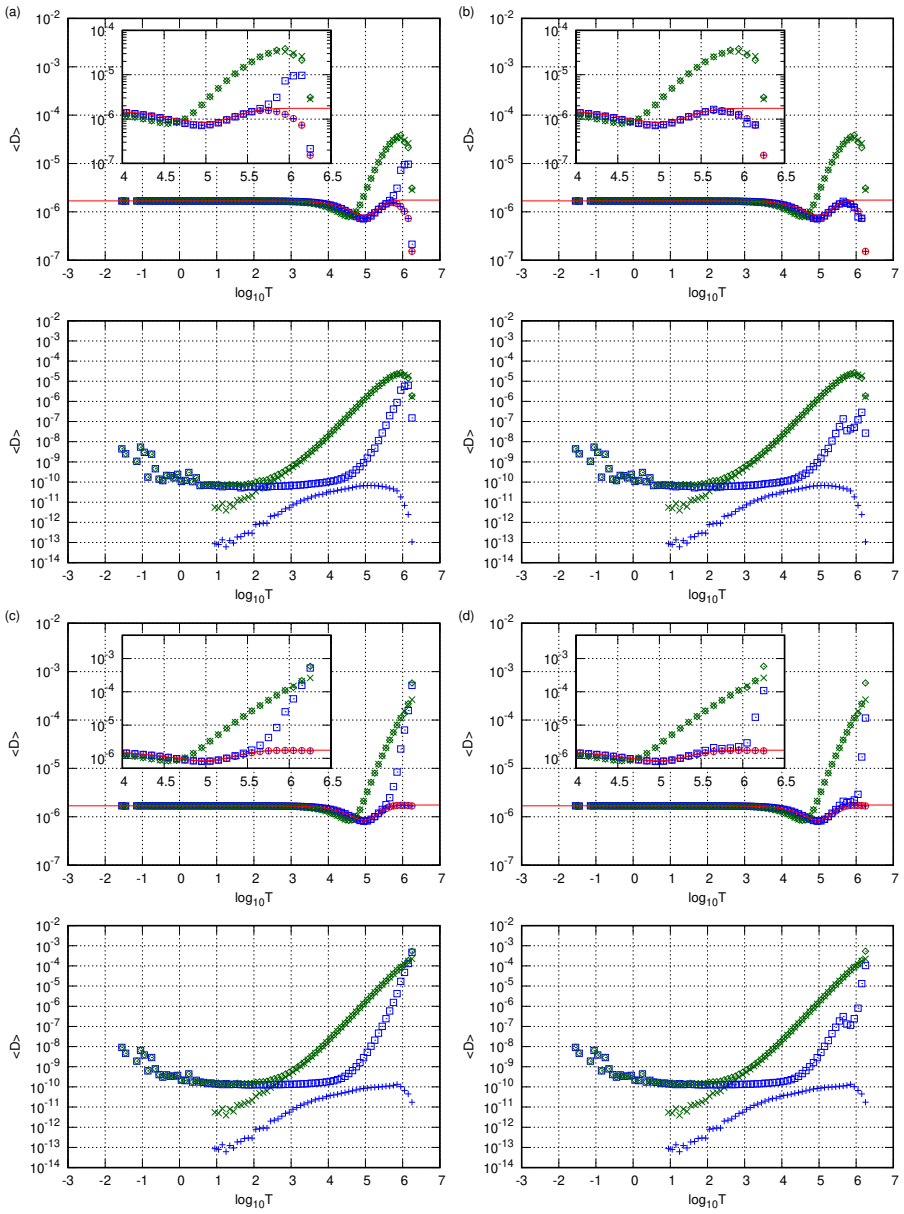

Figure K.6: Average distances for elapsed time T from the last event greater than M. Each symbol represents the same distance as in Fig. 3.6. The results of the numerical updating of $\langle D'' \rangle$ were calculated using statistical amounts in Eq. (3.42) taken from the $\mathcal{N} = 10^5$ time series with (a) $\Delta\tau_M = 0.1$ and $l_c = 0$, (b) $\Delta\tau_M = 0.1$ and $l_c = 5$, (c) $\Delta\tau_M = 0.025$ and $l_c = 0$, and (d) $\Delta\tau_M = 0.025$ and $l_c = 20$.

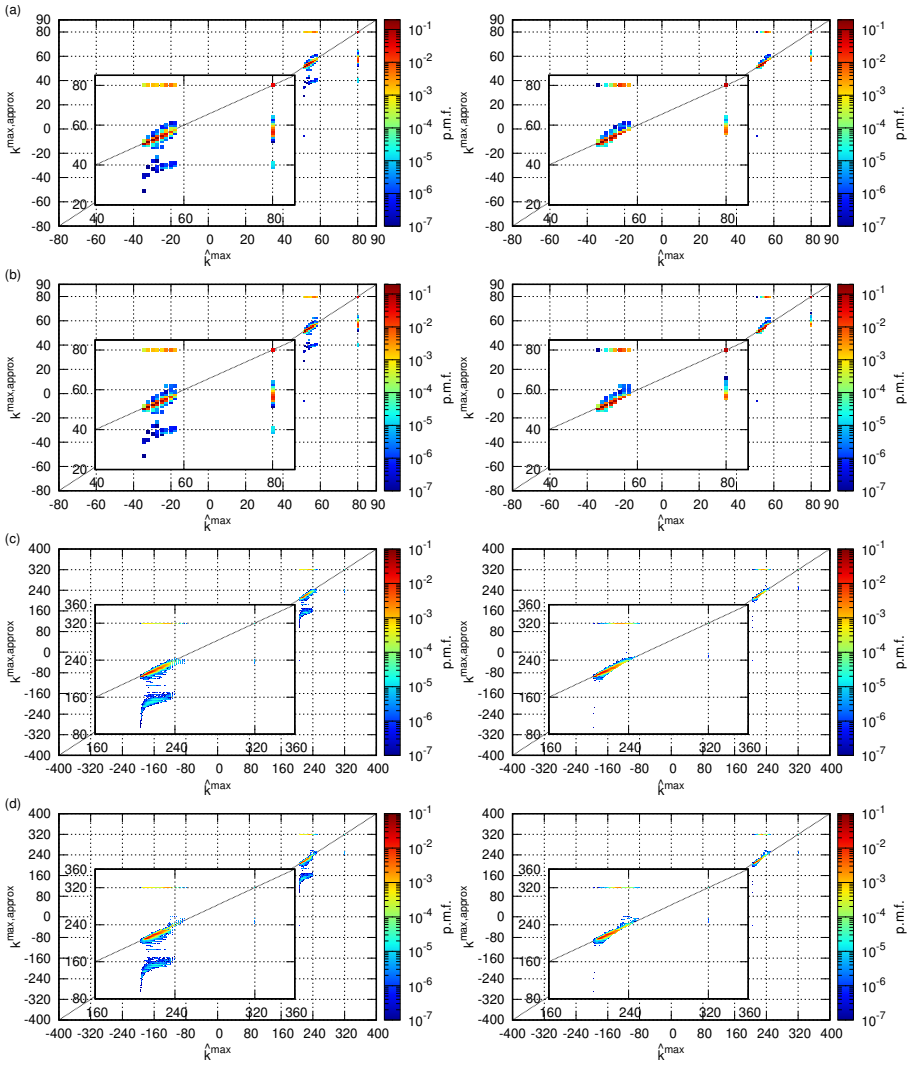

Figure K.7: Joint probability mass function of $(\hat{k}^{\max}, k^{\max,\text{approx}})$. $k^{\max,\text{approx}}$ was obtained using the numerical updating method with statistical amounts in Eq. (3.42) taken from the $\mathcal{N} = 10^5$ time series with (a) $\Delta\tau_M = 0.1$ and $l_c = 0$, (b) $\Delta\tau_M = 0.1$ and $l_c = 5$, (c) $\Delta\tau_M = 0.025$ and $l_c = 0$, and (d) $\Delta\tau_M = 0.025$ and $l_c = 20$. For (a) and (b), the horizontal lines at $k^{\max,\text{approx}} = 80$ and the vertical line at $\hat{k}^{\max} = 80$ correspond to cases when no peak is detected by the peak search. Further, for (c) and (d), the lines at $k^{\max,\text{approx}} = 320$ and at $\hat{k}^{\max} = 320$ correspond to the no peak cases. The left panels are results when the peak search was conducted in the range of $\tau_M > \max\{\tau_m^{(1)}, \cdots, \tau_m^{(n)}\}$ and the right panels in the range of $\tau_M > T$.

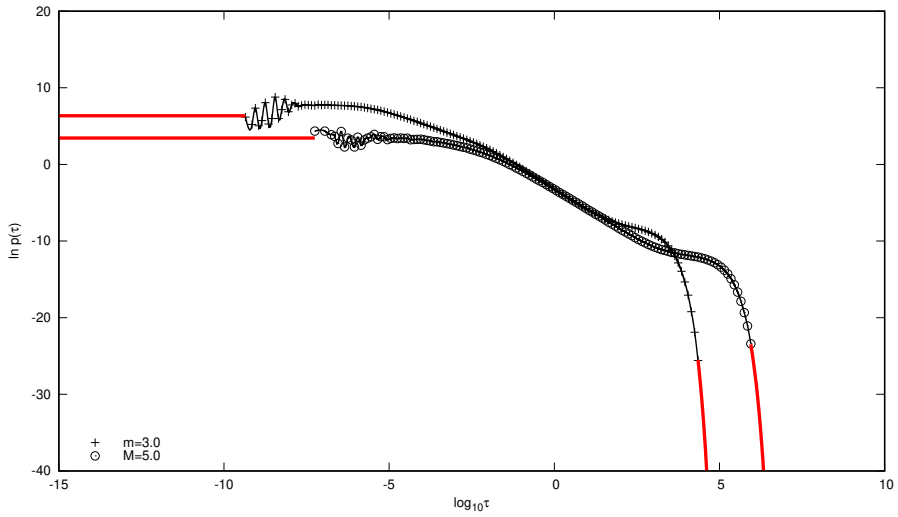

Figure K.8: Inter-event time distributions at magnitude thresholds $m = 3$ and $M = 5$ calculated for the numerically generated time series of the ETAS model, and their interpolation and extrapolation functions. The description of the figure is the same as in Fig. K.1.

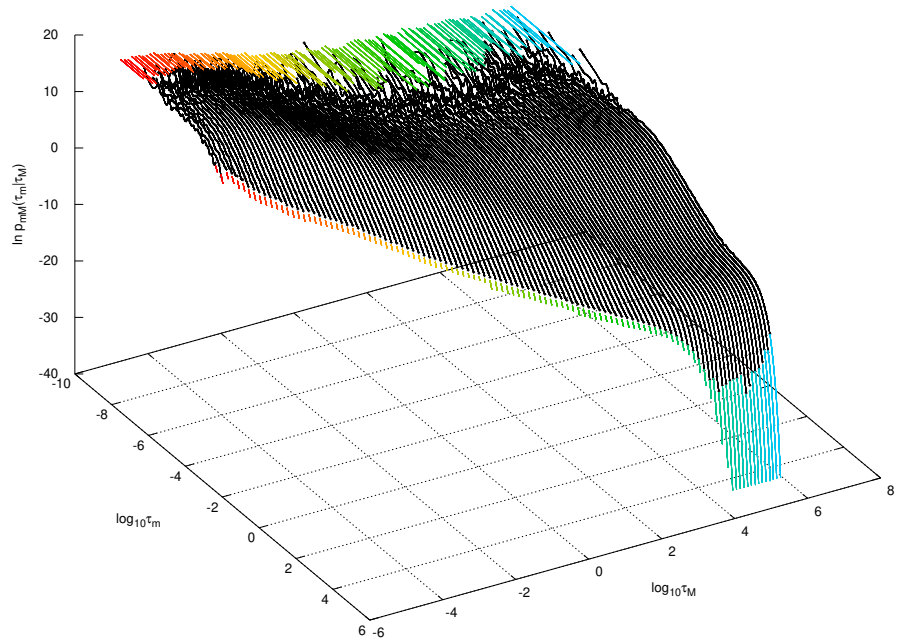

Figure K.9: Conditional probability density function of the lower interval length (τ_m) included in the upper interval of length τ_M for the numerically generated time series of the ETAS model, and its interpolation and extrapolation functions. The description of the figure is the same as in Fig. K.2.

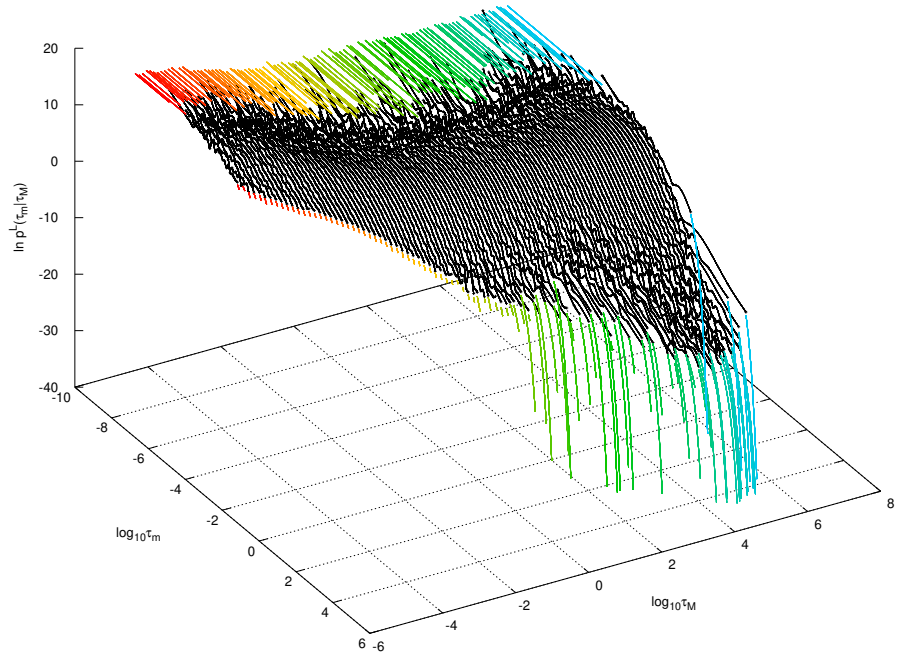

Figure K.10: Conditional probability density function of the leftmost lower interval length included in the upper interval of length τ_M for the numerically generated time series of the ETAS model, and its interpolation and extrapolation functions. The description of the figure is the same as in Fig. K.3.

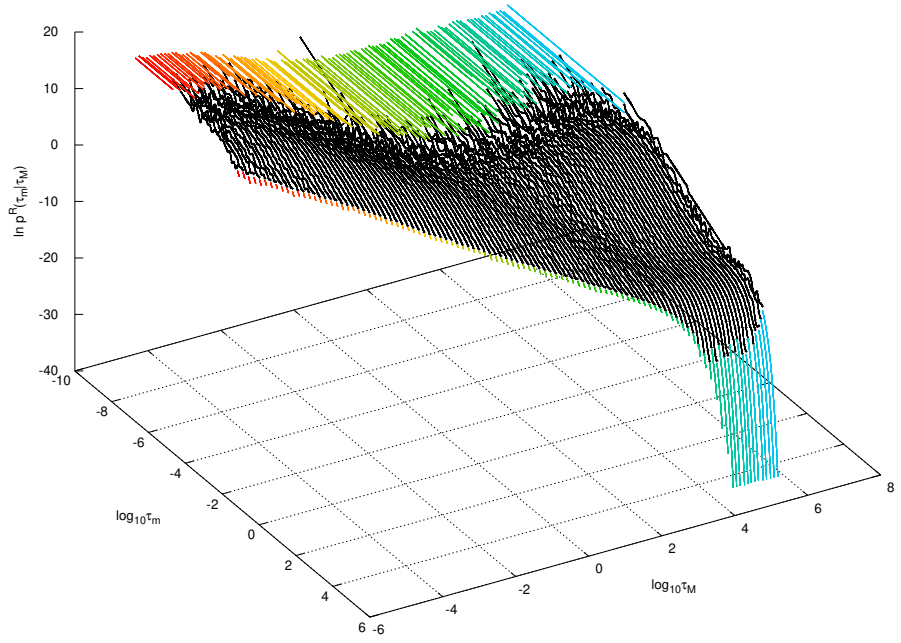

Figure K.11: Conditional probability density function of the rightmost lower interval length included in the upper interval of length τ_M for the numerically generated time series of the ETAS model, and its interpolation and extrapolation functions. The description of the figure is the same as in Fig. K.3.

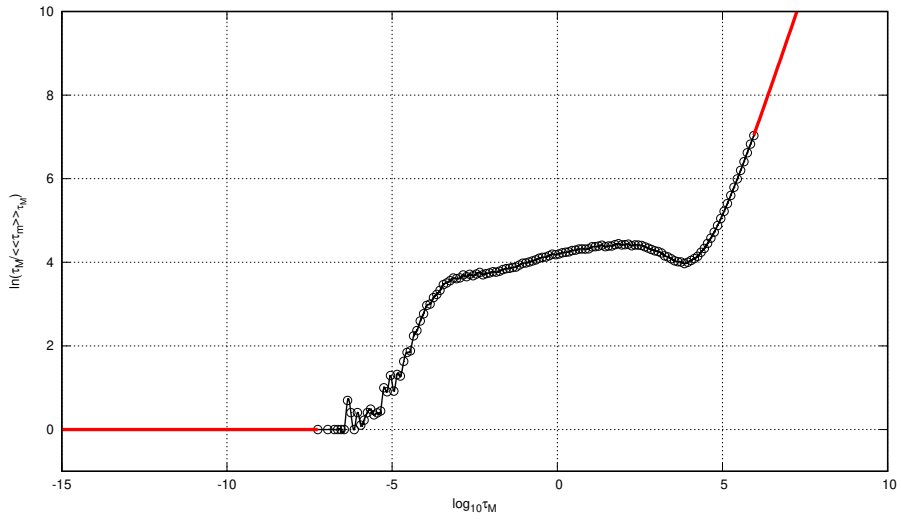

Figure K.12: Average number of lower intervals included in the upper interval of length τ_M ($\tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M}$) for the numerically generated time series of the ETAS model. The description of the figure is the same as in Fig. K.4.

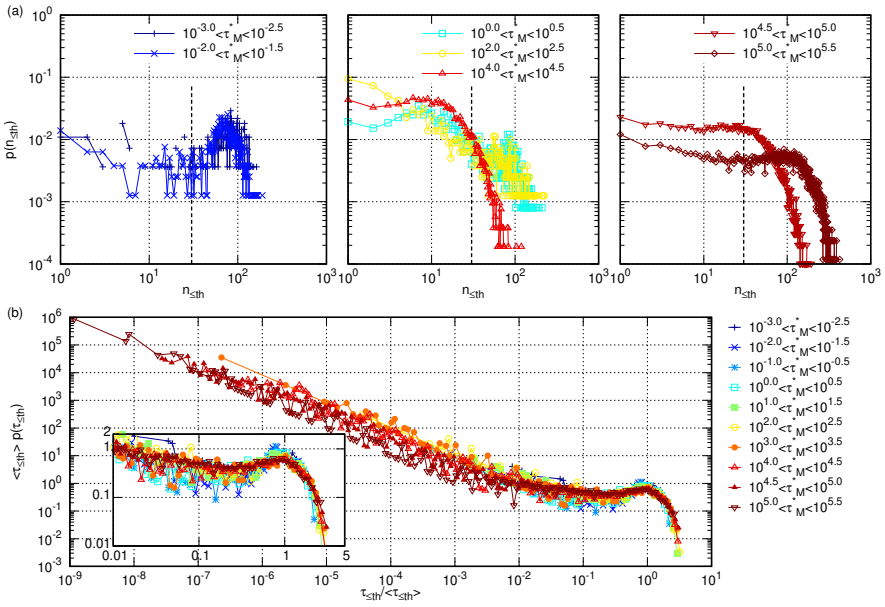

Figure K.13: (a) Probability distributions ($p(n_{\leq \mathrm{th}})$) of the number of consecutive updates ($n_{\leq \mathrm{th}}$) with $|\delta_n| \leq \delta_{\mathrm{th}} = 0.5$ for each τ_M^*. Only the cases with $n_{\leq \mathrm{th}} > 0$ are included in the population. The vertical dotted line indicates $n_{\leq \mathrm{th}} = 30$. (b) Probability density function ($p(\tau_{\leq \mathrm{th}})$) of the duration time ($\tau_{\leq \mathrm{th}}$) during n_{th}-updates with $\delta_{\mathrm{th}} = 0.5$ for each τ_M^*. The distributions rescaled by the averages ($\langle\tau_{\leq \mathrm{th}}\rangle := \int_0^\infty \tau_{\leq \mathrm{th}} p(\tau_{\leq \mathrm{th}}) d\tau_{\leq \mathrm{th}}$) are shown.

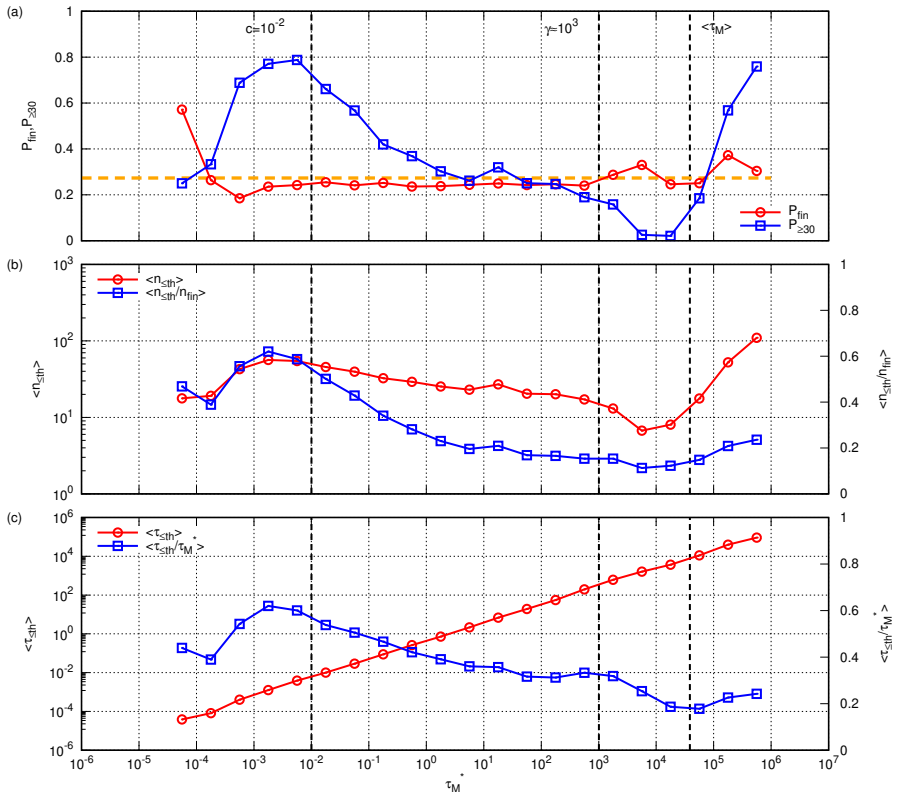

Figure K.14: Statistical results for each τ_M^* with $\delta_{\rm th} = 0.25$. (a) $P_{\rm fin}$, $P_{\geq 30}$, and the average of $P_{\rm fin}(\approx 0.27)$ for the overall τ_M^*; (b) $\langle n_{\leq {\rm th}} \rangle$ and $\langle n_{\leq {\rm th}}/n_{\rm fin} \rangle$; and (c) $\langle \tau_{\leq {\rm th}} \rangle$ and $\langle \tau_{\leq {\rm th}}/\tau_M^* \rangle$ are shown.

Appendix L

Derivations in Chapter 4

This Appendix is based on the Supplemental Material of Ref. [3], paper 3 in the list of the author's papers.

This Appendix describes the process of deriving Eqs. (4.13) and (4.16) from Eq. (4.8) based on the assumptions made in the two cases. In the following, i and j are integers in the range $1 \leq i, j \leq n$.

L.1 Derivation of Eq. (4.13) for the First Case

The probability density functions and the average number ($\langle n(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M}$) in Eq. (4.8) are respectively rewritten as follows.

L.1.1 $p_M(\boldsymbol{X}, \tau_M)$

As the spatial position ($\boldsymbol{X}$) and time interval (τ_M) in an upper spatiotemporal pair are assumed to be independent:

$$p_M(\boldsymbol{X}, \tau_M) = p_M(\boldsymbol{X}) p_M(\tau_M). \tag{L.1}$$

L.1.2 $p_m(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})$

As the lower spatial positions ($\boldsymbol{x}_i$) and time intervals ($\tau_m^{(i)}$) are assumed to be independent of each other:

$$p_m(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) = \prod_{i=1}^{n} p_m(\boldsymbol{x}_i) p_m(\tau_m^{(i)}). \tag{L.2}$$

L.1.3 $p_{Mm}(\boldsymbol{X}, \tau_M | \boldsymbol{x}_{1:n}, \tau_m^{(1:n)})$

First, $\boldsymbol{X}$ and τ_M in an upper spatiotemporal pair are assumed to be conditionally independent given $\{\boldsymbol{x}_{1:n}, \tau_m^{(1:n)}\}$, and therefore:

$$p_{Mm}(\boldsymbol{X}, \tau_M | \boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) = p_{Mm}(\boldsymbol{X} | \boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) p_{Mm}(\tau_M | \boldsymbol{x}_{1:n}, \tau_m^{(1:n)}). \tag{L.3}$$

Furthermore, as the conditional independence of $\boldsymbol{X}$ and $\tau_m^{(i)}$ given $\{\boldsymbol{x}_{1:n}\}$:

$$p_{Mm}(\boldsymbol{X} | \boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) = p_{Mm}(\boldsymbol{X} | \boldsymbol{x}_{1:n}), \tag{L.4}$$

and the conditional independence of τ_M and $\boldsymbol{x}_i$ given $\{\tau_m^{(1:n)}\}$:

$$p_{Mm}(\tau_M | \boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) = p_{Mm}(\tau_M | \tau_m^{(1:n)}). \tag{L.5}$$

Therefore,

$$p_{Mm}(\boldsymbol{X}, \tau_M | \boldsymbol{x}_{1:n}, \tau_m^{(1:n)}) = p_{Mm}(\boldsymbol{X} | \boldsymbol{x}_{1:n}) p_{Mm}(\tau_M | \tau_m^{(1:n)}). \tag{L.6}$$

L.1.4 $\langle n(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M}$

By the assumption of the conditional independence of $\{\boldsymbol{x}_{1:n}\}$ and $\{\tau_m^{(1:n)}\}$ given $\{\boldsymbol{X}, \tau_M\}$:

$$\langle n(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M} = p_{mM}(\boldsymbol{x}_{1:n}|\boldsymbol{X}, \tau_M)\langle n(\tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M}. \tag{L.7}$$

Further, by the conditional independence of $\boldsymbol{X}$ and $\tau_m^{(i)}$ given τ_M:

$$\langle n(\tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M} = \langle n(\tau_m^{(1:n)})\rangle_{\tau_M}. \tag{L.8}$$

As $\tau_m^{(i)}$ and $\tau_m^{(j)}$ $(i \neq j)$ are not conditionally independent given τ_M, Eq. (L.8) can not be further transformed.

On the other hand, by the conditional independence of τ_M and $\boldsymbol{x}_i$ given $\boldsymbol{X}$,

$$p_{mM}(\boldsymbol{x}_{1:n}|\boldsymbol{X}, \tau_M) = p_{mM}(\boldsymbol{x}_{1:n}|\boldsymbol{X}), \tag{L.9}$$

and because the conditional independence of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ $(i \neq j)$ given $\boldsymbol{X}$ is assumed, Eq. (L.9) is further transformed as follows.

$$p_{mM}(\boldsymbol{x}_{1:n}|\boldsymbol{X}) = \prod_{i=1}^{n} p_{mM}(\boldsymbol{x}_i|\boldsymbol{X}). \tag{L.10}$$

Therefore,

$$\langle n(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M} = \langle n(\tau_m^{(1:n)})\rangle_{\tau_M} \prod_{i=1}^{n} p_{mM}(\boldsymbol{x}_i|\boldsymbol{X}). \tag{L.11}$$

Finally, Eq. (4.13) is obtained by substituting Eqs. (L.1), (L.2), (L.6), and (L.11) in Eq. (4.8).

L.2 Derivation of Eq. (4.16) for the Second Case

In this case, the (conditional) independence between $\boldsymbol{X}$ and τ_M is not assumed, and therefore, $p_M(\boldsymbol{X}, \tau_M)$ and $p_{Mm}(\boldsymbol{X}, \tau_M|\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})$ are unchanged. Further, due to the assumption of the independence between lower spatial positions and time intervals, $p_m(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})$ is rewritten in the same way as Eq. (L.2).

$\langle n(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M}$ can also be rewritten in the same way as Eq. (L.7). However, because only the conditional independence of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ $(i \neq j)$ given $\{\boldsymbol{X}, \tau_M\}$ is assumed, the rewritings in Eqs. (L.8) and (L.9) are not allowed. Therefore,

$$\langle n(\boldsymbol{x}_{1:n}, \tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M} = \langle n(\tau_m^{(1:n)})\rangle_{\boldsymbol{X},\tau_M} \prod_{i=1}^{n} p_{mM}(\boldsymbol{x}_i|\boldsymbol{X}, \tau_M). \tag{L.12}$$

Finally, Eq. (4.16) is obtained by substituting Eqs. (L.2) and (L.12) in Eq. (4.8).

Appendix M

Preliminary Seismic Catalog Analysis for Spatiotemporal Bayesian Approach

This Appendix is based on the Supplemental Material of Ref. [3], paper 3 in the list of the author's papers.

This Appendix provides a preliminary examination of a seismic catalog in Southern California [18, 19]. We show the methods of catalog analysis and the numerical results of the inverse probability density function in Eq. (4.18) with $n = 1$, $p_{Mm}(\tau_M|\boldsymbol{x}, \tau_m)$, in comparison with the inter-event time distribution $p_M(\tau_M)$ and the inverse probability density function only for temporal information $p_{Mm}(\tau_M|\tau_m)$. Note that this analysis does not test the entire Bayesian approach: it only examines the inverse probability density function in Bayes' theorem that considers only one spatiotemporal lower pair.

The seismic catalog provides the information on the occurrence time, epicenter (corresponding to $\boldsymbol{x}$ and $\boldsymbol{X}$ in the text), and magnitude of each earthquake in the spatial range of 29 – 38 degrees north latitude and 112 – 123 degrees west longitude (corresponding to the spatial area S in Chapter 4). In this analysis, the time range is from 01/01/1981 to 03/31/2022. Figure M.1 shows $P(M)$ for earthquakes recorded in the seismic catalog within the above spatiotemporal domain. $P(M)$ can be fit by the GR law, in particular in the range $M \in [2.0, 5.0]$. We focus our analysis on $M \geq 2.0$ following the result in Fig. M.1 and set the cut-off magnitude in this range, though note that the completeness magnitude (M_c) fluctuates in space [37, 40, 102] and varies with time [38, 39], and thus it partially exceeds 2 set here. In particular, the tendency of M_c to be higher at the edge of S [37, 40, 102] should be considered in further study (refer to Section 1.1).

Figure M.2 shows the time series of magnitudes ≥ 2.0, and Fig. M.3 shows the corresponding epicenters of earthquakes in the above-described spatiotemporal domain. Large red circles indicate the events with a magnitude ≥ 5.0. The top five cells in terms of the number of events with a magnitude ≥ 5.0 are colored; region (1), $115° - 115.5°W$, $32° - 32.5°N$, region (2), $116° - 116.5°W$, $34° - 34.5°N$, region (3), $118.5° - 119°W$, $34° - 34.5°N$, region (4), $117.5° - 118°W$, $35.5° - 36.0°N$, and region (5), $116.5° - 117°W$, $34° - 34.5°N$. We performed the analysis for the spatiotemporal marked point process shown in Figs. M.2 and M.3. Note that the cut-off magnitudes were set to $(m, M) = (2.0, 4.0)$ and $(m, M) = (2.0, 5.0)$. Also, when computing $p_{Mm}(\tau_M|\boldsymbol{x}, \tau_m)$, $\boldsymbol{x}$'s in the same cell shown in Fig. M.3 were regarded to be identical.

Figure M.4 shows the results of $p_M(\tau_M)$, $p_{Mm}(\tau_M|\tau_m)$, and $p_{Mm}(\tau_M|\boldsymbol{x}, \tau_m)$. $p_{Mm}(\tau_M|\boldsymbol{x}, \tau_m)$ for $\boldsymbol{x}$ in the regions (1) to (5) are shown. $p_{Mm}(\tau_M|\tau_m)$ and $p_{Mm}(\tau_M|\boldsymbol{x}, \tau_m)$ are shown in the range $\tau_m < \tau_M$ and are compared only in the range $10^{-1} \leq \tau_m < 10^{-0.9}$ as representative.

Although there is no significant peak in the inter-event time distribution ($p_M(\tau_M)$), a peak appears in the inverse probability density function $p_{Mm}(\tau_M|\tau_m)$. This shows that the additional information on an inter-event time interval at the lower cut-off magnitude m can contribute to limiting the range of the time to occur the next large earthquake (this can also be confirmed in Chapter 5). On the other hand, the densities $p_{Mm}(\tau_M|\boldsymbol{x}, \tau_m)$ are not notably different from $p_{Mm}(\tau_M|\tau_m)$, which indicates that the spatial information on small earthquakes is not able to contribute significantly to the estimate of the

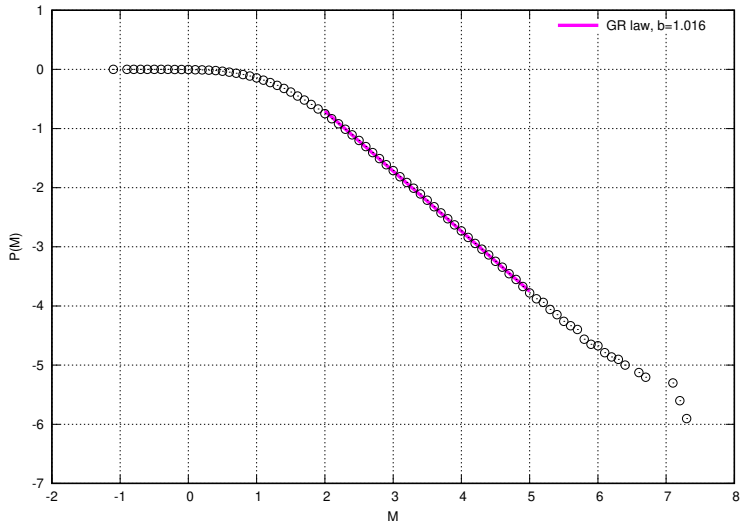

Figure M.1: $P(M)$ for the events in the seismic catalog. The magenta-colored line is the fitting with the GR law in the range $M \in [2.0, 5.0]$.

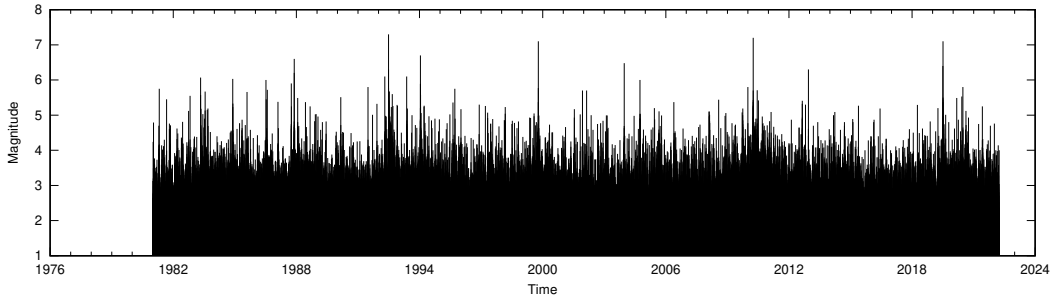

Figure M.2: Time series of magnitudes for the events with a magnitude ≥ 2.0 in the seismic catalog.

large earthquakes' timing. This result implies that it is necessary to be more careful in setting the spatial area S and dividing it into tiny cells. Concerning the former, one may restrict S to the vicinity of a fault zone to clarify the correlations between events. Concerning the latter, it is necessary to divide S into tiny enough cells so that the correspondence between an epicenter $\boldsymbol{x}$ and a cell is sufficiently accurate. Thus, our preliminary analysis suggested that the numerical examination of the spatiotemporal Bayesian approach using seismic catalogs requires careful consideration of the spatial setting and the method of its division.

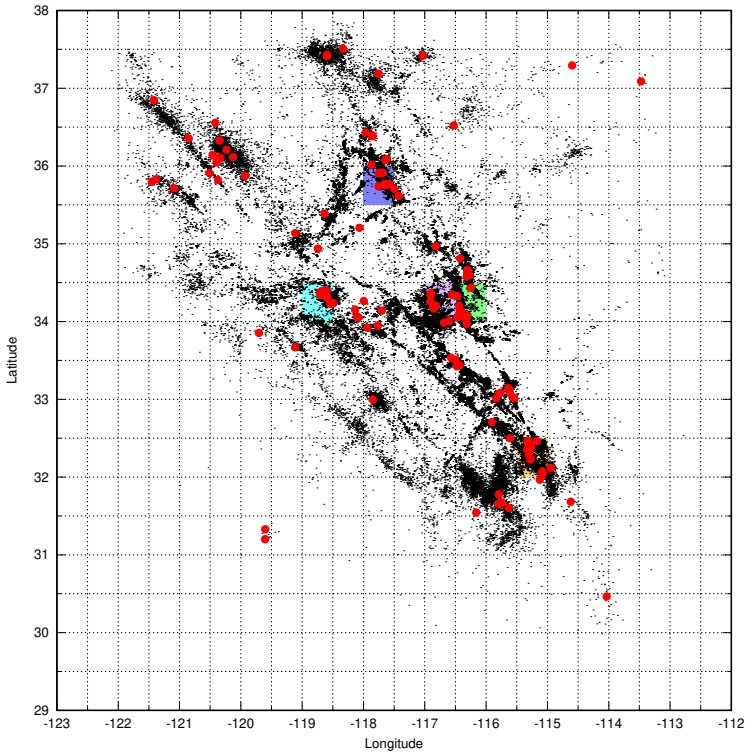

Figure M.3: Epicenters of earthquakes with magnitude ≥ 2.0 recorded in the seismic catalog in Southern California [18, 19], from 01/01/1981 to 03/31/2022. The spatial area S is divided into cells by grid lines drawn at 0.5° increments in longitude and latitude. Five cells are colored; (Orange) region (1), (Green) region (2), (Cyan) region (3), (Blue) region (4), and (Purple) region (5).

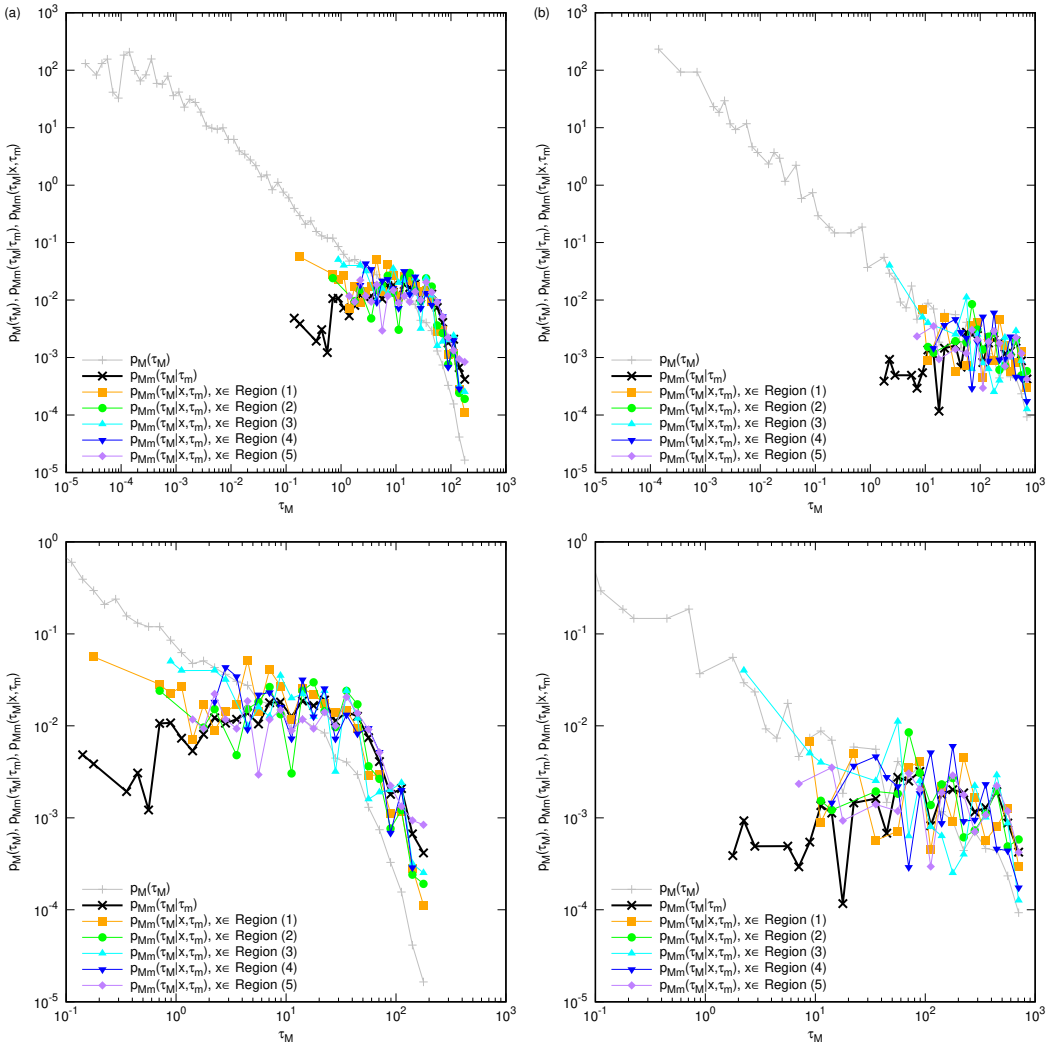

Figure M.4: Comparison between $p_M(\tau_M)$, $p_{Mm}(\tau_M|\tau_m)$, and $p_{Mm}(\tau_M|\boldsymbol{x},\tau_m)$, with the cut-off magnitudes (a) $(m, M) = (2.0, 4.0)$, and (b) $(m, M) = (2.0, 5.0)$. Lower panels are enlarged versions of the upper panels. The figure shows the results of $p_{Mm}(\tau_M|\boldsymbol{x},\tau_m)$ for $\boldsymbol{x}$ in Region (1) – Region (5) indicated in Fig. M.3. $p_{Mm}(\tau_M|\tau_m)$ and $p_{Mm}(\tau_M|\boldsymbol{x},\tau_m)$ are shown in cross sections for the range $10^{-1} \leq \tau_m < 10^{-0.9}$.

Appendix N

Derivation of the Conditional Probability in the Time Series with Weak Inter-event Corrlation

This Appendix is based on the Appendix in Ref. [4], which is paper 4 in the list of author's papers.

This Appendix shows the derivation of the conditional probability (Eq. (5.22)) by substituting Eqs. (5.17) and (5.19) into Eq. (3.10). The denominator of Eq. (3.10) is, from the property of the negative binomial distribution:

$$\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M) = A_{\Delta m}\zeta(y) + 1. \tag{N.1}$$

On the other hand, the numerator is:

$$\begin{aligned}\sum_{i=1}^{\infty} i\rho_{mM}(\tau_m|i,\tau_M)\Psi_{mM}(i|\tau_M) &= \delta(\tau_M-\tau_m)\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)} \\ &+ \theta(\tau_M-\tau_m)\sum_{i=2}^{\infty}\frac{i(i-1)}{\tau_M}\left(1-\frac{x}{y}\right)^{i-2}\left(1+dh_{mM}^{(i)}(\tau_m,\tau_M)\right) \\ &\times \frac{\Gamma\left(i-1+B_{\Delta m}\zeta(y)\right)}{(i-1)!\Gamma\left(B_{\Delta m}\zeta(y)\right)}\left(\frac{B_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{i-1}.\end{aligned} \tag{N.2}$$

The summation without $dh_{mM}^{(i)}(\tau_m,\tau_M)$ of the r.h.s. of Eq. (N.2) is transformed as follows:

$$\begin{aligned}&\sum_{i=2}^{\infty}\frac{i(i-1)}{\tau_M}\left(1-\frac{\tau_m}{\tau_M}\right)^{i-2}\frac{\Gamma\left(i-1+B_{\Delta m}\zeta(y)\right)}{(i-1)!\Gamma\left(B_{\Delta m}\zeta(y)\right)}\left(\frac{B_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{i-1} \\ &= \frac{\left(\frac{B_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)}{\tau_M\Gamma\left(B_{\Delta m}\zeta(y)\right)}\left(\sum_{i=0}^{\infty}\frac{\Gamma(i+1+D)}{i!}C^{i+1}+2\sum_{i=0}^{\infty}\frac{\Gamma(i+D)}{i!}C^{i}\right),\end{aligned} \tag{N.3}$$

where,

$$\begin{aligned}C &:= \frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\left(1-\frac{x}{y}\right). \\ D &:= 1+B_{\Delta m}\zeta(y).\end{aligned}$$

The first infinite series in the r.h.s. of Eq. (N.3):

$$S_1 = 2\sum_{i=0}^{\infty}a_i, \text{ where } a_i = \frac{\Gamma(i+D)}{i!}C^i, \tag{N.4}$$

satisfies the ratio test as follows:

$$\lim_{i\to\infty}\left|\frac{a_{i+1}}{a_i}\right| = \lim_{i\to\infty}\left|\frac{i+D}{i+1}\right||C| = \left|\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right|\left|1-\frac{x}{y}\right| < 1.$$

Therefore, the first infinite series converges as follows:

$$\begin{aligned} S_1 &= 2\Gamma(D)\left[1+\frac{D}{1!}C+\frac{D(D+1)}{2!}C^2+\cdots\right] \\ &= 2\Gamma(D)(1-C)^{-D}. \end{aligned} \tag{N.5}$$

The second infinite series,

$$S_2 = \sum_{i=0}^{\infty} b_i, \text{ where } b_i = \frac{\Gamma(i+1+D)}{i!}C^{i+1}, \tag{N.6}$$

also satisfies the ratio test as follows.

$$\lim_{i\to\infty}\left|\frac{b_{i+1}}{b_i}\right| = \lim_{i\to\infty}\left|\frac{i+1+D}{i+1}\right||C| = \left|\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right|\left|1-\frac{x}{y}\right| < 1.$$

Equation (N.6) can be transformed in the following way:

$$\begin{aligned} S_2 &= C\sum_{i=0}^{\infty}\frac{(i+D)\Gamma(i+D)}{i!}C^i \\ &= C\left(\sum_{i=0}^{\infty}\frac{\Gamma(i+1+D)}{i!}C^{i+1} + D\sum_{i=0}^{\infty}\frac{\Gamma(i+D)}{i!}C^i\right) \\ &= CS_2 + CD\Gamma(D)(1-C)^{-D} \end{aligned}$$

Therefore,

$$S_2 = CD\Gamma(D)(1-C)^{-D-1}. \tag{N.7}$$

From Eqs. (N.5) and (N.7),

$$\begin{aligned} S_1+S_2 = B_{\Delta m}\zeta(y)\,\Gamma\left(B_{\Delta m}\zeta(y)\right)&\left[\frac{(A_{\Delta m}+B_{\Delta m})y}{A_{\Delta m}x+B_{\Delta m}y}\right]^{B_{\Delta m}\zeta(y)+1} \\ &\times\frac{[A_{\Delta m}B_{\Delta m}(y-x)\zeta(y)+A_{\Delta m}(x+y)+2B_{\Delta m}y]}{(A_{\Delta m}x+B_{\Delta m}y)}. \end{aligned}$$

Equation (N.3) is rewritten as follows:

$$\frac{A_{\Delta m}B_{\Delta m}\zeta(y)}{\langle\tau_M\rangle}\left(\frac{B_{\Delta m}y}{A_{\Delta m}x+B_{\Delta m}y}\right)^{B_{\Delta m}\zeta(y)}\frac{\{A_{\Delta m}B_{\Delta m}(y-x)\zeta(y)+A_{\Delta m}(x+y)+2B_{\Delta m}y\}}{(A_{\Delta m}x+B_{\Delta m}y)^2}.$$

Therefore, the conditional probability when $dh^{(i)}_{mM}(\tau_m,\tau_M)\equiv 0$ can be derived as follows:

$$\begin{aligned} p_{mM}(\tau_m|\tau_M) = (A_{\Delta m}\zeta(y)+1)^{-1}&\left\{\delta(\tau_M-\tau_m)\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\right. \\ &+\theta(\tau_M-\tau_m)\frac{[A_{\Delta m}B_{\Delta m}(y-x)\zeta(y)+A_{\Delta m}(x+y)+2B_{\Delta m}y]}{(A_{\Delta m}x+B_{\Delta m}y)^2} \\ &\left.\times\frac{A_{\Delta m}B_{\Delta m}\zeta(y)}{\langle\tau_M\rangle}\left(\frac{B_{\Delta m}y}{A_{\Delta m}x+B_{\Delta m}y}\right)^{B_{\Delta m}\zeta(y)}\right\}. \end{aligned} \tag{N.8}$$

Finally, Eq. (5.22) is obtained by performing the variable transformation $\tau_m \mapsto x$.

Appendix O

Derivation of the Scaling Function of the Inter-event Time Distribution

This Appendix is based on the Appendix in Ref. [4], which is paper 4 in the list of author's papers.

This Appendix shows the derivation of the scaling function of the inter-event time distribution (Eq. (5.27)). First, the conditional probability for $\Delta m \to 0$ is derived. From Eq. (5.18), the denominator of Eq. (3.10) is:

$$\sum_{i=1}^{\infty} i \frac{\left(A_{\Delta m}\zeta\left(y\right)\right)^{(i-1)}}{(i-1)!} e^{-A_{\Delta m}\zeta(y)} = A_{\Delta m}\zeta\left(y\right) + 1. \tag{O.1}$$

On the other hand, the numerator of Eq. (3.10) is, with $dh_{mM}^{(i)}(\tau_m, \tau_M) \equiv 0$,

$$\theta(\tau_M - \tau_m) \sum_{i=2}^{\infty} i \frac{(i-1)}{\tau_M} \left(1 - \frac{x}{y}\right)^{i-2} \frac{\left(A_{\Delta m}\zeta\left(y\right)\right)^{(i-1)}}{(i-1)!} e^{-A_{\Delta m}\zeta(y)} + \delta(\tau_M - \tau_m) e^{-A_{\Delta m}\zeta(y)}. \tag{O.2}$$

The summation in Eq. (O.2) is rewritten as follows:

$$\begin{aligned}
&\sum_{i=0}^{\infty} \frac{(i+2)}{\tau_M} \left(1 - \frac{x}{y}\right)^{i} \frac{\left(A_{\Delta m}\zeta\left(y\right)\right)^{(i+1)}}{i!} e^{-A_{\Delta m}\zeta(y)} \\
&\quad = \frac{1}{\tau_M} A_{\Delta m}\zeta\left(y\right) e^{-A_{\Delta m}\zeta(y)\frac{x}{y}} \sum_{i=0}^{\infty} (i+2) \frac{\left[A_{\Delta m}\zeta\left(y\right)\left(1 - \frac{x}{y}\right)\right]^{i}}{i!} e^{-A_{\Delta m}\zeta(y)\left(1-\frac{x}{y}\right)} \\
&\quad = \frac{1}{\tau_M} A_{\Delta m}\zeta\left(y\right) e^{-A_{\Delta m}\zeta(y)\frac{x}{y}} \left[A_{\Delta m}\zeta\left(y\right)\left(1 - \frac{x}{y}\right) + 2\right].
\end{aligned}$$

Therefore, the conditional probability is:

$$\begin{aligned}
p_{mM}(\tau_m|\tau_M) = \left(A_{\Delta m}\zeta\left(y\right) + 1\right)^{-1} \Big\{ & e^{-A_{\Delta m}\zeta(y)} \delta(\tau_M - \tau_m) \\
& + \frac{1}{\tau_M} A_{\Delta m}\zeta\left(y\right) e^{-A_{\Delta m}\zeta(y)\frac{x}{y}} \left[A_{\Delta m}\zeta\left(y\right)\left(1 - \frac{x}{y}\right) + 2\right] \theta(\tau_M - \tau_m) \Big\}.
\end{aligned} \tag{O.3}$$

Equation (2.3) is rewritten into the relationship between the scaling functions (Eq. (O.4)) by Eqs. (1.8), (O.1) and (O.3) with $N_m/N_M = \langle\tau_M\rangle/\langle\tau_m\rangle = 10^{b\Delta m}$.

$$10^{2b\Delta m} f\left(10^{b\Delta m} x\right) = e^{-A_{\Delta m}\zeta(x)} f(x) + A_{\Delta m} \int_x^{\infty} \frac{\zeta(y)}{y} e^{-A_{\Delta m}\zeta(y)\frac{x}{y}} \left[A_{\Delta m}\zeta(y)\left(1 - \frac{x}{y}\right) + 2\right] f(y) dy. \tag{O.4}$$

Each term in Eq. (O.4) is then followed by operating $\lim_{\Delta m \to 0} \partial_{\Delta m}$:

$$\begin{aligned}
&\partial_{\Delta m} 10^{2b\Delta m} f\left(10^{b\Delta m} x\right) \xrightarrow[\Delta m \to 0]{} b \ln 10 \left(2f(x) + x f'(x)\right), \\
&\partial_{\Delta m} e^{-A_{\Delta m}\zeta(x)} f(x) \qquad \xrightarrow[\Delta m \to 0]{} -b \ln 10 \zeta(x) f(x), \\
&\partial_{\Delta m} A_{\Delta m} \int_x^\infty \frac{\zeta(y)}{y} e^{-A_{\Delta m}\zeta(y)\frac{x}{y}} \left[A_{\Delta m}\zeta(y)\left(1 - \frac{x}{y}\right) + 2\right] f(y) dy \\
&\qquad\qquad \xrightarrow[\Delta m \to 0]{} 2b \ln 10 \int_x^\infty \frac{\zeta(y)}{y} f(y) dy.
\end{aligned}$$

Hence,

$$2f(x) + x f'(x) = -\zeta(x) f(x) + 2\int_x^\infty \frac{\zeta(y)}{y} f(y) dy. \tag{O.5}$$

Equation (O.5) indicates the essential equivalence between $f(y)$ and $\zeta(y)$.

Equation (O.5) is simplified using $G(x)$ as defined in Eq. (O.6), as follows:

$$2G(x) = -xG'(x),$$

where,

$$G(x) := f(x) - \int_x^\infty \frac{\zeta(y)}{y} f(y) dy, \tag{O.6}$$

Therefore, with a constant κ:

$$G(x) = \frac{\kappa}{x^2},$$

or,

$$f'(x) + \frac{\zeta(x)}{x} f(x) = -\frac{2\kappa}{x^3}.$$

Change the variable from x to y, and the solution is with a constant y_0,

$$f(y) = \left(-\int_{y_0}^y \frac{2\kappa}{u^3} e^{\int_{y_0}^u \frac{\zeta(v)}{v} dv} du + f(y_0)\right) e^{-\int_{y_0}^y \frac{\zeta(v)}{v} dv}.$$

In particular, when $\kappa = 0$:

$$f(y) = f(y_0) \exp\left(-\int_{y_0}^y \frac{\zeta(u)}{u} du\right).$$

Using Eq. (5.13) suggested by the catalog analysis as $\zeta(y)$,

$$f(y) = \begin{cases} f(y_0) y_0^\beta e^{\alpha y_0} y^{-\beta} e^{-\alpha y} & (\text{if } y \geq y_0), \\ f(y_0) e^{\gamma y_0} e^{-\gamma y} & (\text{if } y < y_0), \end{cases} \tag{O.7}$$

where $f(y_0)$ is determined by the normalization condition as:

$$f(y_0) = \left(\frac{e^{\gamma y_0} - 1}{\gamma} + \frac{y_0^\beta e^{\alpha y_0}}{\alpha^{1-\beta}} \Gamma\left(1 - \beta, \alpha y_0\right)\right)^{-1}. \tag{O.8}$$

Hence, by Eqs. (O.7) and (O.8), the scaling function of the inter-event time distribution is derived as Eq. (5.27).

Appendix P

Derivation of the Inverse Probability in the Time Series with Weak Inter-event Correlation

This Appendix is based on the Appendix in Ref. [4], which is paper 4 in the list of author's papers.

This Appendix shows the derivation of the inverse probability from Eq. (3.7) with Eqs. (5.28) and (N.8).

From Eqs. (3.5) and (5.28), $z_m(\tau_m)$ and $z_M(\tau_M)$ are:

$$z_m(\tau_m) = \frac{\alpha^{1-\beta}}{\Gamma(1-\beta)} \frac{1}{\langle \tau_m \rangle} \left\{ (A_{\Delta m}+1)x \right\}^{1-\beta} e^{-\alpha(A_{\Delta m}+1)x}, \tag{P.1}$$

$$z_M(\tau_M) = \frac{\alpha^{1-\beta}}{\Gamma(1-\beta)} \frac{1}{\langle \tau_M \rangle} y^{1-\beta} e^{-\alpha y}. \tag{P.2}$$

Also, from Eqs. (3.5) and (N.8), $z_{mM}(\tau_m|\tau_M)$ is:

$$\begin{aligned} z_{mM}(\tau_m|\tau_M) = \left(\frac{x}{y}\right) \Bigg\{ & \delta(\tau_M - \tau_m) \left(\frac{A_{\Delta m}}{A_{\Delta m} + B_{\Delta m}} \right)^{B_{\Delta m}\zeta(y)} \\ & + \theta(\tau_M - \tau_m) \frac{[A_{\Delta m} B_{\Delta m}(y-x)\zeta\,(y) + A_{\Delta m}(x+y) + 2B_{\Delta m} y]}{(A_{\Delta m} x + B_{\Delta m} y)^2} \\ & \times \frac{A_{\Delta m} B_{\Delta m} \zeta(y)}{\langle \tau_M \rangle} \left(\frac{B_{\Delta m} y}{A_{\Delta m} x + B_{\Delta m} y} \right)^{B_{\Delta m}\zeta(y)} \Bigg\}. \end{aligned} \tag{P.3}$$

The inverse probability for $dh^{(i)}_{mM}(\tau_m, \tau_M) \equiv 0$ is derived by substituting Eqs. (P.1), (P.2), and (P.3) into Eq. (3.7):

$$\begin{aligned} p_{Mm}(\tau_M|\tau_m) = \frac{\left(\frac{y}{x}\right)^{-\beta} e^{-\alpha(y-(A_{\Delta m}+1)x)}}{(A_{\Delta m}+1)^{2-\beta}} \Bigg\{ & \delta(\tau_M - \tau_m) \left(\frac{A_{\Delta m}}{A_{\Delta m} + B_{\Delta m}} \right)^{B_{\Delta m}\zeta(y)} \\ & + \theta(\tau_M - \tau_m) \frac{[A_{\Delta m} B_{\Delta m}(y-x)\zeta(y) + A_{\Delta m}(x+y) + 2B_{\Delta m} y]}{(A_{\Delta m} x + B_{\Delta m} y)^2} \frac{A_{\Delta m} B_{\Delta m} \zeta(y)}{\langle \tau_M \rangle} \left(\frac{B_{\Delta m} y}{A_{\Delta m} x + B_{\Delta m} y} \right)^{B_{\Delta m}\zeta(y)} \Bigg\}. \end{aligned} \tag{P.4}$$

Finally, Eq. (5.29) is derived by performing the variable transformation $\tau_M \mapsto y$ to Eq. (P.4).

Appendix Q

Supplementary Information for Chapter 4

This Appendix is based on the Supplemental Material of Ref. [4], which is paper 4 in the list of author's papers.

This Appendix provides the figures for the results related to the main text yet not shown. See the caption of the corresponding figure in the main text for detailed descriptions of the figures.

Q.1 Figures S1- for $\Psi_{mM}(i|\tau_M)$, $\Psi_{\Delta m}(i|y)$, and the Negative Binomial Distribution

Figures S1-(1) to S1-(7) show $\Psi_{mM}(i|\tau_M)$, $\Psi_{\Delta m}(i|y)$, and the negative binomial distributions for each time series with Δm values: (a) $\Delta m = 0.5$, (b) $\Delta m = 1.0$, and (c) $\Delta m = 1.5$, for Fig. S1-(1), S1-(3), and S1-(5) – S1-(7); (a) $\Delta m = 0.2$, (b) $\Delta m = 0.4$, (c) $\Delta m = 0.6$, and (d) $\Delta m = 0.8$ for Fig. S1-(2) and S1-(4).

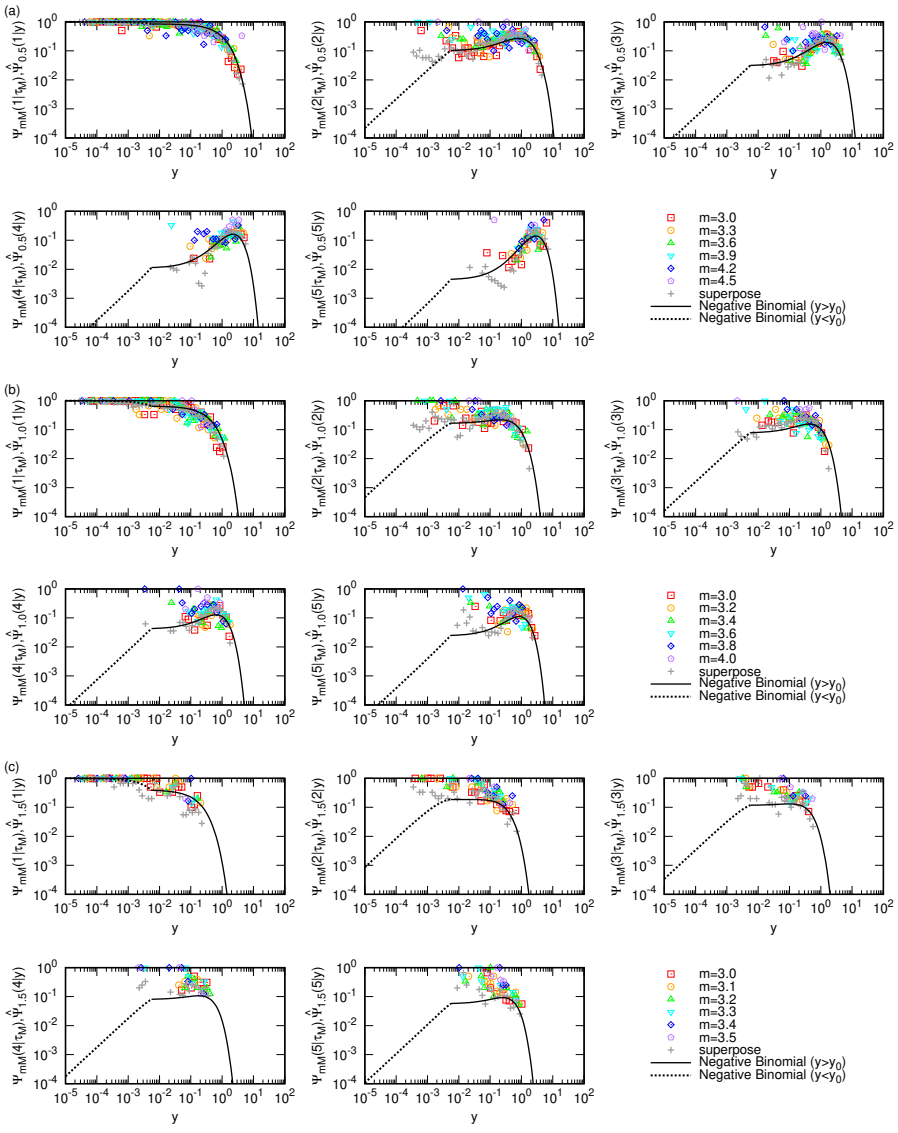

Figure S1-(1): JS

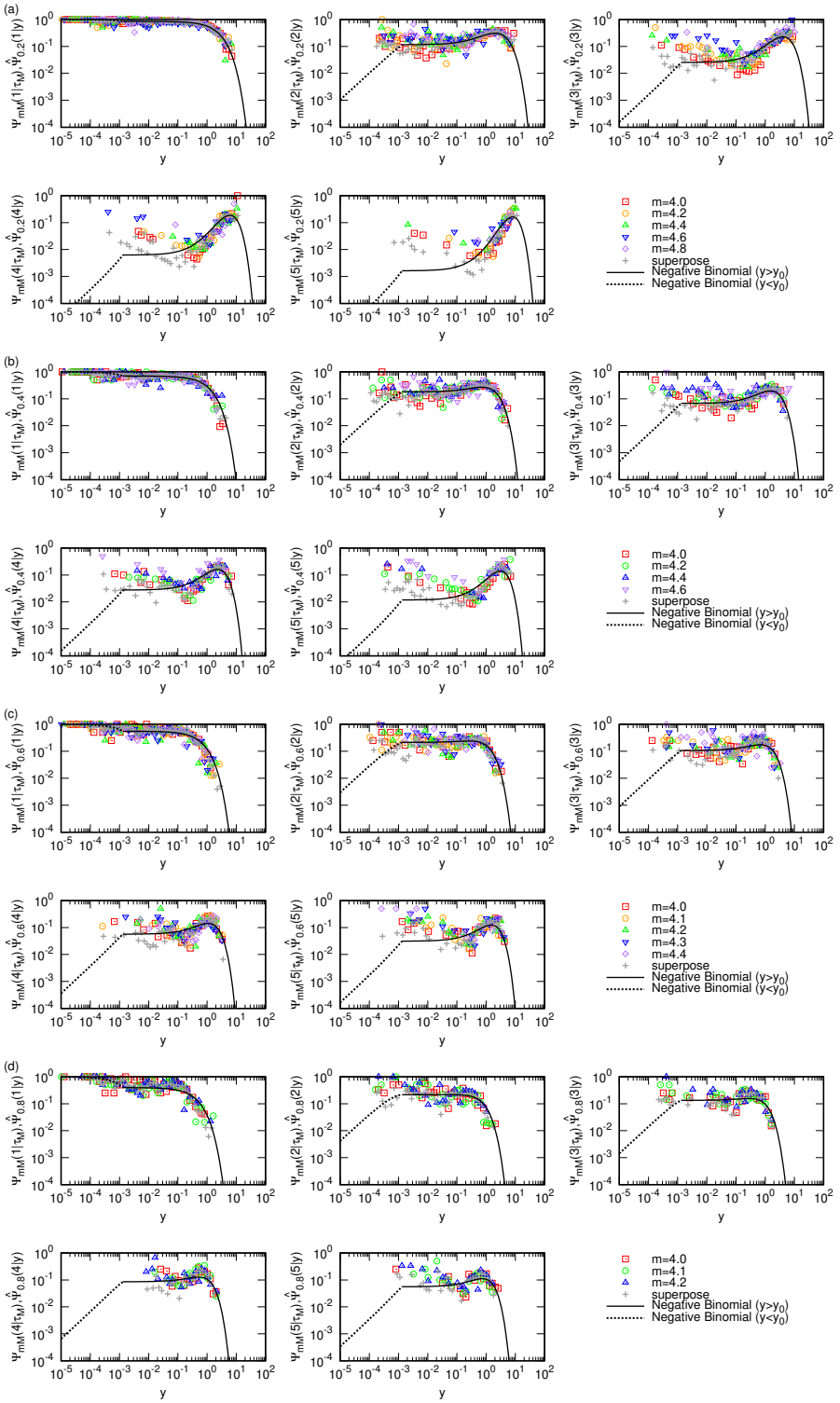

Figure S1-(2): JA

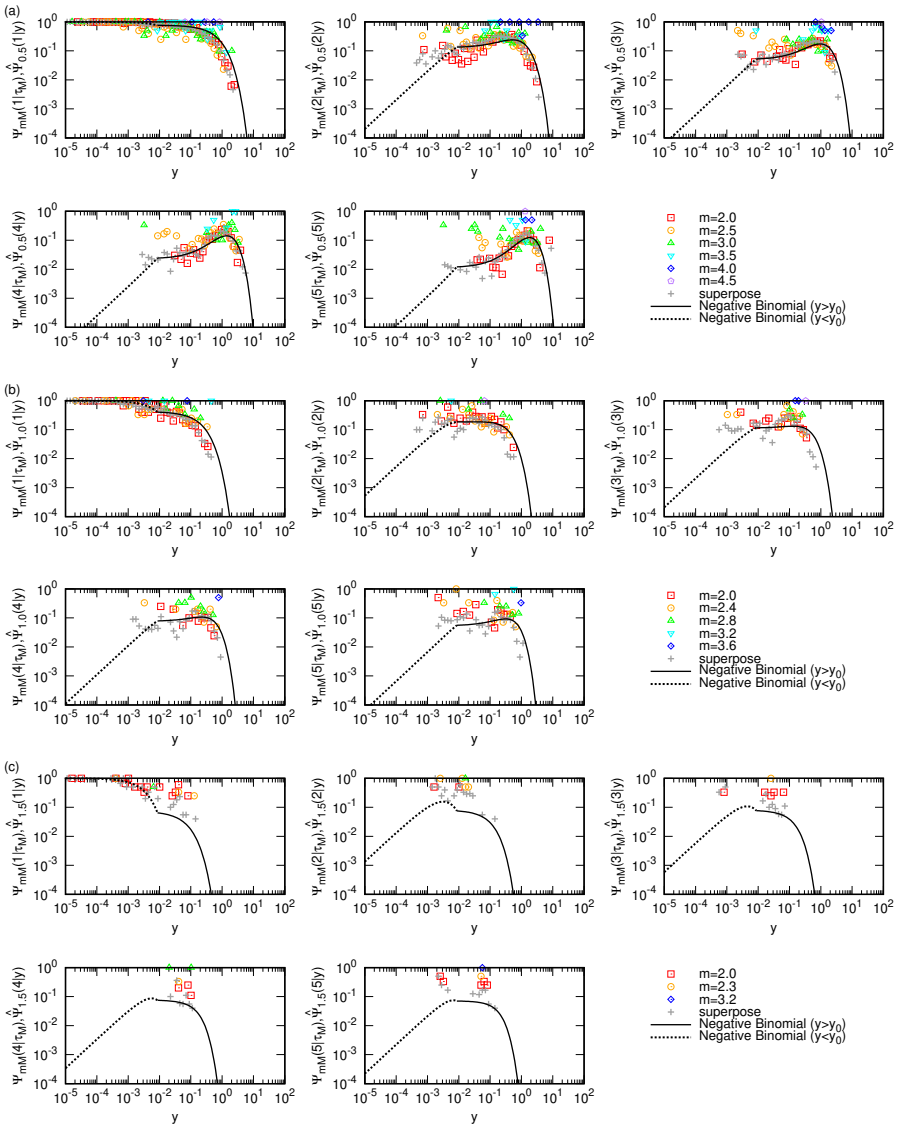

Figure S1-(3): SCA1

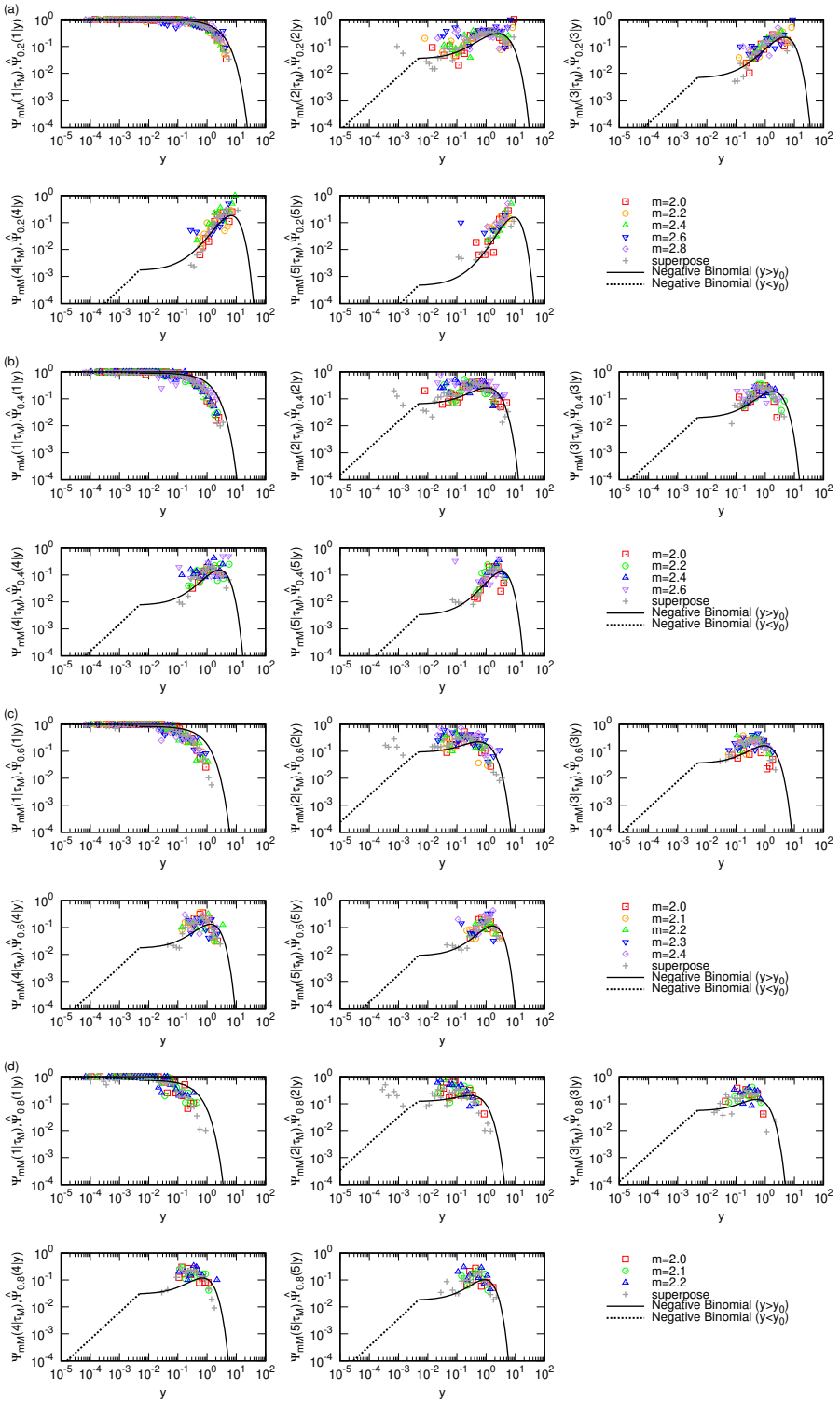

Figure S1-(4): SCA2

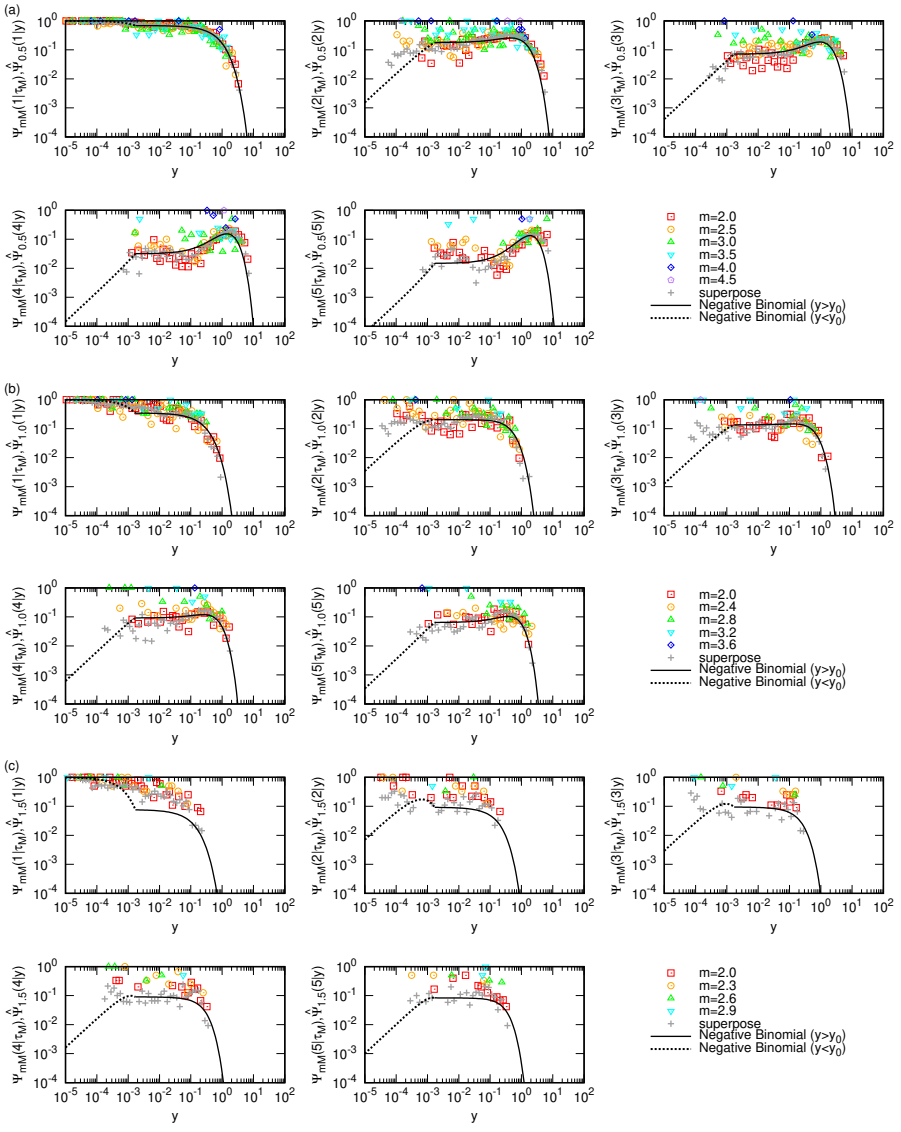

Figure S1-(5): SCA3

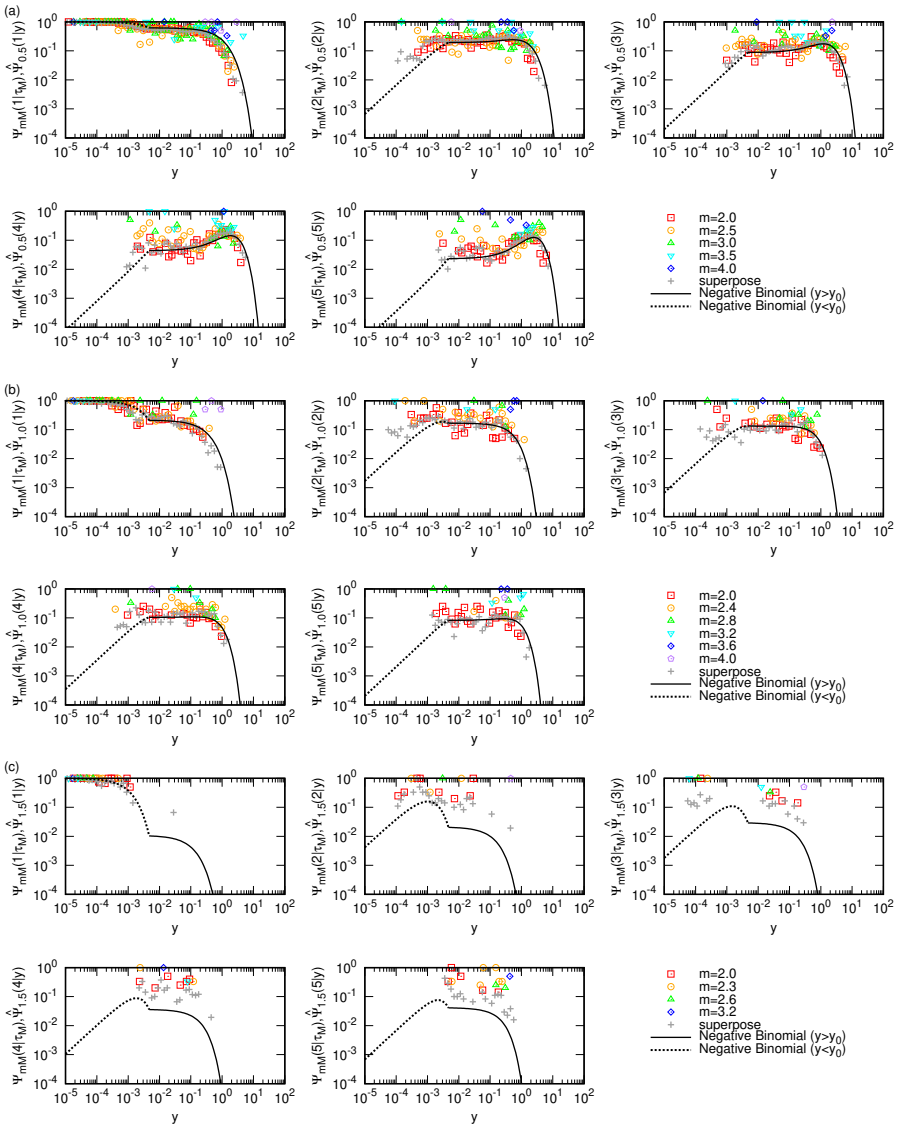

Figure S1-(6): SCS1

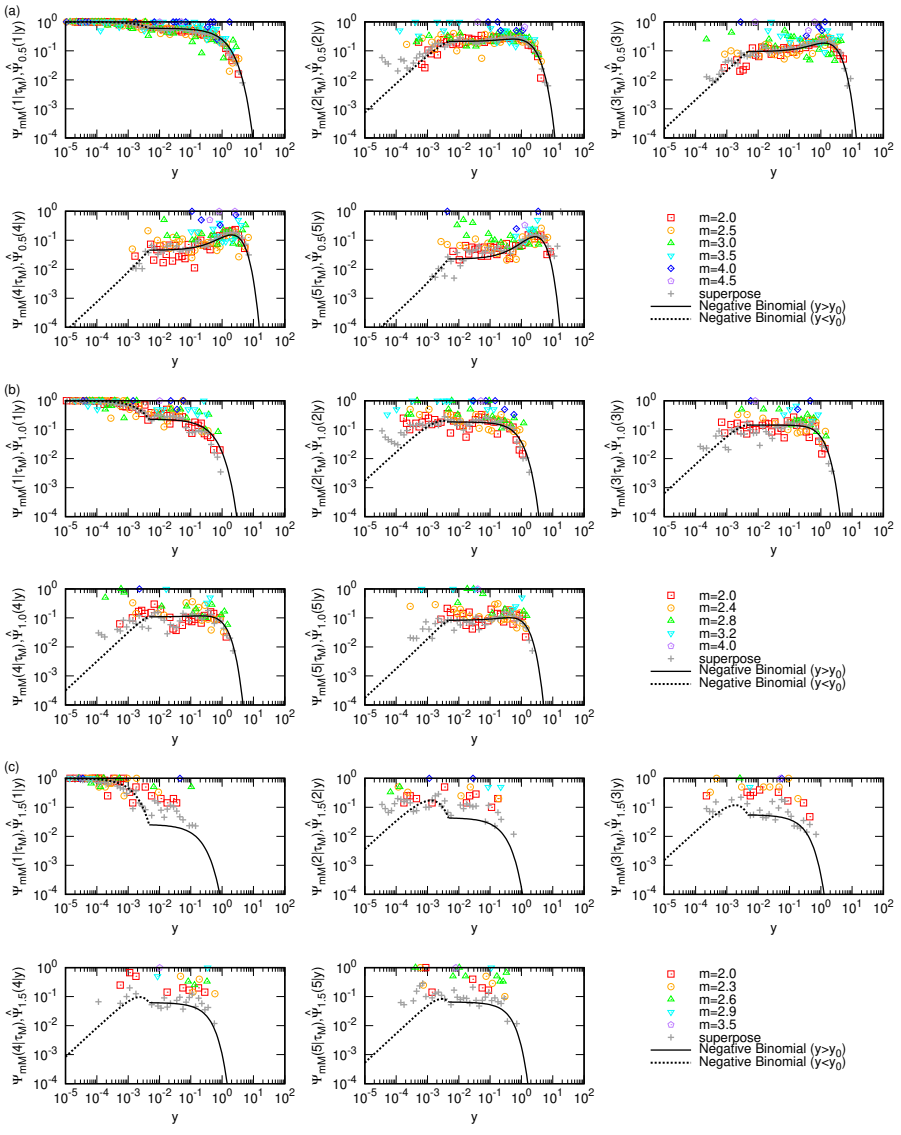

Figure S1-(7): SCS2

Q.2 Figures S2- for $\rho_{mM}(\tau_m|2,\tau_M)$ and $P_{mM}(\tau_m/\tau_M|2,\tau_M)$

Figures S2-(1) to S2-(7) show $\rho_{mM}(\tau_m|2,\tau_M)$ and $P_{mM}(\tau_m/\tau_M|2,\tau_M)$ for each time series. Δm values are: (a) $\Delta m = 0.5$, (b) $\Delta m = 1.0$, and (c) $\Delta m = 1.5$ for Figs. S2-(1), S2-(3), and S2-(5) – S2-(7); (a) $\Delta m = 0.2$, (b) $\Delta m = 0.4$, (c) $\Delta m = 0.6$, and (d) $\Delta m = 0.8$ for Figs. S2-(2) and S2-(4).

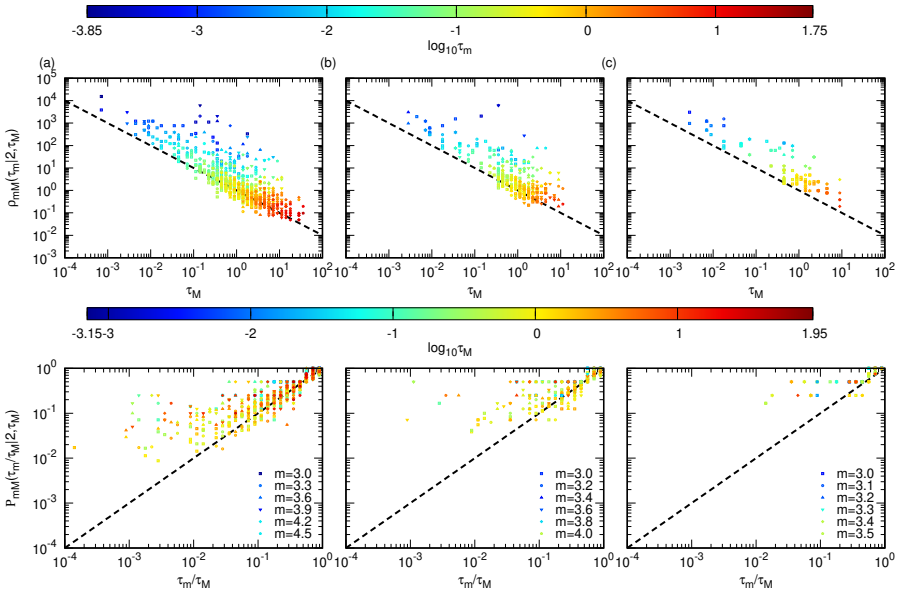

Figure S2-(1): JS

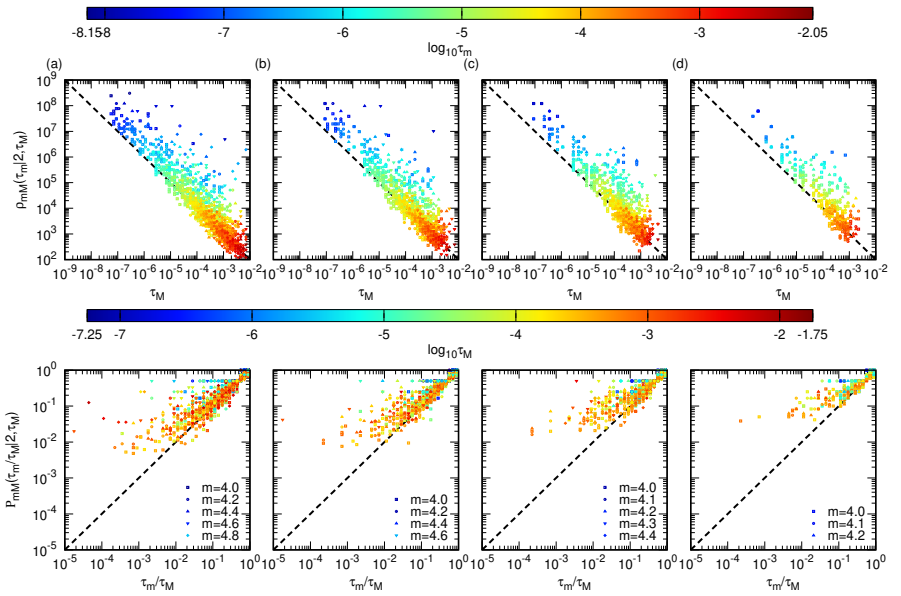

Figure S2-(2): JA

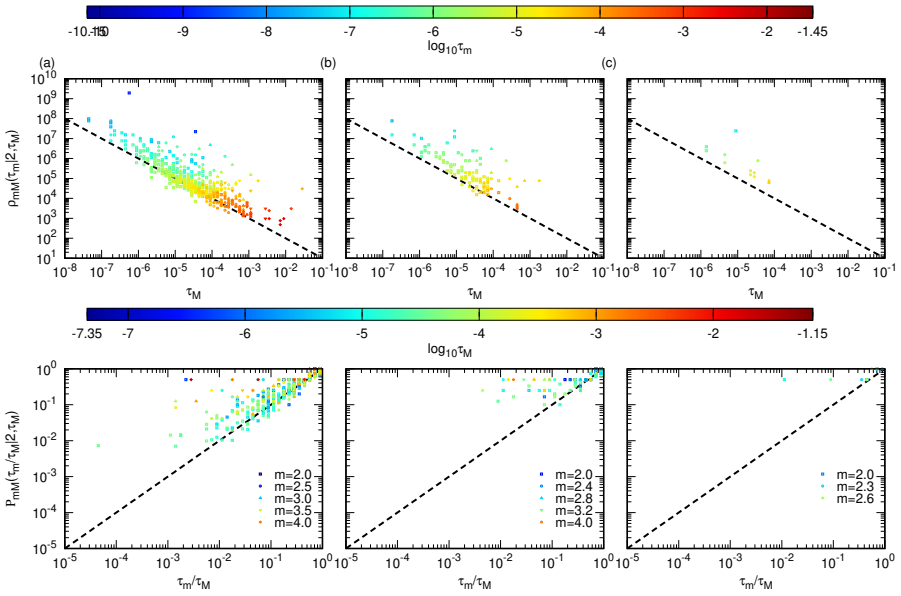

Figure S2-(3): SCA1

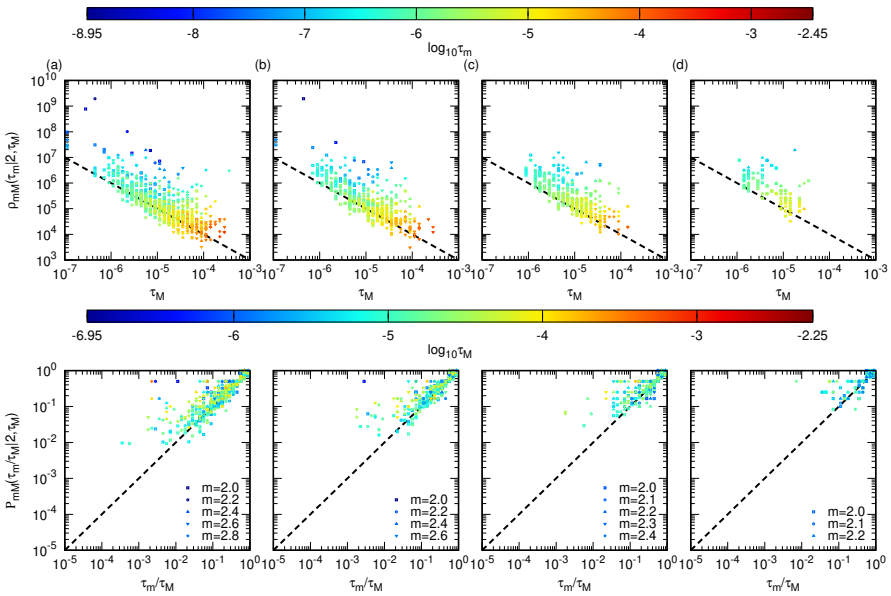

Figure S2-(4): SCA2

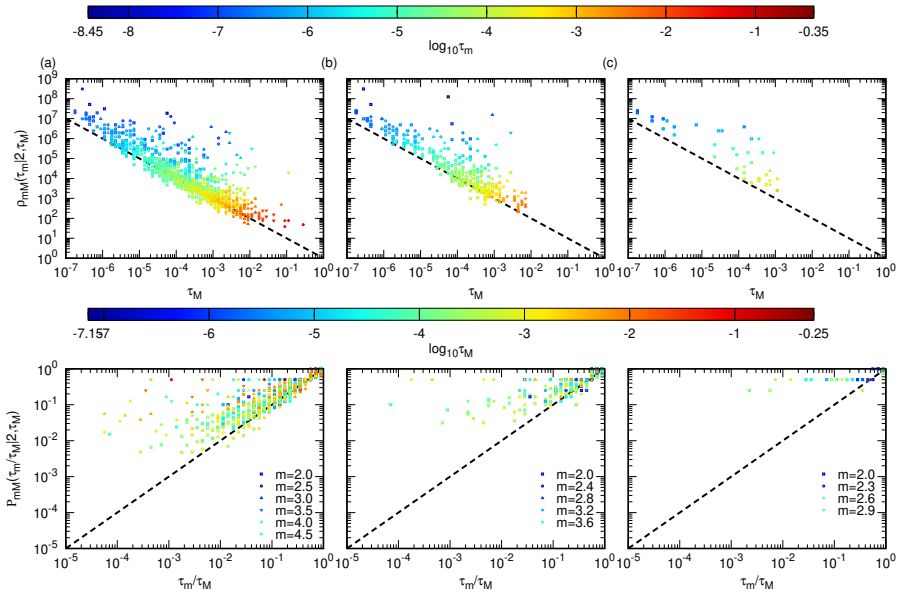

Figure S2-(5): SCA3

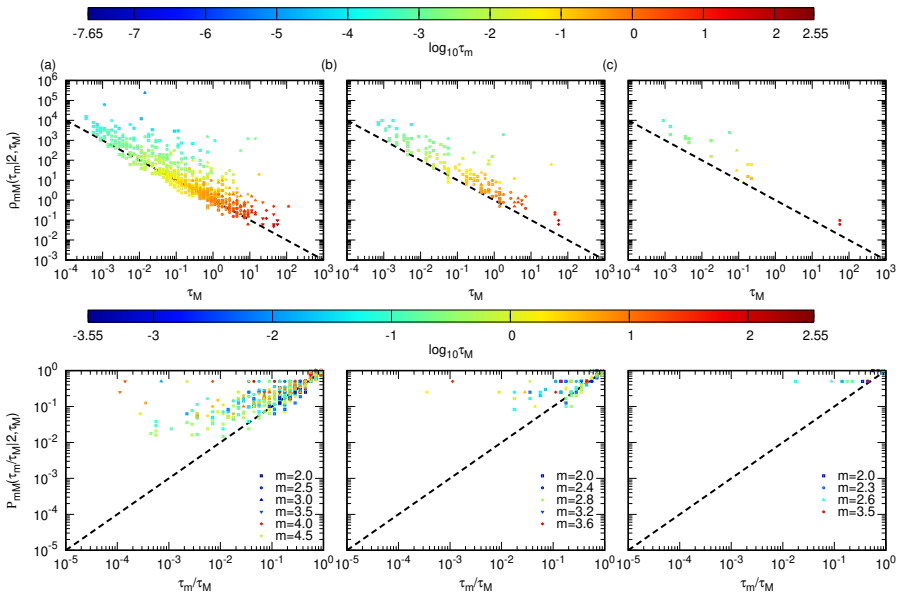

Figure S2-(6): SCS1

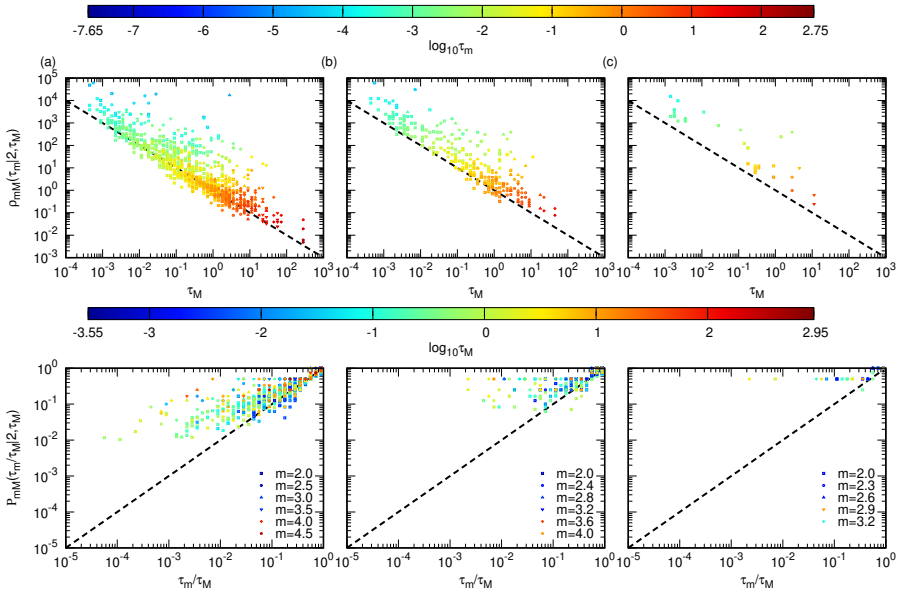

Figure S2-(7): SCS2

Q.3 Figures S3- for $\hat{P}_{mM}(x|y)$ and $\overline{P}_{\Delta m}(x|y)$

Figures S3-(1) to S3-(7) show $\hat{P}_{mM}(x|y)$ and $\overline{P}_{\Delta m}(x|y)$ with: (a) $\Delta m = 0.4$, (b) $\Delta m = 0.8$, (c) $\Delta m = 1.2$, and (d) $\Delta m = 1.6$ for Fig. S3-(1); (a) $\Delta m = 0.2$, (b) $\Delta m = 0.4$, (c) $\Delta m = 0.6$, and (d) $\Delta m = 0.8$ for Fig. S3-(2); (a) $\Delta m = 0.6$, (b) $\Delta m = 1.2$, (c) $\Delta m = 1.8$, and (d) $\Delta m = 2.4$ for Figs. S3-(3) – S3-(7). For Figs. S3-(3) – S3-(7), the y ranges to display are chosen from j's such that at least $n_{\min}$ of the (m, M) pairs indicated in the figure's legend fall within the range $(10^{j\Delta\sigma}, 10^{(j+1)\Delta\sigma})$, where $n_{\min}$ is 5 for Figs. S3-(3), S3-(5) – S3-(7) and 4 for Fig. S3-(4). Note that when the determined number of y ranges is fewer than four, the identical y ranges are displayed multiple times.

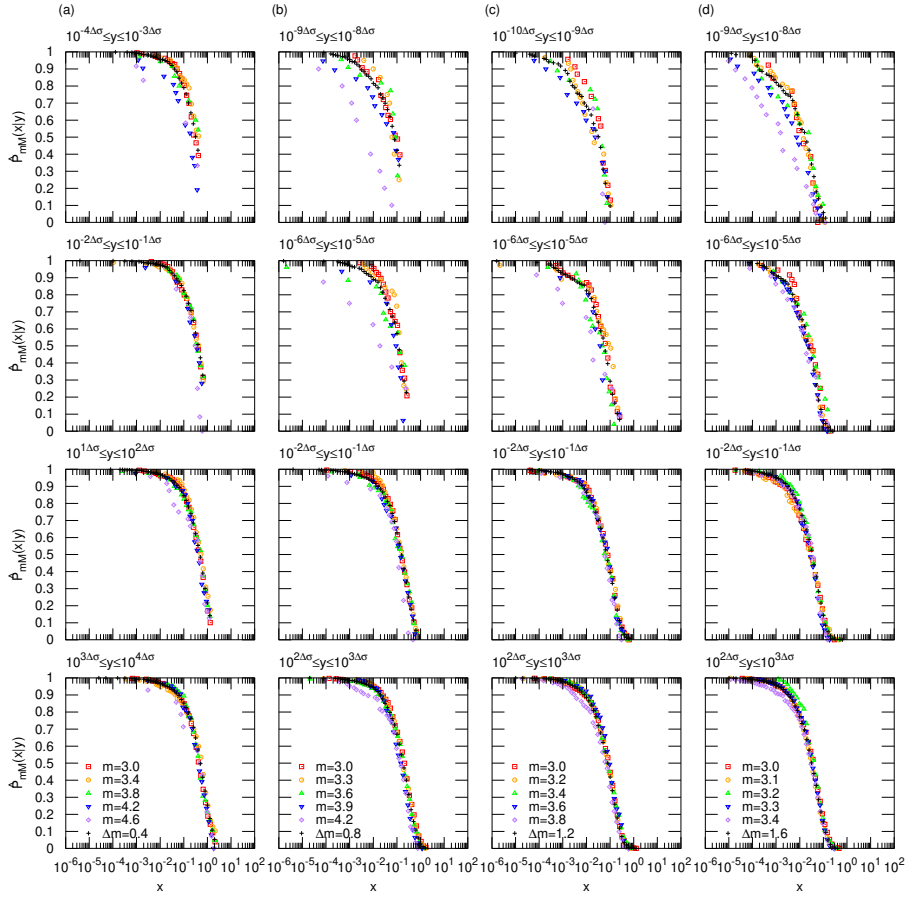

Figure S3-(1): JS

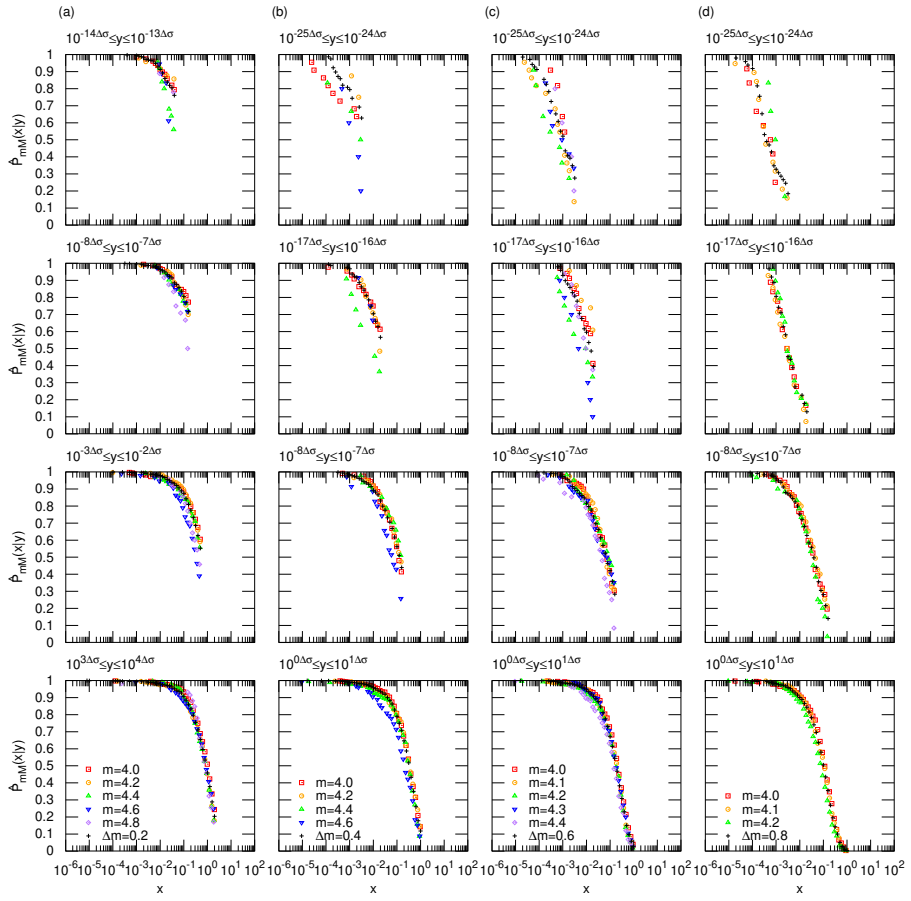

Figure S3-(2): JA

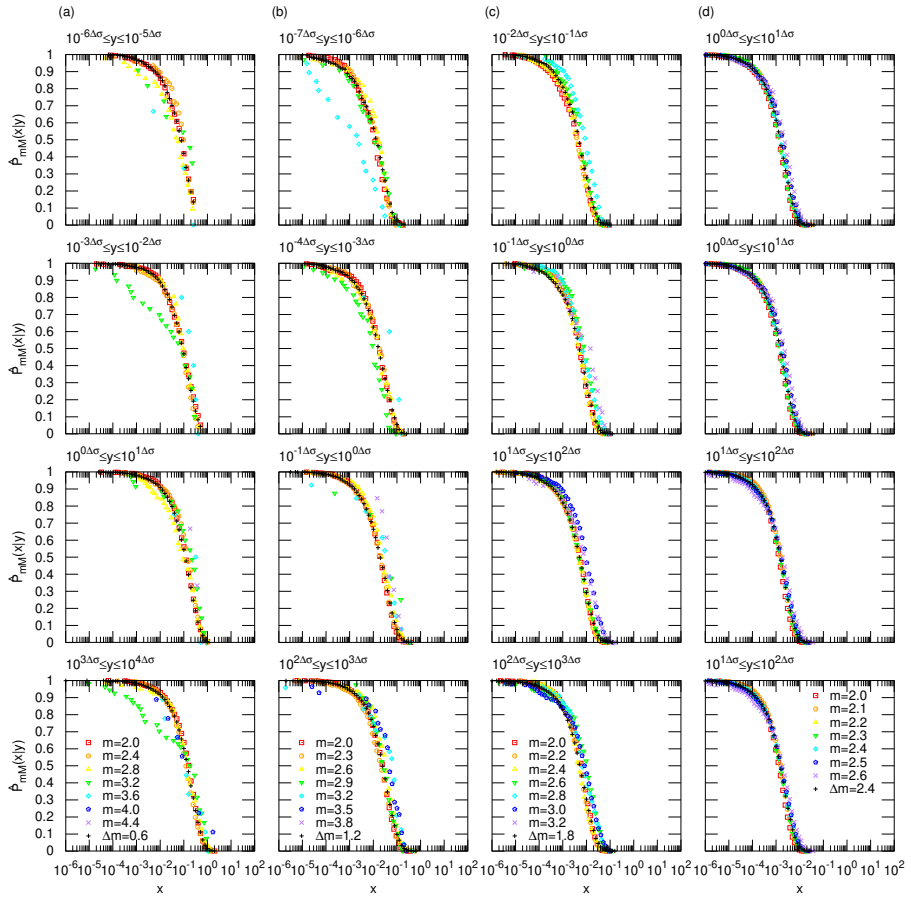

Figure S3-(3): SCA1

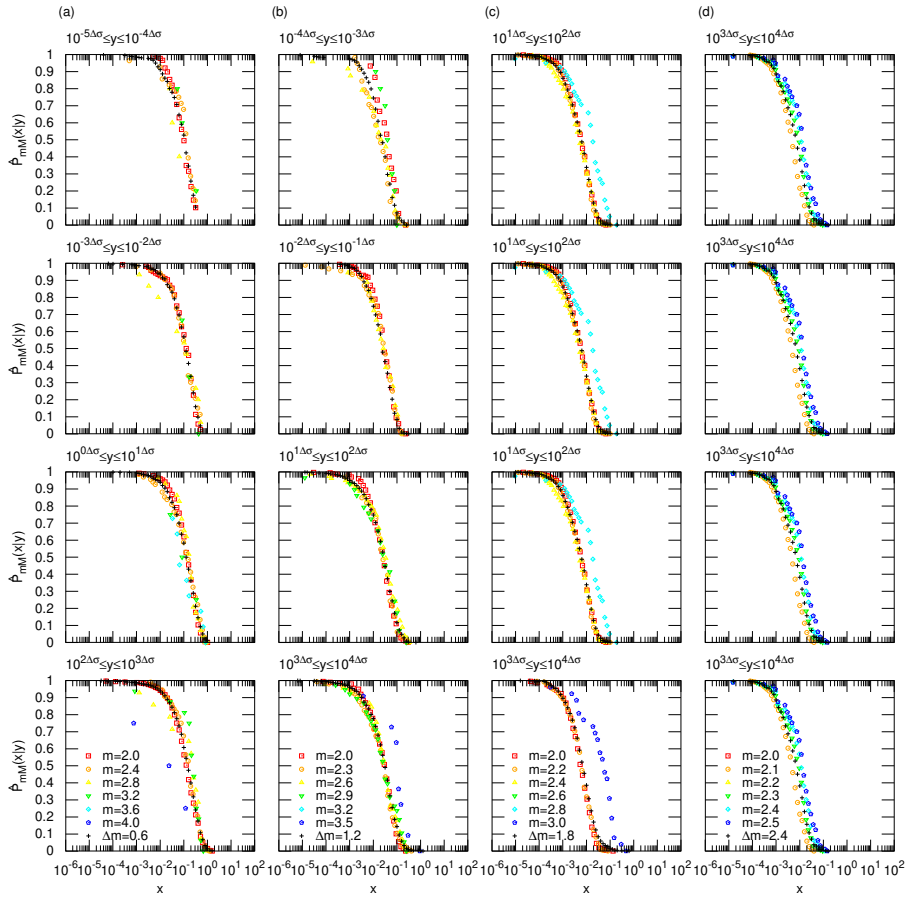

Figure S3-(4): SCA2

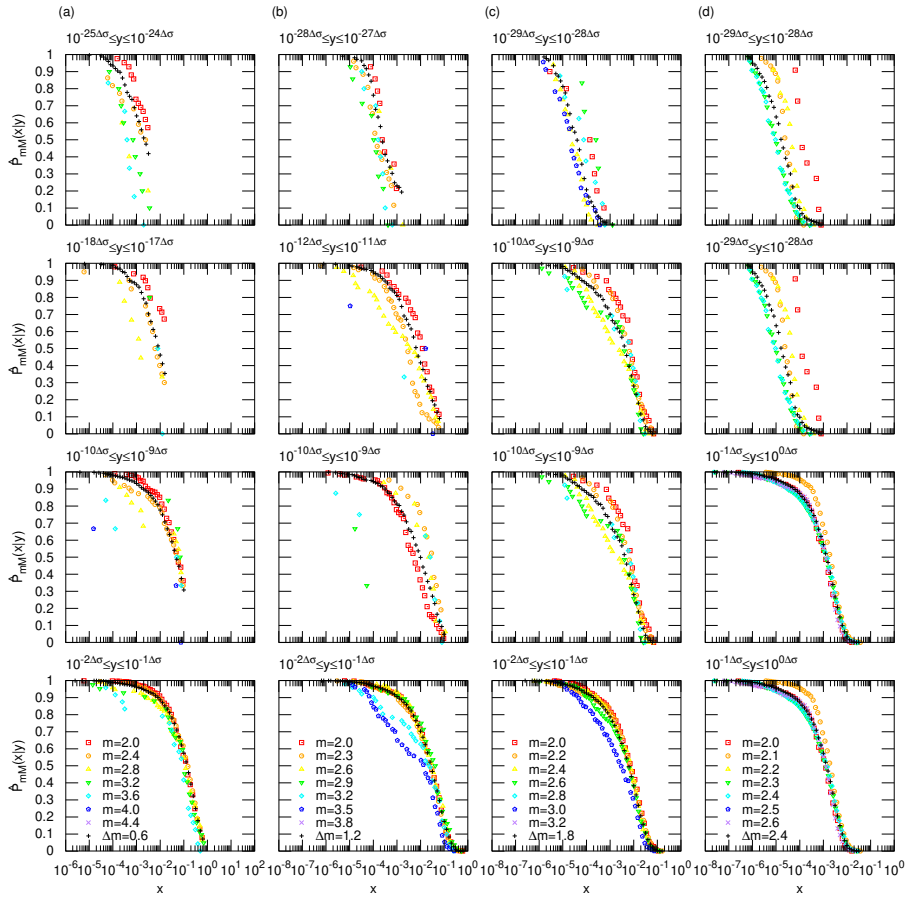

Figure S3-(5): SCA3

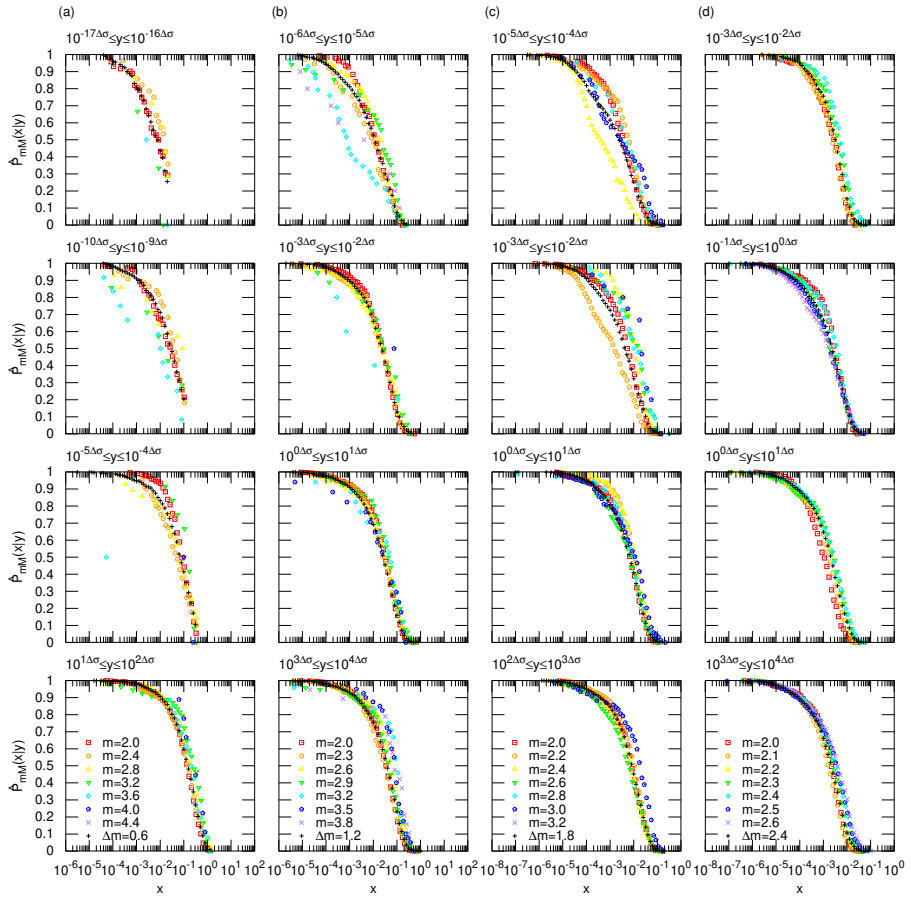

Figure S3-(6): SCS1

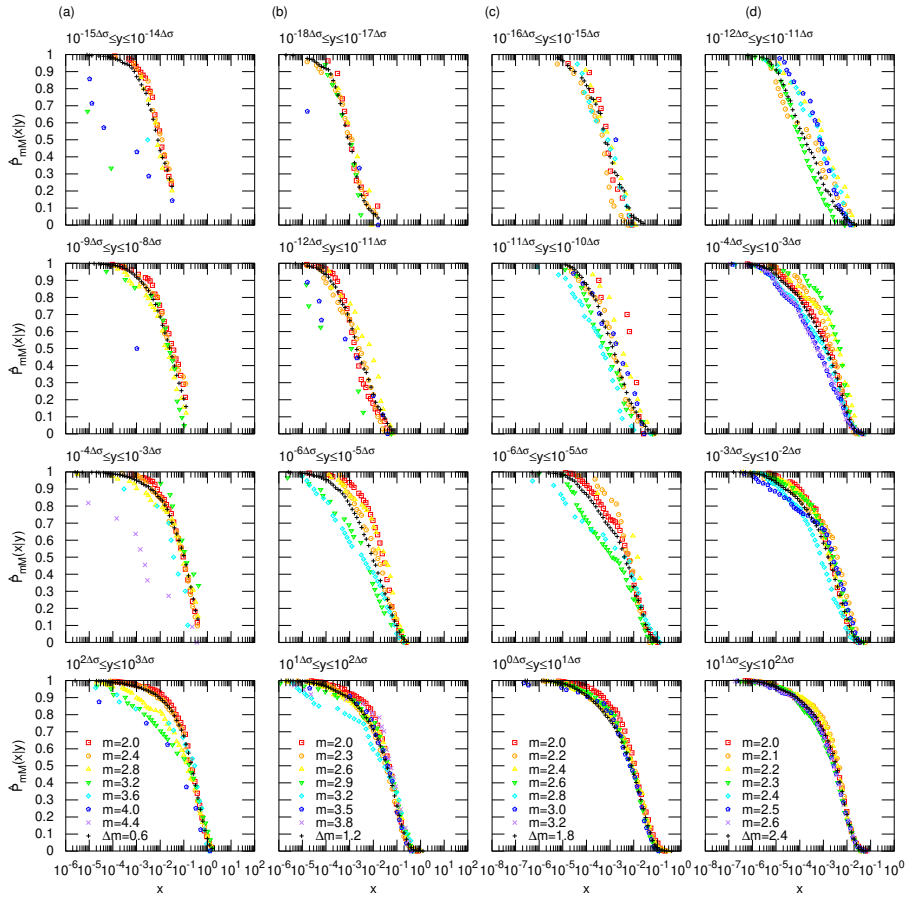

Figure S3-(7): SCS2

Q.4 Figures S4- for $\hat{p}_{\Delta m}(x|y)$ and $\overline{p}_{\Delta m}(x|y)$

Figures S4-(1) to S4-(8) show $\hat{p}_{\Delta m}(x|y)$ and $\overline{p}_{\Delta m}(x|y)$ for each time series with: (a) $\Delta m = 0.4$, (b) $\Delta m = 0.8$, (c) $\Delta m = 1.2$, and (d) $\Delta m = 1.6$ for S4-(1), S4-(2); (a) $\Delta m = 0.2$, (b) $\Delta m = 0.4$, (c) $\Delta m = 0.6$, and (d) $\Delta m = 0.8$ for S4-(3); (a) $\Delta m = 0.6$, (b) $\Delta m = 1.2$, (c) $\Delta m = 1.8$, and (d) $\Delta m = 2.4$ for S4-(4) – S4-(8).

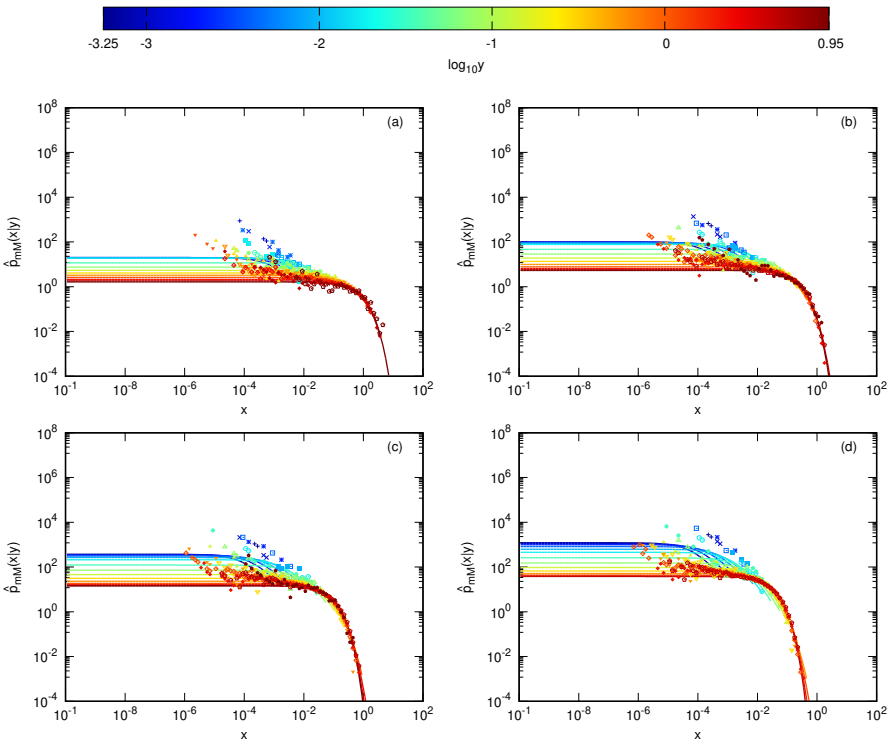

Figure S4-(1): CMT

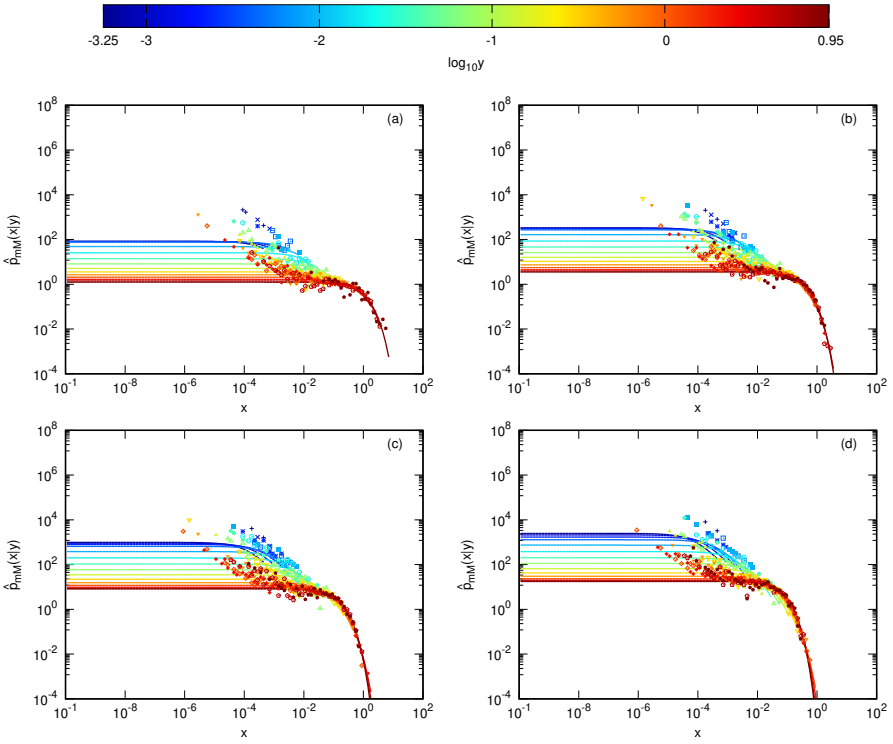

Figure S4-(2): JS

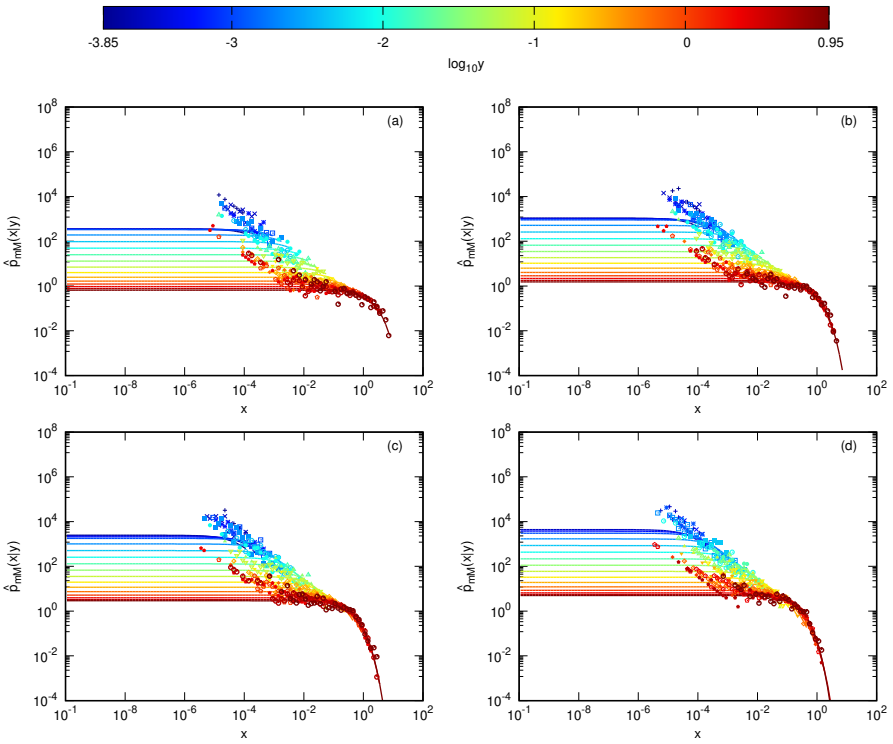

Figure S4-(3): JA

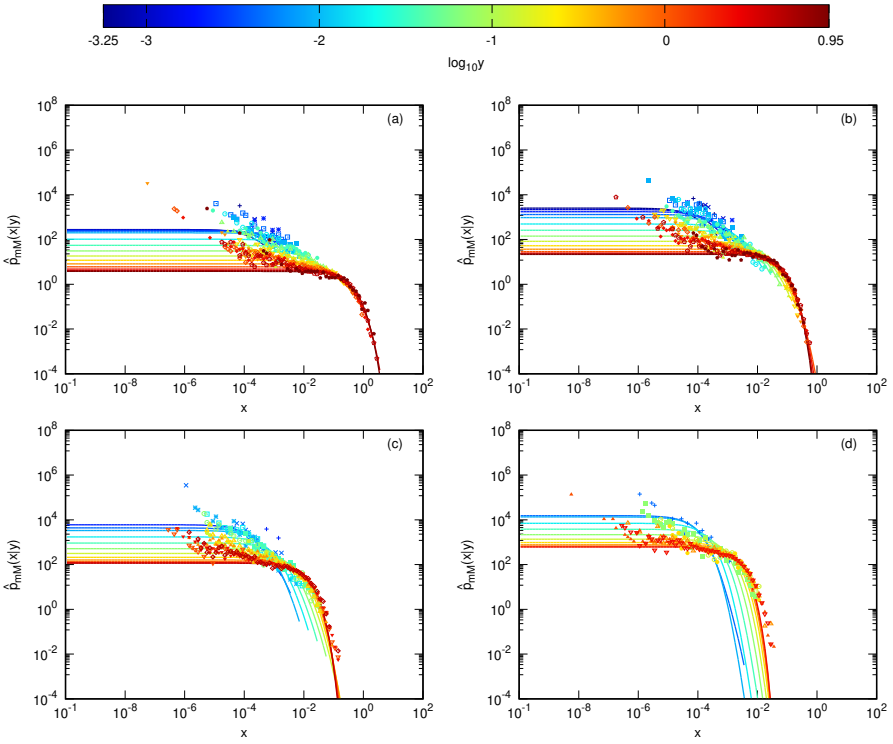

Figure S4-(4): SCA1

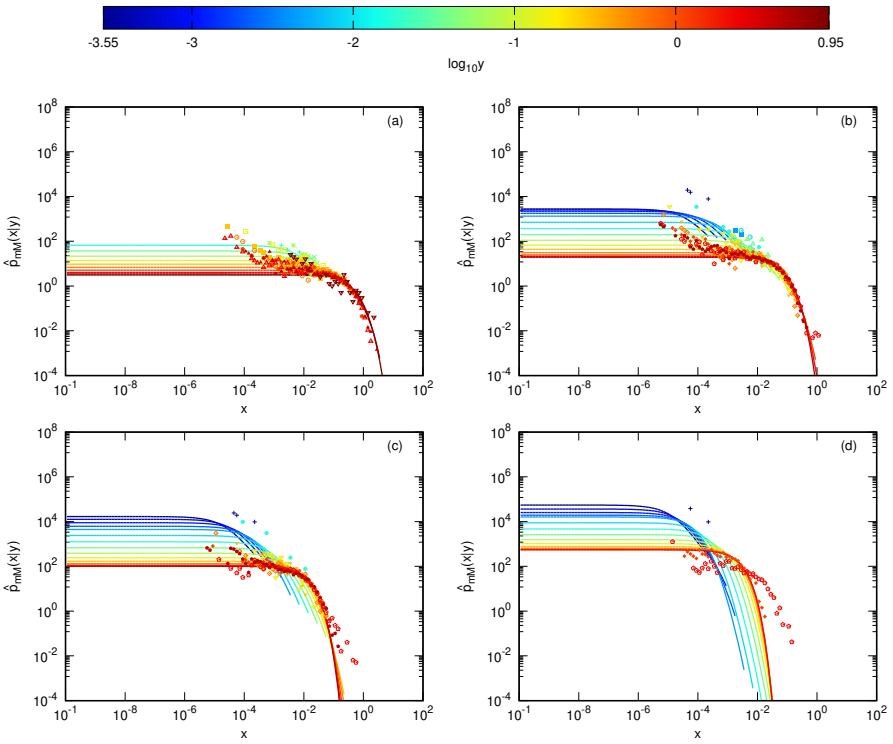

Figure S4-(5): SCA2

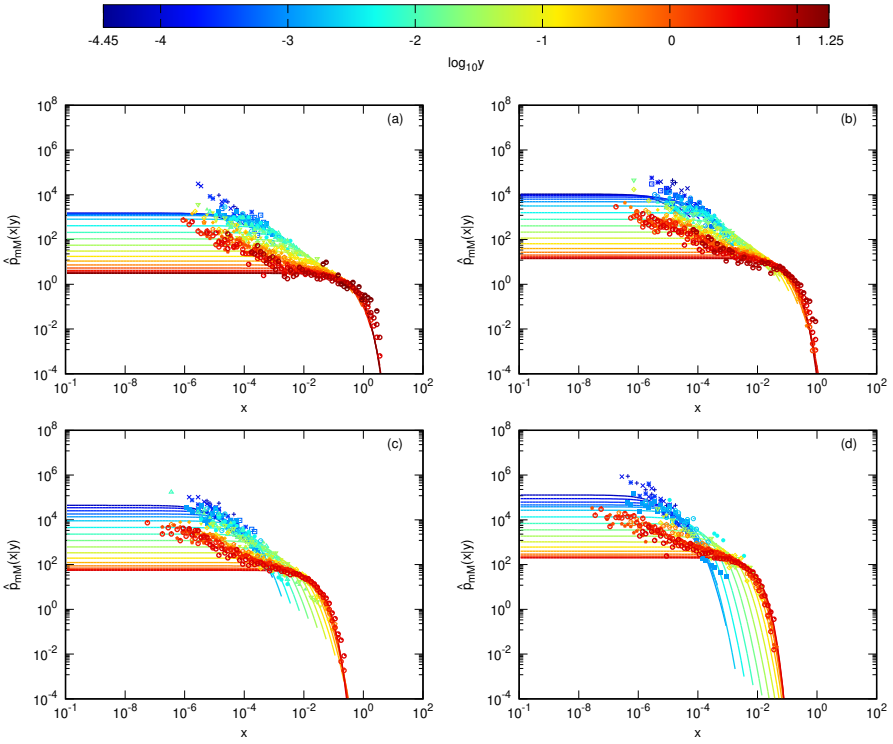

Figure S4-(6): SCA3

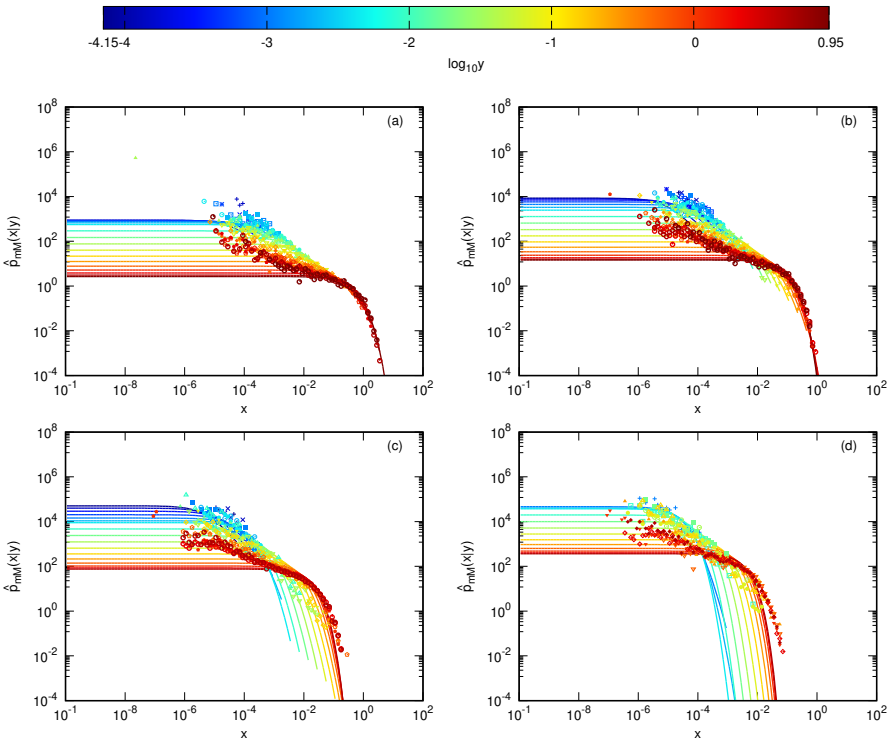

Figure S4-(7): SCS1

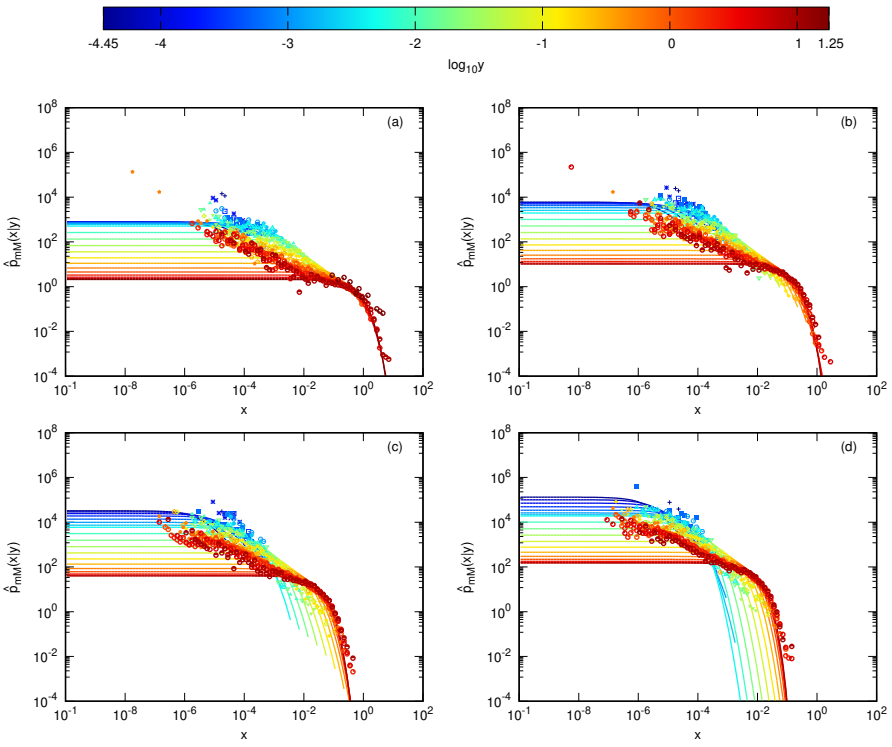

Figure S4-(8): SCS2

Q.5 Figures S5- for $\hat{P}_{Mm}(y|x)$ and $\overline{P}_{\Delta m}(y|x)$

Figures S5-(1) to S5-(7) show $\hat{P}_{Mm}(y|x)$ and $\overline{P}_{\Delta m}(y|x)$ for each time series with: (a) $\Delta m = 0.4$, (b) $\Delta m = 0.8$, (c) $\Delta m = 1.2$, and (d) $\Delta m = 1.6$ for Fig. S5-(1); (a) $\Delta m = 0.2$, (b) $\Delta m = 0.4$, (c) $\Delta m = 0.6$, and (d) $\Delta m = 0.8$ for Fig. S5-(2); (a) $\Delta m = 0.6$, (b) $\Delta m = 1.2$, (c) $\Delta m = 1.8$, and (d) $\Delta m = 2.4$ for Figs. S5-(3) – S5-(7). The displayed y ranges in Fig. S5-(4) are chosen from such j' s that at least 5 of the (m, M)s indicated in the legend are in the range $(10^{j\Delta\sigma}, 10^{(j+1)\Delta\sigma})$. Note the same multi-times display of identical x ranges as in Fig. S3-.

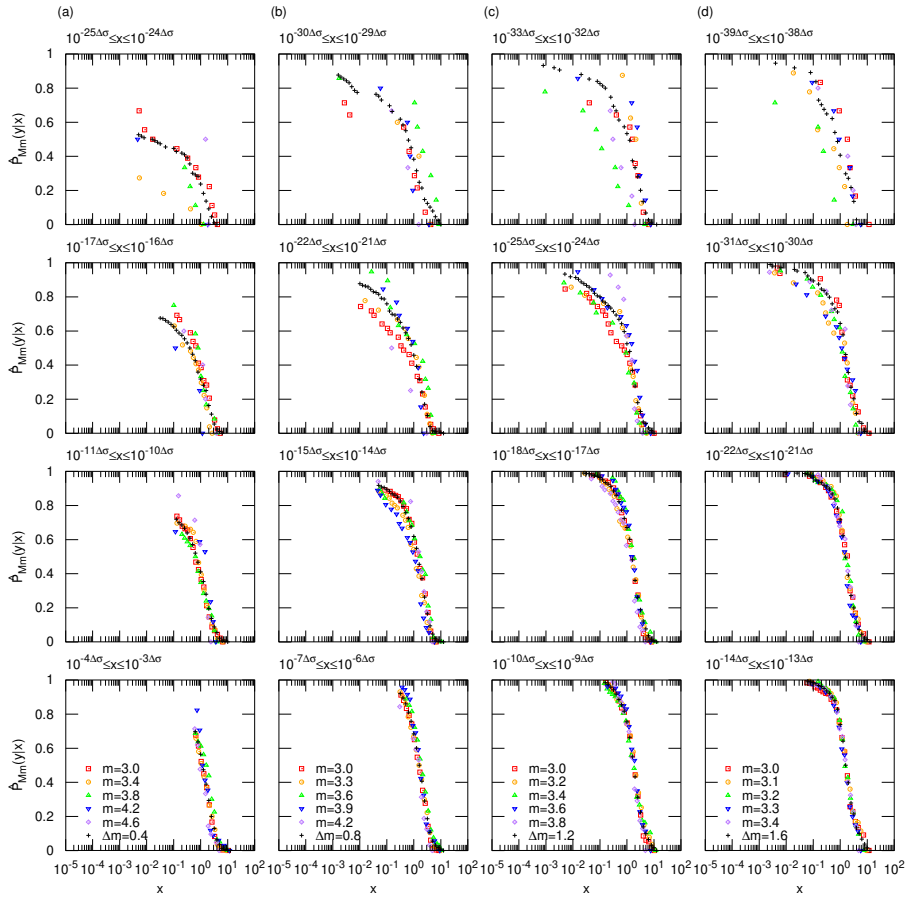

Figure S5-(1): JS

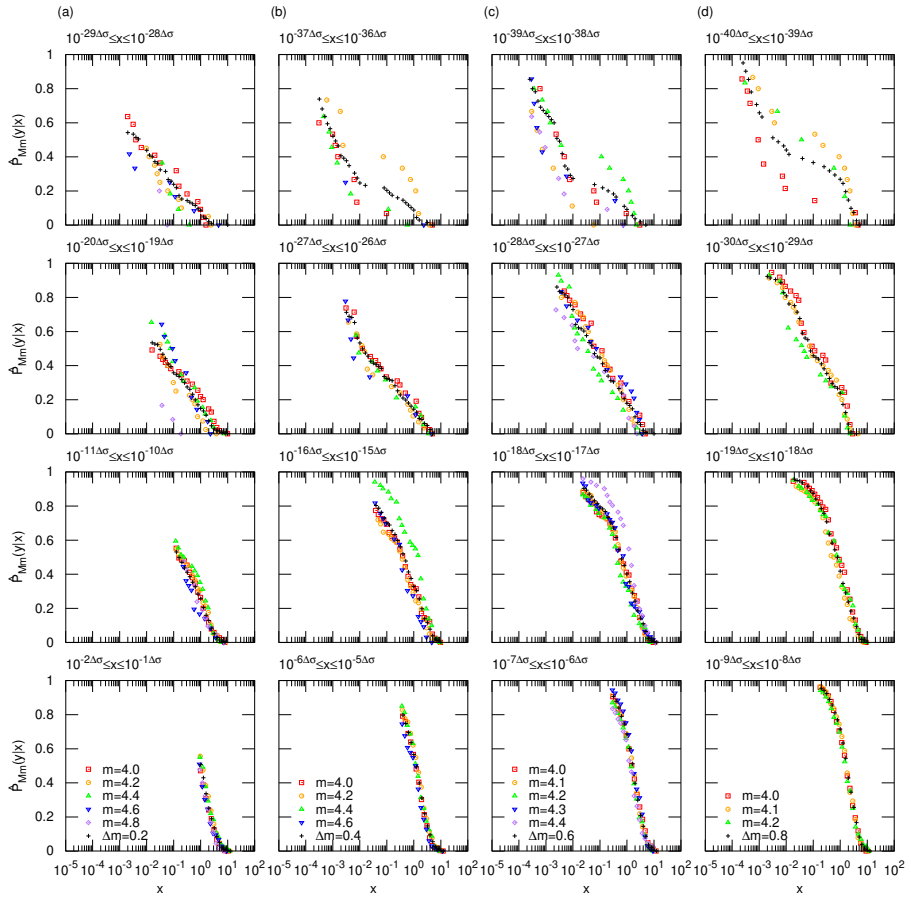

Figure S5-(2): JA

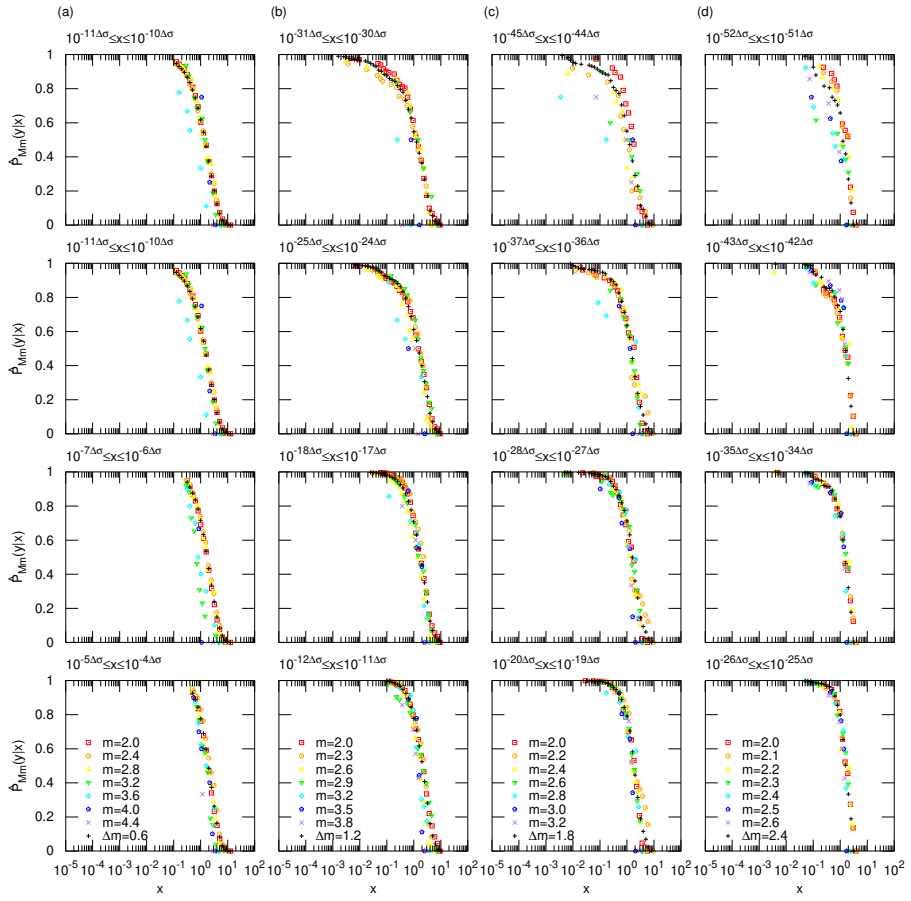

Figure S5-(3): SCA1

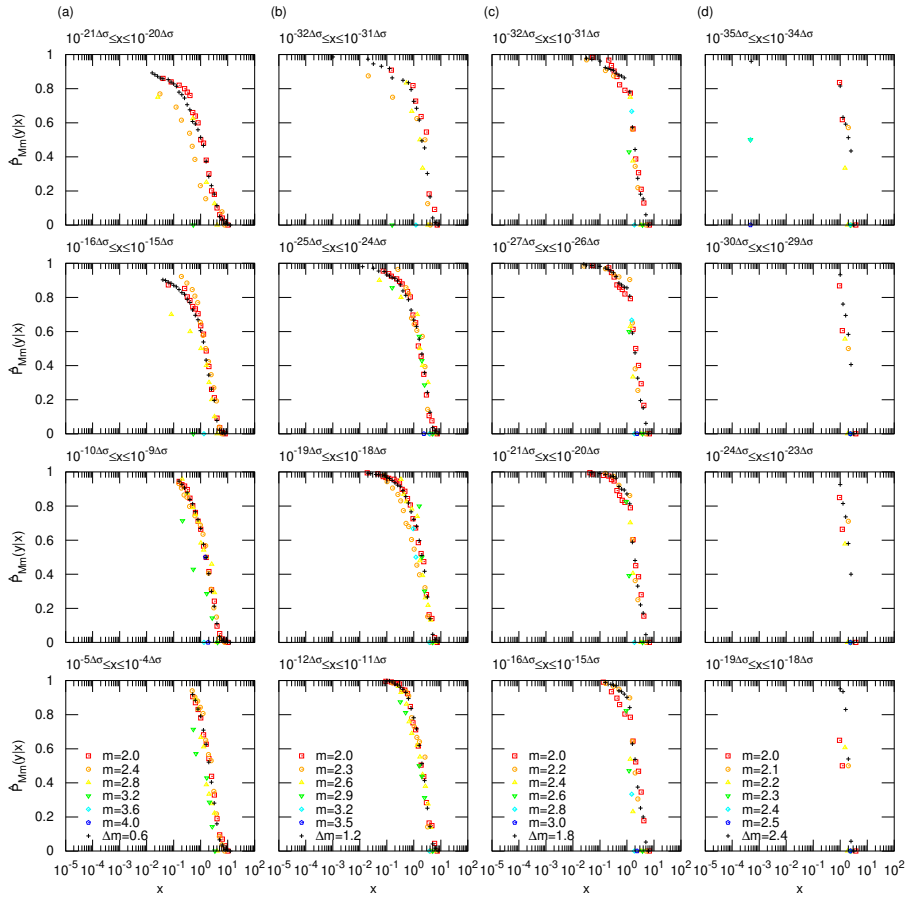

Figure S5-(4): SCA2

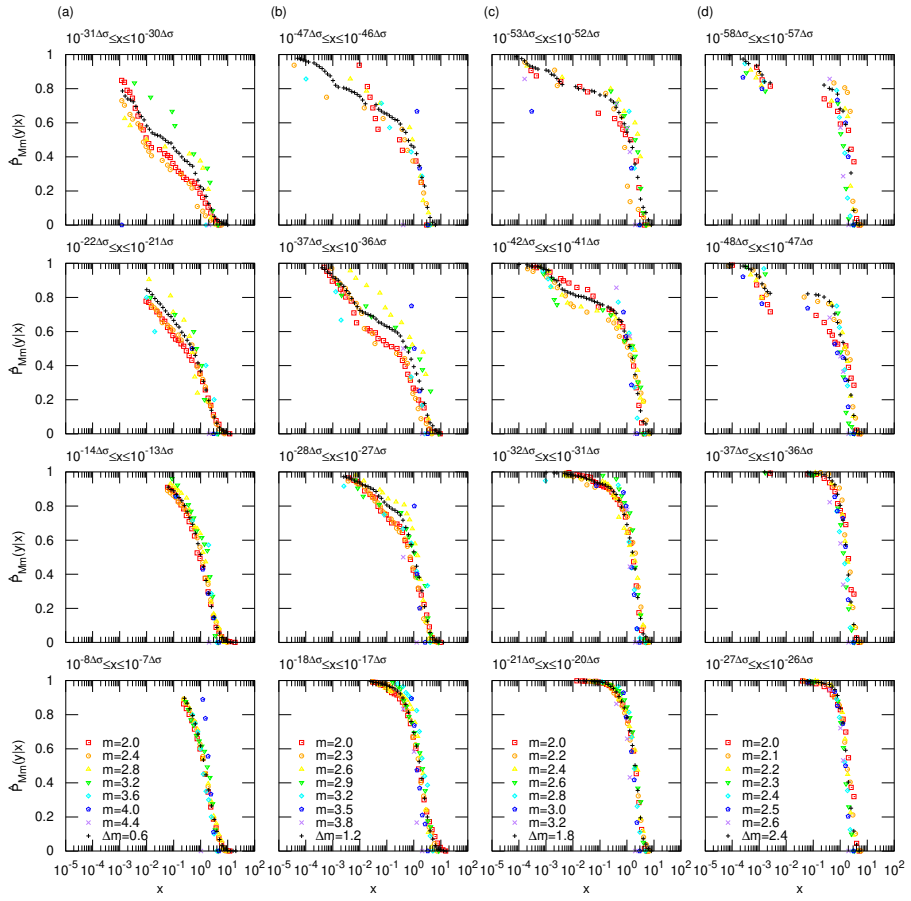

Figure S5-(5): SCA3

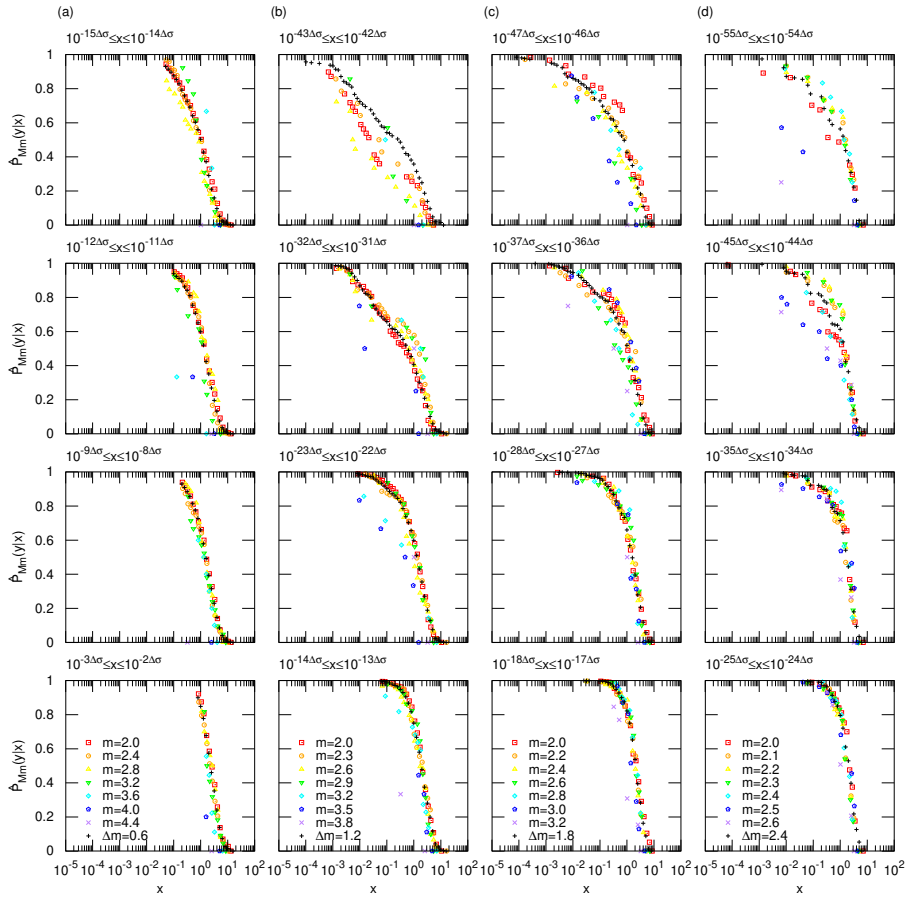

Figure S5-(6): SCS1

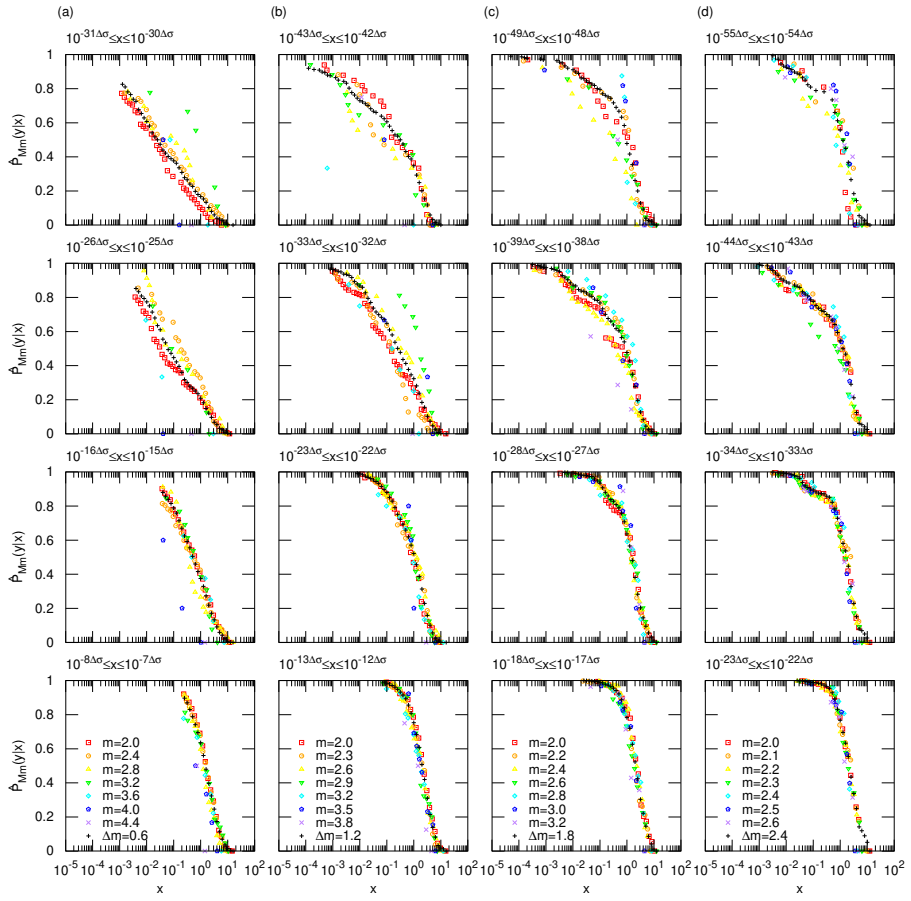

Figure S5-(7): SCS2

Q.6 Figures S6- for $\hat{p}_{\Delta m}(y|x)$ and $\overline{p}_{\Delta m}(y|x)$

Figures S6-(1) to S6-(8) show $\hat{p}_{\Delta m}(y|x)$ and $\overline{p}_{\Delta m}(y|x)$ for each time series with Δm values: (a) $\Delta m = 0.4$, (b) $\Delta m = 0.8$, (c) $\Delta m = 1.2$, and (d) $\Delta m = 1.6$ for Figs. S6-(1), S6-(2); (a) $\Delta m = 0.2$, (b) $\Delta m = 0.4$, (c) $\Delta m = 0.6$, and (d) $\Delta m = 0.8$ for Fig. S6-(3); (a) $\Delta m = 0.6$, (b) $\Delta m = 1.2$, (c) $\Delta m = 1.8$, and (d) $\Delta m = 2.4$ for Figs. S6-(4) – S6-(8).

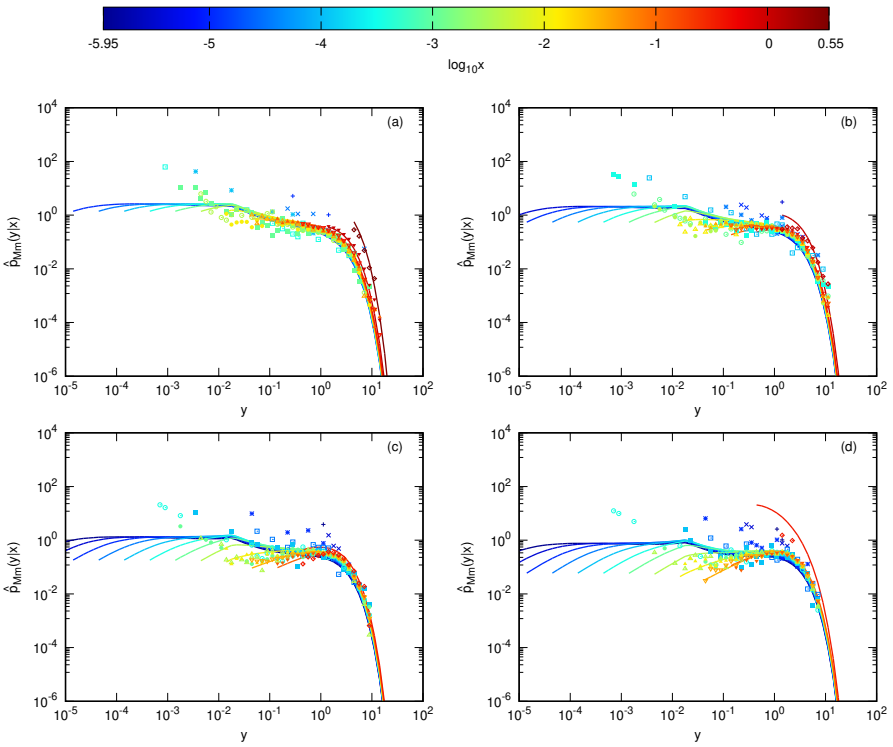

Figure S6-(1): CMT

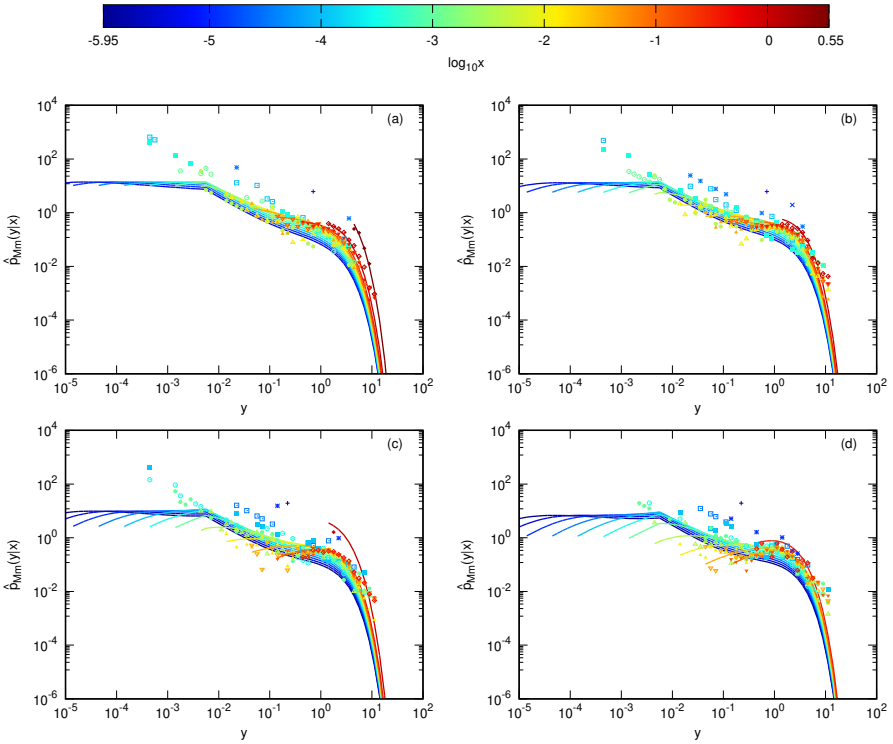

Figure S6-(2): JS

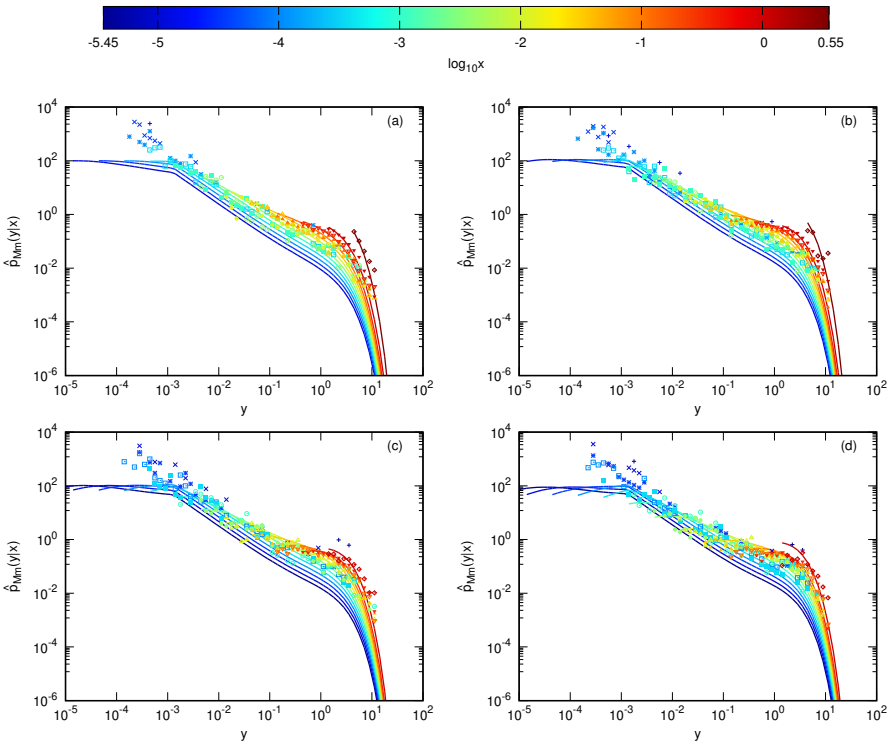

Figure S6-(3): JA

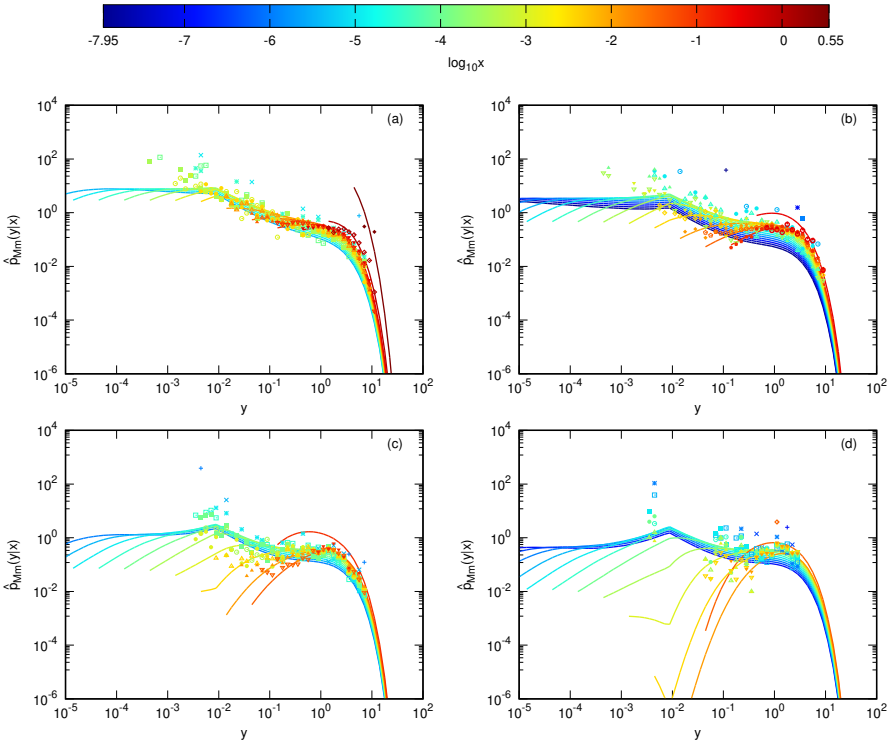

Figure S6-(4): SCA1

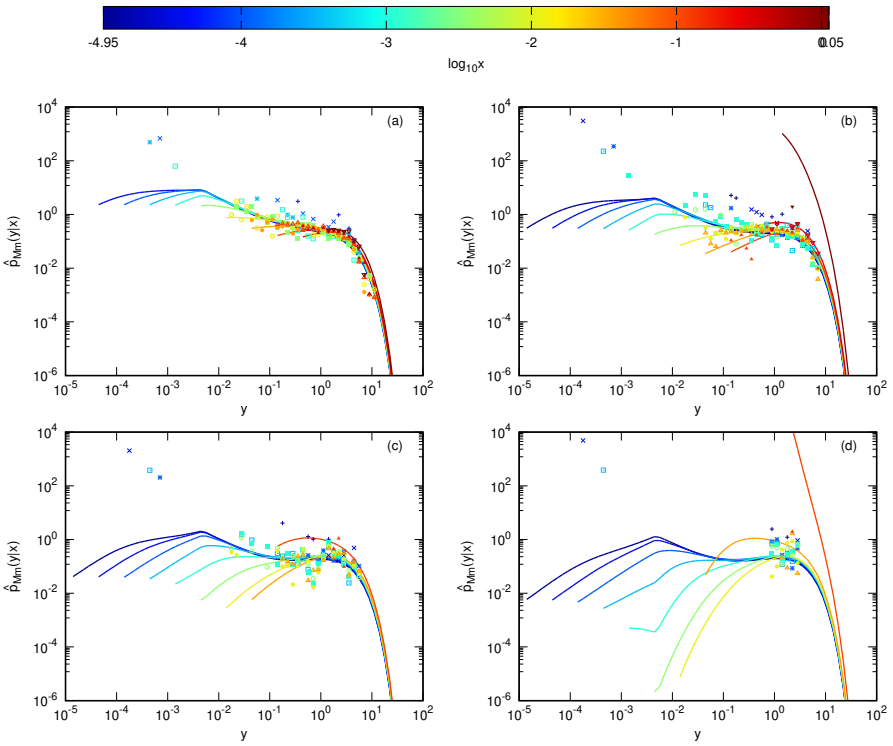

Figure S6-(5): SCA2

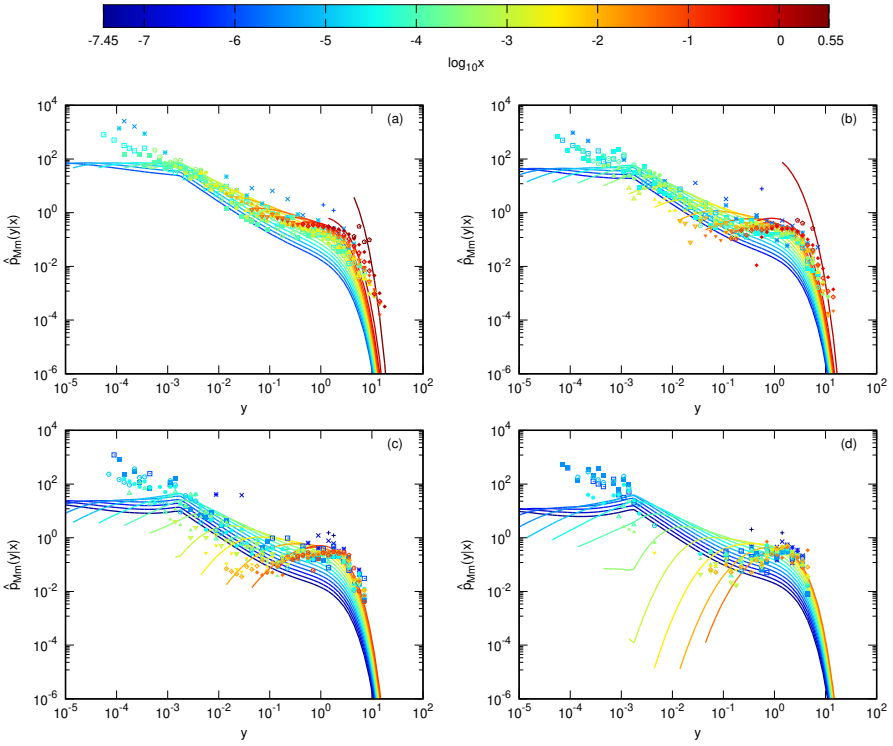

Figure S6-(6): SCA3

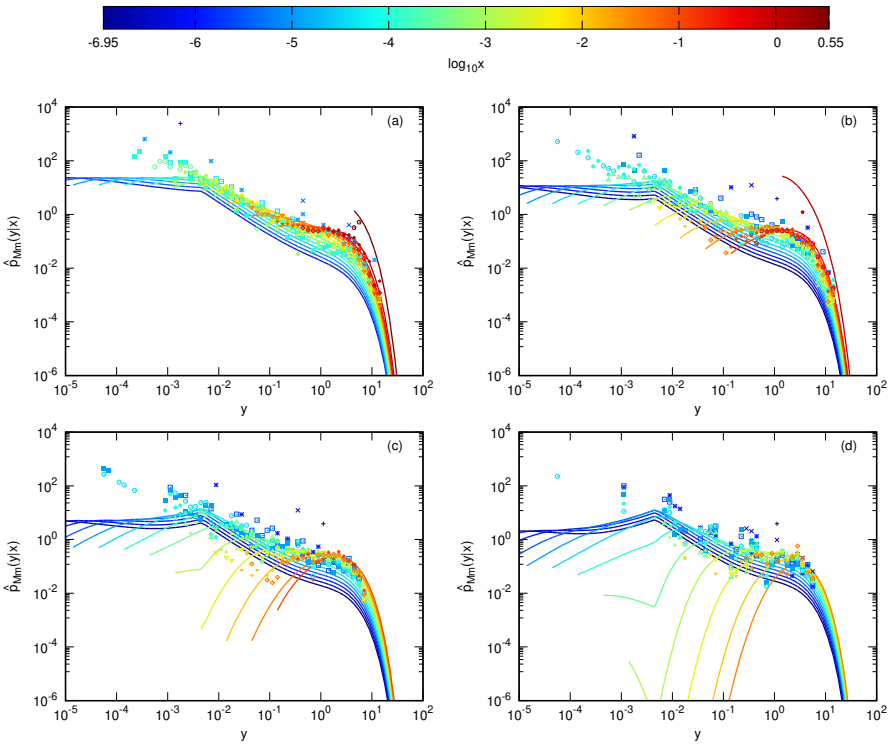

Figure S6-(7): SCS1

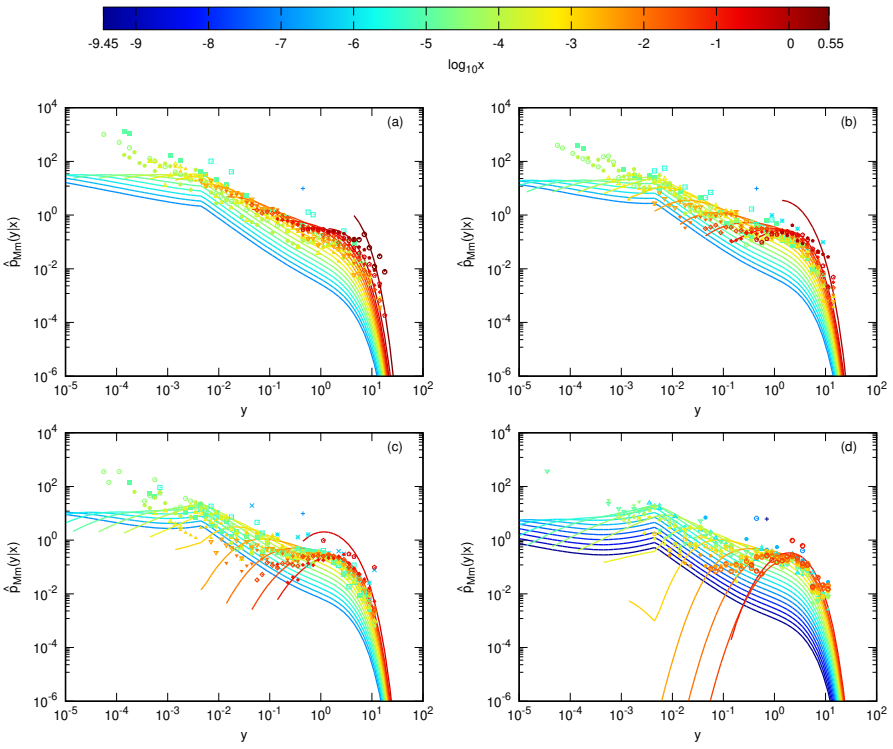

Figure S6-(8): SCS2

Bibliography

[1] H. Tanaka and Y. Aizawa. Detailed Analysis of the Interoccurrence Time Statistics in Seismic Activity. J. Phys. Soc. Jpn. **86**, 024004, 2017.

[2] H. Tanaka and K. Umeno. Bayesian Updating on Time Intervals at Different Magnitude Thresholds in a Marked Point Process and Its Application to Synthetic Seismic Activity. J. Phys. Soc. Jpn. **93**, 024001, 2024.

[3] H. Tanaka and K. Umeno. A Bayesian Inference Method for a Large Magnitude Event in a Spatiotemporal Marked Point Process Representing Seismic Activity. J. Phys. Soc. Jpn. **92**, 113001, 2023.

[4] H. Tanaka. Statistical properties of inter-event times in seismic time series transformed by occurrence rate: An analysis from the viewpoint of hierarchy in the temporal nature of seismicity. ESS Open Archive **December 27**, 2023, DOI: 10.22541/essoar.170365335.59197113/v1.

[5] 宇津徳治. 地震学. 共立出版, 2012 [in Japanese].

[6] C. H. Scholz. *The Mechanics of Earthquakes and Faulting*. Cambridge University Press, 2019, 3rd ed.

[7] 尾形良彦. 地震の確率予測の研究– その展望. 統計数理 **63**(1), 3, 2015 [in Japanese].

[8] S. Abe and N. Suzuki. Small-world structure of earthquake network. Physica A **337**, 357, 2004.

[9] S. Abe and N. Suzuki. Scale-free network of earthquakes. Europhys. Lett. **65**, 581, 2004.

[10] M. Baiesi and M. Paczuski. Scale Free Networks of Earthquakes and Aftershocks. Phys. Rev. E **69**, 066106, 2004.

[11] M. Baiesi and M. Paczuski. Complex networks of earthquakes and aftershocks. Nonlin. Proc. Geophys. **12**, 1, 2005.

[12] M. Naylor, K. Orfanogiannaki, and D. Harte. Exploratory Data Analysis: Magnitude, Space, and Time. Community Online Resource for Statistical Seismicity Analysis, 2010, DOI: 10.5078/corssa-92330203.

[13] J. Woessner, J. Hardebeck, and E. Hauksson. What is an instrumental seismicity catalog? Community Online Resource for Statistical Seismicity Analysis, 2010, DOI: 10.5078/corssa-38784307.

[14] H. Kanamori. The Energy Release in Great Earthquakes. J. Geophys. Res. **82**, 2981, 1977.

[15] T. Hanks and H. Kanamori. A Moment Magnitude Scale. J. Geophys. Res. **84**, 2348, 1979.

[16] Data can be downloaded from [http://evrrss.eri.u-tokyo.ac.jp/tseis/jma1/index-j.html], last access : November 2023; Japan Meteorological Agency Earthquake catalog.

[17] 鶴岡弘. WWW を用いた地震情報検索・解析システムの開発. 情報学基礎, **49**(9), 65, 1998 [in Japanese].

[18] Data can be downloaded from [http://scedc.caltech.edu/research-tools/alt-2011-dd-hauksson-yang-shearer.html], last access: November 2023; Caltech/USGS Southern California Seismic Network (SCSN) DOI: 10.7914/SN/CI; Southern California Earthquake Data Center (SCEDC) DOI: 10.7909/C3WD3xH1.

[19] E. Hauksson, W. Yang, and P. M. Shearer. Waveform Relocated Earthquake Catalog for Southern California (1981 to June 2011). Bull. Seismol. Soc. Am. **102**, 2239, 2012.

[20] National Earthquake Information Center, Preliminary Determination of Epicenters catalog, [http://earthquake.usgs.gov/data/pde.php]. Catalog data is downloaded from [https://earthquake.usgs.gov/earthquakes/search/], by setting the options : Event type = "Earthquakes", Catalog and Contributor = "US - National Earthquake Information Center, PDE", last access : January 2024.

[21] Data can be downloaded from [https://www.globalcmt.org/], last access: October 2023.

[22] G. Ekström, M. Nettles, and A. Dziewoński. The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. Phys. Earth Planet. Inter, **200**, 1, 2012.

[23] A. M. Dziewonski, T.-A. Chou, and J. H. Woodhouse. Determination of Earthquake Source Parameters From Waveform Data for Studies of Global and Regional Seismicity. J. Geophys. Res.: Solid Earth **86**, 2825, 1981.

[24] 井出哲. 絵でわかる地震の科学. 講談社, 2017 [in Japanese].

[25] Y. Aizawa, T. Hasumi, and S. Tsugawa. Seismic Statistics: Universality and Interim Report on the 3. 11 Earthquake (2011) in Fukushima-Miyagi Area. Nonlin. Phenom. Complex Syst. **16**, 116, 2013.

[26] Y. Aizawa and S. Tsugawa. Aftershock Cascade of the 3.11 Earthquake (2011) in Fukushima-Miyagi Area. Nonlin. Phenom. Complex Syst.: From Nano to Macro Scale, NATO Science for Peace and Security Series C: Environmental Security **1**, 21, 2014.

[27] Y. Ogata. Statistical model for standard seismicity and detection of anomalies by residual analysis. Tectonophysics **169**, 159, 1989.

[28] F. Omori. On the After-shocks of Earthquakes. J. Coll. Sci., Imp. Univ. Tokyo **7**, 111, 1894.

[29] T. Utsu. A Statistical Study on the Occurrence of Aftershocks. Geophys. Mag. **30**, 521, 1961.

[30] T. Utsu, Y. Ogata, and R. S. Matsu'ura. The centenary of the Omori Formula for a Decay Law of Aftershock Activity. J. Phys. Earth **43**, 1, 1995.

[31] B. Gutenberg and C. F. Richter. Frequency of earthquakes in California. Bull. Seismol. Soc. Am. **34**, 185, 1944.

[32] D. Amorese. Applying a Change-Point Detection Method on Frequency-Magnitude Distributions. Bull. Seismol. Soc. Am. **97**, 1742, 2007.

[33] K. Aki. Maximum Likelihood Estimate of b in the Formula log N= a-bM and its Confidence Limits. Bull. Earthq. Res. Inst., Tokyo Univ. **43**, 237, 1965.

[34] 宇津徳治. 6. 地震の規模別度数の統計式 log n= a-bM の係数 b を求める一方法. 北海道大学地球物理学研究報告 **13**, 99, 1965 [in Japanese].

[35] A. Mignan and J. Woessner. Estimating the magnitude of completeness for earthquake catalogs. Community Online Resource for Statistical Seismicity Analysis, 2012, DOI: 10.5078/corssa-00180805.

[36] L. Gulia, S. Wiemer, and M. Wyss. Catalog artifacts and quality controls. Community Online Resource for Statistical Seismicity Analysis, 2012, DOI: 10.5078/corssa-93722864.

[37] D. Schorlemmer and J. Woessner. Probability of detecting an earthquake. Bull. Seismol. Soc. Am. **98**, 2103, 2008.

[38] Y. Y. Kagan. Short-Term Properties of Earthquake Catalogs and Models of Earthquake Source. Bull. Seismol. Soc. Am. **94**, 1207, 2004.

[39] A. Helmstetter, Y. Y. Kagan, and D. D. Jackson. Comparison of Short-Term and Time-Independent Earthquake Forecast Models for Southern California. Bull. Seismol. Soc. Am. **96**, 90, 2006.

[40] E. Lippiello, C. Godano, and L. Arcangelis. The earthquake magnitude is influenced by previous seismicity. Geophys. Res. Lett. **39**, L05309, 2012.

[41] B. Gutenberg and C. F. Richter. Magnitude and energy of earthquakes. Ann. Geophys. **9**, 1, 1956.

[42] P. Bak, C. Tang, and K. Wiesenfeld. Self-Organized Criticality: An Explanation of 1/f Noise. Phys. Rev. Lett. **59**, 381, 1987.

[43] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality. Phys. Rev. A **38**, 364, 1988.

[44] J. M. Carlson and J. Langer. Mechanical model of an earthquake fault. Phys. Rev. A **40**, 6470, 1989.

[45] J. M. Carlson and J. Langer. Properties of Earthquakes Generated by Fault Dynamics. Phys. Rev. Lett. **62**, 2632, 1989.

[46] R. Burridge and L. Knopoff. Model and theoretical seismicity. Bull. Seismol. Soc. Am. **57**, 341, 1967.

[47] Z. Olami, H. J. S. Feder, and K. Christensen. Self-Organized Criticality in a Continuous, Nonconservative Cellular Automaton Modeling Earthquakes. Phys. Rev. Lett. **68**, 1244, 1992.

[48] S. Hergarten and H. J. Neugebauer. Foreshocks and Aftershocks in the Olami-Feder-Christensen Model. Phys. Rev. Lett. **88**, 238501, 2002.

[49] A. Helmstetter, S. Hergarten, and D. Sornette. Properties of foreshocks and aftershocks of the nonconservative self-organized critical Olami-Feder-Christensen model. Phys. Rev. E **70**, 046120, 2004.

[50] Y. Ogata. Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. J. Am. Stat. Assoc. **83**, 9, 1988.

[51] D. Vere-Jones. Forecasting earthquakes and earthquake risk. Int. J. Forecasting **11**, 503, 1995.

[52] Y. Ogata. Space-time point-process models for earthquake occurrences. Ann. Inst. Statist. Math. **50**, 379, 1998.

[53] Y. Ogata, K. Katsura, and M. Tanemura. Modelling heterogeneous space–time occurrences of earthquakes and its residual analysis. J. R. Stat. Soc. Ser. C. Appl. Stat. **52**, 499, 2003.

[54] Y. Ogata and J. Zhuang. Space–time ETAS models and an improved extension. Tectonophysics **413**, 13, 2006.

[55] Y. Ogata. Prediction and validation of short-to-long-term earthquake probabilities in inland Japan using the hierarchical space–time ETAS and space–time Poisson process models. Earth Planets Space **74**, 110, 2022.

[56] 宇津徳治, 関彰. 余震区域の面積と本震のエネルギーとの関係. 地震 第 2 輯 **7**, 233, 1955 [in Japanese].

[57] A. Helmstetter. Is earthquake triggering driven by small earthquakes? Phys. Rev. Lett. **91**, 058501, 2003.

[58] A. Helmstetter and D. Sornette. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. J. Geophys. Res.:Solid Earth **107**, 2237, 2002.

[59] 近江崇宏, 野村俊一. 点過程の時系列解析. 統計学 One Point 14, 共立出版, 2021 [in Japanese].

[60] A. Helmstetter and D. Sornette. Importance of direct and indirect triggered seismicity in the ETAS model of seismicity. Geophys. Res. Lett. **30**, 1576, 2003.

[61] A. Helmstetter and D. Sornette. Båth's law derived from the Gutenberg-Richter law and from aftershock properties. Geophys. Res. Lett. **30**, 2069, 2003.

[62] A. Helmstetter and D. Sornette. Foreshocks explained by cascades of triggered seismicity. J. Geophys. Res.: Solid Earth **108**, 2457, 2003.

[63] A. Helmstetter, D. Sornette, and J.-R. Grasso. Mainshocks are aftershocks of conditional foreshocks: How do foreshock statistical properties emerge from aftershock laws. J. Geophys. Res.:Solid Earth **108**, 2046, 2003.

[64] Z. Xiong and J. Zhuang. SETAS: A Spherical Version of the Space–Time ETAS Model. Seismol. Res. Lett. **94**, 1676, 2023.

[65] Y. Ogata, K. Katsura, H. Tsuruoka, and N. Hirata. Exploring Magnitude Forecasting of the Next Earthquake. Seismol. Res. Lett. **89**, 1298, 2018.

[66] D. A. Rhoades and F. F. Evison. Long-range Earthquake Forecasting with Every Earthquake a Precursor According to Scale. Pure Appl. Geophys. **161**, 47, 2004.

[67] C. Smyth and J. Mori. Statistical models for temporal variations of seismicity parameters to forecast seismicity rates in Japan. Earth Planets Space **63**, 231, 2011.

[68] E. Lippiello, L. de Arcangelis, and C. Godano. Influence of Time and Space Correlations on Earthquake Magnitude. Phys. Rev. Lett. **100**, 038501, 2008.

[69] 地震調査研究推進本部 `https://www.jishin.go.jp/reports/research\_report/choukihyoka\_01b/]`in Japanese], last access: November 2023.

[70] C. H. Scholz. *The Mechanics of Earthquakes and Faulting*. Cambridge University Press, 2002, 2nd ed., Section 7.4.3.

[71] P. Bak, K. Christensen, L. Danon, and T. Scanlon. Unified Scaling Law for Earthquakes. Phys. Rev. Lett. **88**, 178501, 2002.

[72] Á. Corral. Local distributions and rate fluctuations in a unified scaling law for earthquakes. Phys. Rev. E **68**, 035102, 2003.

[73] Á. Corral. Universal local versus unified global scaling laws in the statistics of seismicity. Physica A **340**, 590, 2004.

[74] E. Lippiello, C. Godano, and L. de Arcangelis. Dynamical Scaling in Branching Models for Seismicity. Phys. Rev. Lett. **98**, 098501, 2007.

[75] Y. Aizawa. Foundations of earthquake statistics in view of non-stationary chaos theory. 物性研究 **97**, 309, 2011.

[76] J. Davidsen and C. Goltz. Are seismic waiting time distributions universal? Geophys. Res. Lett. **31**, L21612, 2004.

[77] Á. Corral. Long-Term Clustering, Scaling, and Universality in the Temporal Occurrence of Earthquakes. Phys. Rev. Lett. **92**, 108501, 2004.

[78] S. Touati, M. Naylor, and I. G. Main. Origin and Nonuniversality of the Earthquake Interevent Time Distribution. Phys. Rev. Lett. **102**, 168501, 2009.

[79] Á. Corral. Dependence of earthquake recurrence times and independence of magnitudes on seismicity history. Tectonophysics **424**, 177, 2006.

[80] Á. Corral. Statistical Features of Earthquake Temporal Occurrence. Lect. Notes Phys. **705**, 191, Springer, 2007.

[81] G. Molchan. Interevent Time Distribution in Seismicity: A Theoretical Approach. Pure Appl. Geophys. **162**, 1135, 2005.

[82] S. Hainzl, F. Scherbaum, and C. Beauval. Estimating Background Activity Based on Interevent-Time Distribution. Bull. Seismol. Soc. Am. **96**, 313, 2006.

[83] A. Saichev and D. Sornette. “Universal” Distribution of Interearthquake Times Explained. Phys. Rev. Lett. **97**, 078501, 2006.

[84] A. Saichev and D. Sornette. Theory of earthquake recurrence times. J. Geophys. Res.: Solid Earth **112**, B04313, 2007.

[85] D. Sornette, S. Utkin, and A. Saichev. Solution of the nonlinear theory and tests of earthquake recurrence times. Phys. Rev. E **77**, 066109, 2008.

[86] M. Bottiglieri, L. de Arcangelis, C. Godano, and E. Lippiello. Multiple-Time Scaling and Universal Behavior of the Earthquake Interevent Time Distribution. Phys. Rev. Lett. **104**, 158501, 2010.

[87] E. Lippiello, Á. Corral, M. Bottiglieri, C. Godano, and L. de Arcangelis. Scaling behavior of the earthquake intertime distribution: Influence of large shocks and time scales in the Omori law. Phys. Rev. E **86**, 066119, 2012.

[88] L. de Arcangelis, C. Godano, J. R. Grasso, and E. Lippiello. Statistical physics approach to earthquake occurrence and forecasting. Physics Reports **628**, 1, 2016, Chapter 3.

[89] E. Marekova. Scaling Analysis of Time Distribution between Successive Earthquakes in Aftershock Sequences. Acta Geophys. **64**, 885, 2016.

[90] T. Hasumi, T. Akimoto, and Y. Aizawa. The Weibull–log Weibull distribution for interoccurrence times of earthquakes. Physica A **388**, 491, 2009.

[91] T. Hasumi, C.-c. Chen, T. Akimoto, and Y. Aizawa. The Weibull–log Weibull transition of interoccurrence time for synthetic and natural earthquakes. Tectonophysics **485**, 9, 2010.

[92] H. Tanaka and Y. Aizawa. Addendum to “Detailed Analysis of the Interoccurrence Time Statistics in Seismic Activity” [J. Phys. Soc. Jpn. 86, 024004 (2017)]. J. Phys. Soc. Jpn. **86**, 057001, 2017.

[93] H. Tanaka and K. Umeno. Bayesian Updating for Time-Intervals of Different Magnitude Thresholds in Marked Point Process and Its Application to Time-Series of ETAS Model. ESS Open Archive **December 16**, 2021, DOI: 10.1002/essoar.10509635.1.

[94] Private communication with Y. Aizawa. The idea of Eqs. (3.5) and (3.6) were inspired by Y. Aizawa.

[95] 吉川徹, 岡部篤行. 線分をランダムな点で区切った区間の長さの確率分布に関する分析. 総合都市研究 **43**, 99, 1991 [in Japanese].

[96] D. J. Webb. The Statistics of Relative Abundance and Diversity. J. Theor. Biol. **43**, 277, 1974.

[97] Y. Zhang, H. Zhao, X. He, F.-D. Pei, and G.-G. Li. Bayesian prediction of earthquake network based on space–time influence domain. Phys. A Stat. Mech. Appl. **445**, 138, 2016.

[98] O. Bohigas and M.-J. Giannoni. Chaotic motion and random matrix theories. in: Dehesa, J.S., Gomez, J.M.G., Polls, A. (eds) Mathematical and Computational Methods in Nuclear Physics. Lecture Notes in Physics, **209**, Springer, Berlin, Heidelberg, 1984.

[99] 長谷川洋. 量子系の準位統計– 量子カオス序論. 物理学最前線 28 大槻義彦編, 共立出版, 1991, pp.16 – 17 [in Japanese].

[100] 斎藤信彦. カオスの物理. 物理学最前線 30 大槻義彦編, 共立出版, 1992, pp.74 – 78 [in Japanese].

[101] Y. Ogata and K. Shimazaki. Transition from aftershock to normal activity: the 1965 Rat Islands earthquake aftershock sequence. Bull. Seismol. Soc. Am. **74**, 1757, 1984.

[102] S. Wiemer and M. Wyss. Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. Bull. Seismol. Soc. Am. **90**, 859, 2000.

[103] [https://scedc.caltech.edu/earthquake/significant.html].

[104] Á. Corral. Renormalization-Group Transformations and Correlations of Seismicity. Phys. Rev. Lett. **95**, 028501, 2005.

Papers by the Author Related to This Thesis

1. Hiroki Tanaka and Yoji Aizawa. Detailed Analysis of the Interoccurrence Time Statistics in Seismic Activity. J. Phys. Soc. Jpn. **86**, 024004, 2017. (DOI: 10.7566/JPSJ.86.024004)

2. Hiroki Tanaka and Ken Umeno. Bayesian Updating on Time Intervals at Different Magnitude Thresholds in a Marked Point Process and Its Application to Synthetic Seismic Activity. J. Phys. Soc. Jpn. **93**, 024001, 2024. (DOI: 10.7566/JPSJ.93.024001)

3. Hiroki Tanaka and Ken Umeno. A Bayesian Inference Method for a Large Magnitude Event in a Spatiotemporal Marked Point Process Representing Seismic Activity. J. Phys. Soc. Jpn. **92**, 113001, 2024. (DOI: 10.7566/JPSJ.92.113001)

4. Hiroki Tanaka. Statistical properties of inter-event times in seismic time series transformed by occurrence rate: An analysis from the viewpoint of hierarchy in the temporal nature of seismicity. preprint available at ESS Open Archive, **December 27**, 2023. (DOI: 10.22541/essoar.170365335.59197113/v1)

- Chapter 1 is based on papers 1 – 4.
- Chapter 2 is based on paper 1.
- Chapter 3 is based on paper 2.
- Chapter 4 is based on paper 3.
- Chapter 5 is based on paper 4.