Special Proceeding of the 22nd Annual Conference of SSCA held at Savitribai Phule Pune University, Pune, during January 02-04, 2020; pp 11-24

Review of Temporal Point Process Models for Modeling Earthquake Aftershocks

Rupal Shah and K. Muralidharan
Department of Statistics, The M.S. University of Baroda, Gujarat, India

Received: 29 March 2020; Revised: 10 April 2020; Accepted: 14 April 2020

Abstract

It is well known that the occurrence of earthquake is likely to increase another earthquake or the number (or sequence) of aftershocks in the nearby space and time. Similarly, prior to the next major earthquake, pre-seismic foreshocks are expected to occur in the focal region. Thus, it may be of interest for seismologists to study the pattern of sequence of foreshocks and aftershocks for prediction of earthquake. To investigate this problem with the help of quantitative information like, magnitude of earthquake, latitude, longitude, time etc., it might be possible to develop statistical models for the definition and detection of the occurrence of earthquake and sequence of aftershocks. The point process models are most frequently used to model such time to event data. In the time to event data, because of their temporal context, the point process models are also referred to as temporal point process models. Some of the frequently used models are Self-Exciting point process models, Epidemic Type Aftershock-Sequences models, stress release model etc. In this paper, we review various other models specific to temporal events like earthquake and other situations.

Key words: Ground intensity function; Trigger function; Self-exciting point process; Marked point process; ETAS models; Stress release model (SRM).

1. Introduction

In variety of applications, like, occurrence of crime events, posts or likes or clicks on social media, occurrence of earthquakes and its aftershocks, trading in financial markets etc., we come across the sequence of events which are asynchronous in nature, unlike the time series in which the observations are taken at regular time intervals. The time points at which such events of interest occur are also known as temporal points and such data can carry some important relevant information other than the temporal points.

Some examples are:

1. In study of crime patterns, an act of violence by an individual or a group might provoke or stimulate counter attacks by another group of individuals; where the temporal events are the times when violence by one gang took place and relevant information may be the retaliatory attacks by rivalry gang, the locations, the race of victims etc.
2. One post on social media is followed by number of related posts or one video is repeatedly watched by number of viewers; here the temporal events are time of post or video and

Corresponding Author: K. Muralidharan
Email: muralikustat@gmail.com

relevant information may be number of likes or number of times it is watched or tagged, number of positive or negative comments, the geographic location of the respondents etc.

3. In market research, the sales of a product might get a jump following repeated advertisements; where the temporal events are the time of purchase of a pre specified consumer product and the relevant information may be the impact of television advertisements of the product, income, size of the household etc.
4. In medical context, a targeted treatment like chemotherapy can impact the diseased cells; the temporal events in this case may be the time points at which the patient is given a dose of chemotherapy and relevant information may be the dose, the effects (or side effects) etc.
5. In earthquake modeling, an earthquake of higher magnitude might induce more aftershocks as compared to that with a smaller magnitude; the temporal events in this case may be the time points at which earthquakes and major aftershocks (of magnitude beyond some predefined intensity) occur and the relevant information may be longitude and latitude of the location, magnitude of the aftershock etc.
6. In financial markets, when the stock prices are very much volatile or fluctuating (ultra-high frequency); the time points at which the prices rise or decline by a predefined number may be considered to be temporal points and the relevant information may be the factors like international crude oil price, political situation or stability of the country, some event which has international relevance or concerns etc.

In general, temporal events may be either statistically independent, following a Poisson process, or temporally correlated. The term "event dependence" captures the idea that an initial event can increase or decrease the likelihood of subsequent events in the future. Point processes are often found to be appropriate for modeling a series of asynchronous events occurring at points in time (Ascher and Feigngold, 1984) and are also useful for modeling some peculiar pattern of the events observed in time. Among the few examples cited above, we focused on the earthquake data from the perspective of statistical modeling and analysis.

The behavior of the temporal point process may be described with the help of the conditional intensity function, conditional on the history of the process over time. The conditional intensity function proposes the probability of occurrence of the subsequent event in the upcoming instance of time given the history of the past events until the present event. Thus it represents the rate for the occurrence of a new event, conditioned on the history of the process up to time *t*, say, $\mathcal{H}_t = \{t_i : t_i < t\}$, as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{E[N([t, t+\Delta t]) \,|\mathcal{H}_t]}{\Delta t} \tag{1}$$

where $N[\dots]$ is the number of events occurring in $(t, t+\Delta t)$.

The Nonhomogeneous Poisson process (NHPP) models are the most frequently used models to represent the temporal events. The NHPP models are models for reliability growth due to the nature of their intensity function which is a compromise between 'as good as new' and 'as bad as old' models [Duane (1964), Bassin (1969, 1973), Cox and Lewis (1966), Lewis (1970, 1972)]. At times we come across the temporal events with branching structure wherein the occurrence of an event of one type increases the chance of occurrence of similar kind of event in future. For example, each earthquake gives rise to aftershock activity in the neighboring areas depending upon the magnitude of the earthquake. Hawkes (1971) proposed a process to model such a branching structure of temporal events which was named Self-Exciting Point Process (SEPP) which implemented the idea that an earthquake can trigger aftershocks. As it can be observed from historical data on earthquakes, it is obvious to observe that the earthquakes with high intensity induces more aftershocks as compared to those with low intensity, that is, the earthquake and its aftershock sequence shows an epidemic behavioral pattern. Ogata (1988) developed a point process model to model such an epidemic behavior of earthquakes and named it Epidemic Type Aftershock Sequences (ETAS) model which is an extension of SEPP model.

Another point process model that is specifically defined to model the earthquake occurrences is a ‘stress release model’ (SRM) proposed by Knopoff (1971) and Vere-Jones (1978). The SRM is considered to be a stochastic version of the so-called ‘elastic rebound theory’. It considers the increased pressure in an area and the pressure released during an earthquake over a period of time. It is possible to measure the probability of earthquake occurrence using the conditional intensity function of the SRM.

Also the historical data of various earthquakes demonstrates that the high intensity earthquakes are normally followed by high intensity earthquakes in surrounding area and this resists the aftershocks. The interactions among such affected areas have influence on time and intensity of earthquake occurrences in the presence of stress movement. Bebbington and Harte (2001) proposed Linked Stress Release Model (Linked SRM) to model temporal events which exhibits such a characteristic.

Hawkes and Oakes (1974) defined cluster point process assuming that, at every point in temporal point process on $(0, \infty)$, a cluster of activities starts. This point can be interpreted as the first arrival time of a point process which triggers a random stream of events of the similar types. This model exhibits chain ladder type of characteristic which may be used to forecast the total number of events and the time of occurrence of events in the future course of time. Vere-Jones and Davies (1966) and Vere-Jones (1970) proposed trigger models which assumes that a series of primary events (say, main shocks) is distributed completely randomly in time and each of these primary events are capable of generating a series of secondary events (say, aftershocks).

Sometimes the temporal events might be just one of the components of the complex model which is carrying much more information apart from just the time of occurrence of events under study. This additional information may themselves have a stochastic structure and stochastic dependency. Daley and Vere-Jones (2003) classified such temporal point processes as marked point processes (MPP) or marked temporal point processes (MTPP). The behavior of the MTPP may also be described with the help of the conditional intensity function, conditional on the history of the process over time, having two components: the ground intensity function and the mark distribution. The ground intensity function describes the rate of occurrence of events with respect to time and the mark distribution describes the behavior of other variables (referred to as marks), that are associated with the event, and will also be usually dependent on the history of the occurrence of events.

The remaining part of this paper is organized this way: Section 2 offers various models and its technical descriptions. The likelihood estimation and related inferences are presented in

Section 3, followed by data analysis in Section 4. The Section 5 concludes with some useful discussion including the limitations.

2. Models and Descriptions

In case of earthquake occurrences, it can be commonly observed that an earthquake with very high magnitude is usually followed by a sequence of aftershocks and some of the aftershocks themselves may also be of such a high magnitude that they might have aftershocks caused by them. Obviously, both the frequency and magnitude diminish over a period of time, or with respect to space (distance from epicenter) etc. Also prior to a major earthquake or aftershock, pre-seismic activities and foreshocks might take place. But foreshocks are not very easy to identify as compared to aftershocks. Some researchers made attempts to study pre-seismic quiescence and the gaps between two major earthquakes of aftershocks. On the other hand, some researchers were of the opinion that the pre-seismic quiescence and the gaps are nothing but the result of decaying activity of aftershocks which were followed by the last major shock and so they are not of much importance to be studied. Thus, most of the studies focus on the main shocks and their aftershock sequences. In the context of earthquake studies, we consider the time points at which the main shock or major aftershocks occur as temporal points, or equivalently, the times of occurrence of main shocks or major aftershocks as temporal events.

2.1. NHPP models

In general, temporal events may be statistically independent or temporally correlated. Let $N(t_i, t_j)$ represents the number of events occurring between time t_i and t_j and let $(t_1, t_2), (t_3, t_4), \ldots, (t_{k-1}, t_k)$ be the disjoint sets where $t_1 < t_2 \leq \ldots \leq t_{k-1} < t_k$ are times of occurrence of temporal events, then N will be a Poisson process, if each of $N(t_1, t_2), N(t_3, t_4), \ldots, N(t_{k-1}, t_k)$ have a Poisson distribution and are independent, that is, $Cov[N(t_1, t_2), N(t_2, t_3)] = 0$ for any $t_1 < t_2 < t_3$. A Poisson process always has a deterministic conditional intensity $\lambda(t)$. If it is stationary, then $\lambda(t)$ is constant and the process is known as homogeneous Poisson process (HPP); otherwise the process is known as NHPP.

NHPP with a power-law intensity function is a frequently used model for temporal point process. According to the magnitude of the power law, one can ascertain whether the rate of occurrence of an event is decreasing, constant or increasing function of time. Crow (1974) proposed a model for which system failure times are assumed to occur according to a time dependent Poisson process with a Weibull intensity function of the form,

$$\lambda(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}, \text{ where } \theta > 0 \text{ and } \beta > 0. \tag{2}$$

Such a model is referred to as the *Weibull Process* or *Power Law Process* (PLP) and is used to model reliability growth. Another useful NHPP model was proposed was proposed by Cox and Lewis (1966) with intensity function of the form $\lambda(t) = e^{\alpha + \beta t}$, which is referred to as Log-Linear Process (LLP). This kind of models work well for any temporal event data exhibiting stationary characteristics in the long run.

2.2. Self-exciting point process models

When the temporal events are correlated, an initial event can increase or decrease the chances of occurrence of similar events in future. Hawkes (1971) introduced the *self-exciting process* as a point process for which $Cov[N(t_1, t_2), N(t_2, t_3)] > 0$ for any $t_1 < t_2 < t_3$. This means that if an event occurs, another event becomes more likely to occur in time and space (Hawkes and Oakes (1974), Daley and Vere-Jones (2003)). Hawkes (1971) defined the self-exciting process with an intensity function has the form:

$$\lambda(t) = \nu + \int_0^t g(t-s)\, dN(s) = \nu + k_0 \sum_{t_k<t} g(t-t_k) \tag{3}$$

where $\nu > 0$ is a baseline intensity or the background rate of events which is assumed to be constant in time and the second term describes the self-exciting part of the process having two components k_0 and g; k_0 reflects the magnitude of self-excitation and the function g measures the influence of an event on the intensity process or density at which self-excitation is triggered. For the earthquake study, the pre-seismic activities may be considered as the baseline intensity ν; intensity of the main shock which triggers the aftershocks may be considered as the magnitude of self-exciting part, that is k_0 and g is the density function at which self-excitation is triggered. It is assumed that $g(x) \geq 0$ for all $x \geq 0$, $g(x) = 0$ for $x < 0$, and $\int_0^t g(u)\, du < 1$, where t is the time up to which the events are observed. Many forms of $g(t)$ have been proposed and studied in literature, and in most of the cases, the choices and importance of the density depends on the situations and contexts. Egesdal et. al. (2010) considered g as an exponential distribution in their study, and they considered the intensity function having the form

$$\lambda(t) = \mu + k_0 \sum_{t_i<t} w\, e^{w\,(t_i-t)} \tag{4}$$

This intensity function describes the rate at which events occur over time, and is not only influenced by the current time, but also by the events that have occurred before the current time. The current events subsequently decrease over time exponentially. While studying earthquake and its aftershocks, similar pattern is likely to be observed and therefore it might be reasonable to consider the above form of intensity function while using SEPP to model earthquake occurrences.

2.3. ETAS model

The models with intensity function as proposed in SEPP were considered by Ogata (1988) for modeling the data regarding earthquake and its aftershocks, where it was assumed that the earthquake aftershock sequences can be modeled like an epidemic, that is, the earthquakes with larger magnitudes might have a sequence of more aftershocks in a given interval of time and also it may continue for a longer time after the main shock. He named such models as *Epidemic Type Aftershock-Sequences* (*ETAS*) models. Ogata (1988) studied various statistical models for the standard activity of sequence of earthquakes and compared them using likelihood methods. He proposed Epidemic-Type models with reference to the age-dependent birth and death process introduced by Kendall (1949), in which only the births or events are allowed to occur at a constant rate per unit time according to Poisson process and with each birth or event (with reference to earthquake study, main shock) there is associated a cluster of subsidiary events (say, aftershocks) formed by the births of all of the descendants of all generations of the immigrant (Hawkes and Oakes, 1974). Thus one of the differences between the epidemic-type model and the trigger model is that the trigger model assumes only the first generation offspring whereas the epidemic-type model assumes that each event has the possibility of possessing offspring. Ogata (1988) in his study proved that the epidemic-type models that include the effect of magnitude of earthquake give a better fit to the data than any of the trigger models (considered in a restricted form).

These models can be constructed by assuming two activities associated with each occurrence of an earthquake, background events and aftershock events. It is assumed that the background events occur independently according to a stationary Poisson process $\mu(x, y)$, with magnitudes distributed independently of μ and each occurrence of an earthquake increases the risk of aftershocks. Also, it is reasonable to assume that the increased risk of aftershocks also spreads in the neighboring locations in space and time according to the kernel $g(t)$.

Ogata (1988) considered the model (4) with $(t) = \frac{K}{(t+c)^p}$, as defined by the modified Omori formula (Utsu, 1961), where K, c and p are parameters, t is the time since occurrence of shock. The number K depends on the lower bound of the magnitude of aftershocks counted in $N(t)$, whereas c and p are known to be independent of the choice of lower bound. The epidemic-type model is defined in terms of the conditional intensity rate, or seismic risk as a function of time, based on the following assumptions: (a) the background seismic activity is generated according to a stationary Poisson process with a constant hazard rate (b) each shock has a risk of stimulating aftershocks in proportion of the quantity $e^{\beta M}$, where M is the magnitude of the main shock, and (c) the hazard rate of aftershocks decreases with time according to the modified Omori law, $\frac{K}{(t+c)^p}$.

Ogata (1985, 1988, 1989) demonstrated that the ordinary seismic activity of a wide region can be described in terms of the conditional intensity by the superposition of a constant rate for background seismicity and the modified Omori functions of any shocks i which occurred at time t_i, in such a way that

$$\lambda(t \mid \mathcal{H}_t) = \mu + \sum_{t_i<t} \frac{K_i}{(t-t_i+c)^p} \tag{5}$$

where μ is the rate of occurrence of the background seismic activity. The sum $\sum_{t_i<t}$ is taken for all shocks i which occurred before time t, and the parameter K_i for each shock i contributes to the size of the corresponding aftershock. More importantly the parameter K_i is dependent on the magnitude M_i of the aftershock as well as the cut-off or threshold magnitude M_0 of the data set according to the exponential function form

$$K_i = A\, e^{\alpha(M_i - M_0)} \tag{6}$$

Above form is based on the empirical formula obtained by Utsu and Seki (1955) regarding the linear relation between the logarithms of aftershock areas and the magnitudes M of the main shock. It suggests that the number N of aftershocks with magnitudes over a threshold M_0 for a fixed time span can be roughly estimated as proportional to $exp\big(\alpha\,(M - M_0)\big)$ for a constant α. Model (5) with (6) for the ordinary seismicity in terms of the rate of occurrence of aftershocks is an ETAS model with $\mathcal{H}_t = \{(t_i, M_i);\ t_i < t\}$ as the history of occurrence times $\{t_i\}$ up to time t and their corresponding magnitudes $\{M_i\}$, with intensity function

$$\lambda(t \mid \mathcal{H}_t) = \mu + A \sum_{t_i<t} e^{\alpha(M_i - M_0)} \left(\frac{c}{t-t_i+c}\right)^p \tag{7}$$

The ETAS model assumes that the aftershocks are generated as events occurring according to Poisson process with rate μ. The parameters (μ, A, α, c, p) are all positive, t_i is the time of occurrence of the i^{th} event with magnitude M_i and M_0 is a threshold magnitude of aftershocks considered in the study. The term $e^{\alpha(M_i - M_0)}$ conveys the meaning that the main shock or aftershock with larger magnitude raise the intensity of occurrence of aftershocks more, and the term $\left(\frac{c}{t-t_i+c}\right)^p$ determines the length (time) till which the aftershock sequence will continue to occur. There will have to be put certain constraints on parameters, otherwise the aftershock sequence (epidemic) could explode and never die out. The parameters α and p characterizes the temporal pattern of seismicity. The parameter p represents the rate at which the aftershock sequence decays, and the parameter α represents the vulnerability of magnitude of an earthquake in generating the aftershocks. Ogata (1987) also proved that the swarm-type activity has a smaller value of α, that is, a small α value indicates the presence of main shock and aftershock activity whereas the larger value of α indicates that there are only few large aftershocks or magnitude of main shock is much larger than the maximum magnitude of aftershocks. Thus it can be said that α^{-1} represents the average time until a next aftershock occurs.

The appropriate selection of parameter values is very crucial part of the modeling process. The distance in space and time over which the risk of main shocks or aftershocks spreads, the number of aftershocks, the dependence of the increased risk of aftershocks on magnitude size of main shock, etc. can have great impact on the power of a point process model in predicting the space and time of next major earthquake.

2.4. Stress release models

The elastic rebound theory proposed by Reid (1910) is a classical model for earthquake mechanisms which postulates that elastic stress in a seismically active region accumulates due to movement of tectonic plates, and is released when the stress exceeds the strength of the medium. Thus this theory suggests that a large earthquake should be followed by a period of quiescence (passive period), whereas in reality a strong earthquake can be followed by a period of activation (another earthquake of comparable magnitude). This elastic stress within a region can be extracted by collecting various kinds of information but obtaining the information regarding the temporal variations of seismic activity might be the more suitable approach as it is expected to reflect directly the nature of earthquake generating stress. Vere-Jones (1978) proposed the stress release model as a stochastic version of the elastic rebound theory, incorporating the deterministic stress build-up within a region and its stochastic release through earthquakes, by developing the stochastic model for the occurrence of sequence of main shocks which was proposed by Knopoff (1971). Many researchers applied stress release model to analyze several historical earthquakes, particularly to identify statistically distinct regions to which different stress models can be applied. One of the most interesting finding from these studies is that large earthquakes are often found to be followed by large earthquakes quite distant from the first. This seems consistent with a general consensus that the earthquakes taking place in the Earth's crust forms a tightly linked, near-critical process that exhibits the self-similarity, long-range correlation and power-law distributions. Thus it can be considered that there exists a class of models exhibiting self-organized criticality due to competition between local strengthening and weakening through interactions.

Suppose ρ is the loading rate which describes stress accumulation caused by large-scale tectonic plate movement, which is assumed to be constant and positive and $X(0)$ is the initial pressure, which is assumed to be positive. The stress is released when it exceeds the energy limit of plate strength in the form of earthquakes. Stress will increase linearly with the loading rate ρ at time t with initial stress $X(0)$ and thus the following equation can be obtained,

$$X(t) = X(0) + \rho t \tag{8}$$

Let t_i and S_i be the time and stress released, respectively, at the time of occurrence of event $i; i = 1, 2, \dots, n$ where n is the number of events occurred in the time interval $(0, t)$ under consideration. The value of stress S_i is correlated with the corresponding magnitude which is proportional to the seismic energy which is released during the occurrence of earthquake. Then the accumulated stress released by all events in the period $(0, t)$ can be expressed as $S(t) = \sum_{i;\, t_i<t} S_i$. An important variable in the stress release model is the level of stress in a certain area which controls the occurrence of an earthquake. The stress level at time t increases deterministically and decreases stochastically due to an earthquake. Therefore the stress released at time t, $X(t)$, is the difference between the accumulated stress that increases linearly with the loading rate ρ and the accumulated stress that is released in the period $(0, t)$, that is,

$$X(t) = X(0) + \rho\, t - S(t) \tag{9}$$

This is referred to as a Stress Release Model.

Further, the conditional intensity function of the above model can be obtained through its hazard function, say $\Psi(x)$, which denotes the probability of occurrence of an earthquake in the time interval $(t, t + \Delta t)$. Assuming that the hazard function is having an exponential form

$$\Psi(x) = \exp(\alpha + \beta x)\,;\ \alpha \in R, \beta \geq 0 \tag{10}$$

where the parameter α describes the initial stress value and parameter β describes the combination of strength and heterogeneity of the earth's crust in the area. Using equation (10), the conditional intensity function of the stress release model with the history $\mathcal{H}_t = \{(t_i, M_i);\ t_i < t\}$ is defined as

$$\lambda(t \mid \mathcal{H}_t) = \Psi\big(\mathrm{X(t)}\big) = exp\big(\alpha + \beta\,(X(0) + \rho\, t - S(t)\,)\big) \tag{11}$$

Taking $\alpha + \beta X(0) = a, \beta \rho = b$ and $1/\rho = c$, equation (11) reduces to

$$\lambda(t \mid \mathcal{H}_t) = exp\big(a + b\,(t - c\,S(t)\,)\big) \tag{12}$$

2.5. Linked stress release model

Apart from aftershocks, large events are often followed by large events quite distant from the first. On the other hand, very large events can resist occurrence of subsequent events or aftershocks on the same or other fault lines. Thus, interaction among areas can influence the time and magnitude of earthquake occurrence. Also, the regional stress itself evolves over centuries in the area which might be the result of the cumulative effects of all previous earthquakes and changes in tectonic loading. But the presence of stress movement and interaction among areas has not been incorporated in stress release model. The omission of this interaction between areas might underestimate the activity by the time and magnitude-predictable stress release model. These shortcomings of stress release model motivated its extension to include interactions among areas, by means of stress transfer and reduction and were named as Linked stress release model [Lu et al. (1999), Bebbington and Harte (2001)].

Zheng and Vere-Jones (1994) found that large geographical regions give better fits to the stress release model when broken down into subunits, and further noted some hints of clustering relating to some form of action at a distance, i.e. stress transfer and interaction. Thus, the multivariate extension to the stress release model was proposed as a linked stress release model by defining the evolution of stress $X_i(t)$ in the i^{th} region as

$$X_i(t) = X_i(0) + \rho_i\, t - \textstyle\sum_j \theta_{ij}\, S^{(j)}(t) \tag{13}$$

where $S^{(j)}(t)$ is the accumulated stress release in region j over the period $(0, t)$, and the coefficient θ_{ij} measures the fixed proportion of stress drop, initiated in region j, which is transferred to region i. Here, θ_{ij} may be positive or negative, resulting in damping or excitation respectively. It is convenient to set $\theta_{ii} = 1$ for all i while ignoring aftershocks. If $\theta_{ij} = 0$ for all $i \neq j$, the model is reduced to an independent combination of simple forms as that of stress release model. Then the point process conditional intensity function of the linked stress release model will be of the form

$$\lambda_i(t) = \Psi\left(X_i(t)\right) = exp\left(\alpha_i + \beta_i\left(X_i(0) + \rho_i t - \textstyle\sum_j \theta_{ij}\, S^{(j)}(t)\right)\right) \tag{14}$$

for each region i, where α_i, β_i, ρ_i and θ_{ij} are the parameters to be estimated. The above form of intensity function can be re-parameterized as

$$\lambda_i(t) = exp\big(a_i + b_i\big(t - \textstyle\sum_j c_{ij}\, S^{(j)}(t)\big)\,\big) \tag{15}$$

with $a_i = \alpha_i + \beta_i\, X_i(0)$, $b_i = \beta_i\, \rho_i$ and $c_{ij} = {}^{\theta_{ij}}/_{\rho_i}$.

2.6. Marked point processes

The conditional intensity function of a point process, conditional on the past history of the temporal events, describes the instantaneous Poisson rate. Suppose we consider the events that are occurring in two-dimensional space, for example longitude and latitude in case of study regarding earthquakes, according to time. Then the history up to but not including time t of events may be denoted by $\mathcal{H}_t$ and may be defined as $\mathcal{H}_t = \{(t_i, x_i, y_i)$ for all i for which $t_i < t\}$. Here t_i is the time of occurrence of the i^{th} event and (x_i, y_i) is its spatial location, that is, longitude and latitude of the location of main shock or aftershock. Then the conditional intensity function as described in (1) may be written as

$$\lambda(t, x, y \mid \mathcal{H}_t) = \lim_{\delta,\xi,\eta \to 0} \frac{1}{\delta\xi\eta} \; P\left[\mathrm{N}_{\delta\xi\eta}(t, x, y) > 0 \mid \mathcal{H}_t\right] \tag{16}$$

where $\mathrm{N}_{\delta\xi\eta}(t, x, y)$ is the number of aftershocks occurring in the space $[t, t+\ \delta) \times [x, x+\ \xi) \times [y, y+\ \eta)$. As stated in section 1, the conditional intensity function of the marked point process should have two components, the ground intensity function which describes the rate at which the aftershocks occur over time and the history of main shock or aftershocks which occurred before the current time and mark distribution. Let $\mathrm{N}_\delta(t)$ be the number of aftershocks in time interval $[t, t+\ \delta)$, then the ground intensity function for the marked point process can be defined as

$$\lambda_g\left(t, \underline{\theta} \mid \mathcal{H}_t\right) = \lim_{\delta \to 0} \frac{1}{\delta} \; P[\mathrm{N}_\delta(t) > 0 \mid \mathcal{H}_t] \tag{17}$$

where $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_m) \in \Theta_\mathrm{m}$ are the parameters. $\lambda_g\left(t, \underline{\theta} \mid \mathcal{H}_t\right)$ can also be simply denoted as $\lambda_g(t \mid \mathcal{H}_t)$. Suppose the mark distribution of the variables x (univariate or multivariate) considered as marks may be denoted as $f(x \mid \mathcal{H}_t)$. Then the conditional intensity function of the marked point process may be considered to have form

$$\lambda(t, x \mid \mathcal{H}_t) = \lambda_g(t \mid \mathcal{H}_t)\, f(x \mid \mathcal{H}_t) \tag{18}$$

In the next section, we consider the estimation of model parameters in a couple of models described above.

3. Likelihood Functions and Parameter Estimation

The likelihood function for the temporal point process models where the conditional intensity function is a function of time only, can be derived in the following manner: Let τ be the time of the occurrence of the last event before time *t*, $F(t \mid \mathcal{H}_\tau) = P\{T \geq t \mid \mathcal{H}_\tau\}$ denote the conditional distribution of the time of occurrence of the next event after time *t* and $f(t \mid \mathcal{H}_\tau)$ be the corresponding conditional density function. Then

$$\lambda(t \mid \mathcal{H}_\tau) = \frac{f(t \mid \mathcal{H}_\tau)}{1 - F(t \mid \mathcal{H}_\tau)}$$

Solving the differential equation,

$$F(t\,|\mathcal{H}_\tau) = 1 - exp\left\{-\int_\tau^t \lambda(u\mid \mathcal{H}_\tau)\,du\right\},$$

hence

$$f(t\,|\mathcal{H}_\tau) = \lambda(t\mid \mathcal{H}_\tau)\,exp\left\{-\int_\tau^t \lambda(u\mid \mathcal{H}_\tau)\,du\right\}$$

Let $\ldots < t_{-2} < t_{-1} < t_0 < T_1 < t_1 < t_2 < \cdots < t_n < T_2 < t_{n+1} < t_{n+2} < \cdots$, where $t_i\,;\, i \in \mathbb{Z}$ are times of occurrence of main shock or aftershocks, and those main shock or aftershock events which occurs within the time interval $[T_1, T_2]$ are explicitly included in the likelihood. The events which occurred before time T_1, if any, are included in the history of the process. Then the log likelihood function of such models will be

$$\log L = \sum_{i:T_1 \le t_i \le T_2} log\,\lambda\left(t_i \mid \mathcal{H}_{t_i}\right) - \int_{T_1}^{T_2} \lambda(t\mid \mathcal{H}_t)\,dt$$

The likelihood function for the marked point process models for which the conditional intensity function is as defined in (18), can be derived similarly as above and will be of the form

$$\log L = \sum_{i:\,t_i \in \mathcal{T}} log\,\lambda\left(t_i, x_i, y_i \mid \mathcal{H}_{t_i}\right) - \int_{\mathcal{T}}\int_{\mathcal{Y}}\int_{\mathcal{X}} \lambda(t, x, y\,|\mathcal{H}_t)\;dx\,dy\,dt$$

where $\mathcal{T} \subseteq \mathbb{R}^+$ is a time interval and $\mathcal{X}$ and $\mathcal{Y}$ are the domains of x and y, respectively which represents the mark variables (for example, longitudes and latitudes in the earthquake study). This process can also be extended to have more than two mark variables and in that case the log likelihood function can be considered as

$$\log L = \sum_{i:\,t_i \in \mathcal{T}} log\,\lambda\left(t_i, \underline{x_i} \mid \mathcal{H}_{t_i}\right) - \int_{\mathcal{T}}\int_{\mathcal{X}} \lambda\left(t, \underline{x}\,|\mathcal{H}_t\right)\;d\underline{x}\,dt$$

where $\underline{x}$ is a multivariate variable referring the mark variables. The general form of the log likelihood function of a marked point process can be expressed as

$$\log L = \sum_{i:\,t_i \in \mathcal{T}} log\,\lambda_g\left(t_i \mid \mathcal{H}_{t_i}\right) - \int_{\mathcal{T}} \lambda_g(t\,|\mathcal{H}_t)\,dt + \sum_{i:\,t_i \in \mathcal{T}} \log f\left(\underline{x_i} \mid \mathcal{H}_{t_i}\right)$$

In different studies the mark density $f\left(\underline{x} \mid \mathcal{H}_t\right)$ is taken as exponential or gamma or Weibull etc. The model parameters can be estimated using the method of maximum likelihood (MLE), by obtaining the system of likelihood equations simultaneously by differentiating the log likelihood function with respect to the parameters to be estimated and equating those equations to zero. The system of likelihood equations for the models discussed in this study does not possess the explicit solution and therefore a suitable iterative method for nonlinear optimization may have to be adopted for solving them.

4. Data Analysis

As an illustration, we consider the data of Kathmandu earthquake happened in the year 2015. The deadly earthquake of magnitude 7.9 measured on Richter scale, shook Nepal and sent tremors through Indian subcontinent, on April 25, 2015 at 11:56 (Nepalese time). The Complete description about the data, data set, and many inference-based studies on this data are available in Shah et al. (2019). The model check and validation were carried out using information functions like: Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). Goodness-of-fit tests based on power law intensity and exponential intensity failed to fit the data. Therefore, we have fitted marked temporal process models with ETAS and SEP models as the ground intensity functions, considering marks as the magnitudes of earthquakes or aftershocks with intensity more than 5. In both the cases, the underlying distribution is assumed as Gamma. The summary of estimates is presented in Table 1.

Table 1. Parameter Estimates for Gamma model

Ground intensity function			
ETAS		SEP	
Parameters	**Estimates**	**Parameters**	**Estimates**
μ	0.001	μ	0.028
A	0.048	k_0	2.800
α	1.963	w	0.067
c	0.769	θ	0.400
p	1.305	β	2.700
θ	2.519	$\log L$	-339.770
β	0.169	AIC	689.540
$\log L$	-193.860	BIC	699.570
AIC	401.720		
BIC	415.769		

5. Conclusions and Future Directions

Temporal data are very sensitive to the assumptions underlying the situation. So one has to attempt to fit different models satisfying the history of events and the event occurrence mechanism, and then choose the best one. Among various competing models it might be possible that some other model still exists which provides the better fit to the data than the models which have already considered. Thus, it is logical to check whether the major features of the given data are captured and reproduced by the assumed models. Graphical procedures are developed for intensifying the features of the data that deviate from the model, if any, as follows: If $\hat{\lambda}_g(t \mid \mathcal{H}_\tau)$ is the estimated ground intensity function which is fitted on the given data, after obtaining the MLEs of parameters of the temporal point process, then the transformed time points may be assumed to follow a stationary Poisson process with rate one. Thus if $t_i; i = 1, 2, \ldots, n$ are the event times, then the sequence of transformed times τ_i will be $\tau_i = \int_0^{t_i} \hat{\lambda}_g(t \mid \mathcal{H}_\tau)\,\mathrm{dt}$. A deviation from a property of $\{\tau_i\}$ from that which is expected from a stationary Poisson process implies the existence of a corresponding feature of the data $\{t_i\}$ that is not captured by the underlying model. The intensity $\hat{\lambda}_g(t \mid \mathcal{H}_\tau)$ represents a model for prediction, whereas the transformed data $\{\tau_i\}$ may be regarded as "noise" or "residuals" of the point process data $\{t_i\}$. This sequence of $\{\tau_i\}$ is referred to as the residual process by Ogata (1988). This is under investigation.

To check the goodness of fit of the model, the event number i can be plotted versus the transformed time τ_i. The plotted points should approximately follow a straight line. Significant departure from the straight line indicates a weakness in the model. Moreover, the slope of the line less than one implies that the transformed times τ_i are too small indicating that the fitted ground intensity function $\hat{\lambda}_g(t \mid \mathcal{H}_\tau)$ is too small, and the slope of the line greater than one implies that $\hat{\lambda}_g(t \mid \mathcal{H}_\tau)$ is too large. The changing pattern of ground intensity function is also an eye opener for understanding the seismological fluctuations of the occurrence. This is also under investigation. Although challenges are many, with the computing power of complex models, this issue can be resolved.

References

Ascher H. and Feingold H. (1984). *Repairable Systems Reliability – Modeling, Inference, Misconceptions and Their Causes.* Marcel Dekker, New York.

Bassin W. M. (1973). A Bayesian optimal overhaul interval model for the Weibull restoration process. *Journal of the American Statistical Association*, **68**, 575-578.

Bassin W. M. (1969). Increasing hazard functions and overhauls policy. *ARMS, IEEE-69 C 8-R*, 173-180.

Bebbington M. and Harte D. (2003). The linked stress release model for spatio-temporal seismicity: formulations, procedures and applications. *Geophysical Journal International*, **154**, 925-946.

Bebbington M. and Harte D. (2001). On the statistics of the linked stress release process. *Journal of Applied Probability*, **38A**, 176-187.

Cox D. R. and Lewis P. A. (1966). *The Statistical Analysis of Series of Events*. Metheun, London.

Crow L. H. (1974). Reliability analysis for complex repairable systems. In *Reliability and Biometry*, Eds. F. Proschan and R. J. Serfling, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 379-410.

Daley D. J. and Vere-Jones D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Second Edition. Springer-Verlag, New York. ISBN 0-387-95541-0.

Duane J. T. (1964). Learning curve approach to reliability monitoring. *IEEE Transactions*, **A-2**, 563-566.

Egesdal M., Fathauer C., Louie K. and Neuman J. (2010). *Statistical and Stochastic Modeling of Gang Rivalries in Los Angeles*. SIAM Undergraduate Research Online.

Hawkes A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society*, **B33**, 438-443.

Hawkes A. G. and Oakes D. A. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, **11**, 493-503.

Kendall D. G. (1949). Stochastic processes and population growth. *Journal of the Royal Statistical Society*, **B11**, 230-264.

Knopoff L. (1971). A stochastic model for the occurrence of main sequence events. *Reviews of Geophysics and Space Physics*, **9**, 175–188.

Lewis P. A. W. (1972). Recent results in the statistical analysis of univariate point processes. In *Stochastic Point Processes*, Ed. P.A.W. Lewis, 1-54. New York: Wiley.

Lewis P. A. W. (1970). Remarks on the theory, computation and application of the spectral analysis of series of events. *Journal of Sound and Vibration*, **12**, 353-375.

Lu C., Harte D. and Bebbington M. (1999). A linked stress release model for historical Japanese earthquakes: coupling among major seismic regions. *Earth Planets Space*, **51**, 907–916.

Ogata Y. (1989). Statistical model for standard seismicity and detection of anomalies by residual analysis. *Tectonophysics*, **169**, 159-174.

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

Ogata Y. (1987). *Long Term Dependence of Earthquake Occurrences and Statistical Models for Standard Seismic Activity (in Japanese)*. Suri Zisin Caku (Mathematical Seismology) II (Ed. M. Saito), ISM Cooperative Research Report 3, 115-124, Institute of Statistical Mathematics, Tokyo.

Ogata Y. and Tanemura M. (1985). Estimation of interaction potentials of marked spatial point patterns through the maximum likelihood method. *Biometrics*, **41**, 421-433.

Reid H. F. (1910). The mechanism of the earthquake. In *The California Earthquake* of April 18, 1906, Report of the State Earthquake Investigation Commission, 2, 16–28, Carnegie Institute of Washington, Washington, DC.

Shah R., Muralidharan K. and Parajuli A. (2019). Temporal point process models for Nepal earthquake aftershocks. *International Journal of Statistics and Reliability Engineering* (to appear).

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

Utsu T. and Seki A. (1955). Relation between the area of aftershock region and the energy of the main shock, *Zisin (Journal of the Seismological Society of Japan)*, 2nd Series, ii. 7, 233-240.

Vere-Jones D. (1978). Earthquake prediction - a statistician's view. *Journal of Physics of the Earth*, **26**, 129–146.

Vere-Jones D. (1970). Stochastic models for earthquake occurrence (with discussion). *Journal of the Royal Statistical Society*, **B32**, 1-62.

Vere-Jones D. and Davies R. B. (1966). A statistical survey of earthquakes in the main seismic region of New Zealand. Part 2. Time series analyses. *New Zealand Journal of Geology and Geophysics*, **9**, 251-284.

Zheng X. and Vere-Jones D. (1994). Further applications of the stochastic stress release model to historical earthquake data. *Tectonophysics*, **229**, 101-121.