International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:7, No:6, 2013

# Hazard Rate Estimation of Temporal Point Process, Case Study: Earthquake Hazard Rate in Nusatenggara Region

Sunusi N., Kresna A. J., Islamiyati A., and Raupong

***Abstract*—**Hazard rate estimation is one of the important topics in forecasting earthquake occurrence. Forecasting earthquake occurrence is a part of the statistical seismology where the main subject is the point process. Generally, earthquake hazard rate is estimated based on the point process likelihood equation called the Hazard Rate Likelihood of Point Process (HRLPP). In this research, we have developed estimation method, that is hazard rate single decrement HRSD. This method was adapted from estimation method in actuarial studies. Here, one individual associated with an earthquake with inter event time is exponentially distributed. The information of epicenter and time of earthquake occurrence are used to estimate hazard rate. At the end, a case study of earthquake hazard rate will be given. Furthermore, we compare the hazard rate between HRLPP and HRSD method.

***Keywords*—**Earthquake forecast, Hazard Rate, Likelihood point process, Point process.

## I. Introduction

POINT process is a stochastic model that can explain the natural phenomena that are random in both space and time. The earthquake is one example of point process. In this model, the earthquake is seen as a random collection of points in space, where each point stated time or / and location of an event.

The earthquake occurrence is generally viewed as a Poisson process. In this process, the occurrence is memory less and independent from the other. In this study, the time between successive earthquakes as random variable decrypted. Stochastic study of earthquake occurrence has been used since many years ago. Approach the probabilistic prediction magnitude earthquake on a particular fault proposed by Rikitake and Ogata [10], [9].

Ferraes estimated the interval of waiting until the occurrence of the next earthquake (earthquake recurrence time) using the concept of conditional probability [6]. This concept explains that if an earthquake does not occur in the time interval since the last earthquakes, the earthquake will occur at the maximum conditional probability. This model explains that the typical large earthquake will be repeated along the same segment of a fault or plate boundary.

In terms of forecasts, the difficulty that arises is the involvement of the difference in the distribution of time since the occurrence of the last event. A number of distribution models often used are: Gaussian, Weibull, log normal, gamma, and Pareto [14]. However, it has yet to be claimed the most appropriate distribution for inter event time of two successive earthquakes.

Earthquake hazard rate estimation studies have been carried out by experts. Vere-Jones [13] and Ogata [9] estimated the earthquake hazard rate using a parametric approach, namely through the point process likelihood equations called Hazard Rate Likelihood Point Process (HRLPP).

The likelihood equation is expressed by Vere-Jones [13] and Ogata [9] as non-linear equations, where the solution is often solved numerically. The method used by Vere-Jones and Ogata is limited to the estimated hazard rate for observation interval. Furthermore, Sunusi et al. [11] using a temporal point process likelihood equations are constructed via Riemann Stieltjes approach to estimate the hazard rate for waiting time is exponentially distributed.

This method uses a set time of occurrence of events in a time interval of observation. Study of hazard rate estimation is also expressed by Darwis et al. [3] and Sunusi [12] using maximum likelihood and match it with the Gompertz models. To renew hazard models on location and time, a sequential approach was used least squares (least square sequential). In addition to the parametric approach, several experts using the non-parametric approach to estimate the hazard rate to be periodic Poisson process [7], [8].

## II. Hazard Rate Estimation

### *A. Hazard Rate Likelihood of Point Process*

Generally, likelihood function is product of probability density function. In point processes, the probability density function is not known. Hence, the likelihood function of point process is approximated by Poisson. The stationary Poisson process on the line is completely defined by [2], in which we use $N(a_i,\ b_i]$ to denote the number of events of the process falling in the half-open interval $(a_i, b_i]$ for $a_i < b_i \leq a_i + 1$:

Sunusi N is with the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Indonesia (e-mail: ntitisanusi@gmail.com).

Kresna A.J is with the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Indonesia (akresnajaya@yahoo.com).

Islamiyati A is with the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Indonesia.

Raupong is with the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Indonesia.

$$P(N(a_i, b_i] = n_i,\ \ i = 1, 2, \dots, k) = \prod_{i=1}^{k} \frac{[\lambda(b_i - a_i)]^{n_i}}{n_i!} e^{-\lambda(b_i - a_i)} \quad (1)$$

where $\lambda(b_i - a_i) = \int_{a_i}^{b_i} \lambda(t)dt$.

Therefore, following directly from (1), that

$$P(N(0, \tau] = 0) = e^{-\lambda(\tau)} \quad (2)$$

is the probability of finding no points in an interval of length $\tau$. This may also be interpreted as the probability that the random interval extending from the origin to the point first appearing to the right of the origin has length exceeding $\tau$. Besides that, the (2) shows that the interval under consideration has an exponential distribution.

Construction of likelihood function takes the more general form [2]:

$$L_{(0,T]}(N; t_1, t_2, \dots, t_N) = exp\left(-\int_0^T \lambda(t)dt + \int_0^T \log \lambda(t) N(dt)\right) \quad (3)$$

Let the times of occurrence are $t_1, t_2, \dots, t_n$ in time interval [S,T] and hazard rate function in parameter form $\lambda(t|\mathcal{H}_t)$, so likelihood of point process is written by:

$$L(\theta|t_1, t_2, \dots, t_n; S, T) = \{\prod_{i=1}^{n} \lambda_\theta(t_i|\mathcal{H}_t)\} exp\left\{-\int_s^T \lambda_\theta(t|\mathcal{H}_t)dt\right\} \quad (4)$$

### *B. Hazard Rate Single Decrement*

In this section, we presented a method to estimate the hazard rate of temporal point process called HRSD. HRSD include likelihood estimation of single decrement and the momen estimation of single decrement [3], [6]. HRSD estimation different from estimation methods have been introduced and used in statistical seismology, that is HRLPP.

HRSD estimation considered as a waiting time estimation problem so that it can be used to forecast earthquake occurrence time for the next period, whereas the previous method of looking at the hazard rate estimation as a problem of estimation of event time occurrence in a time interval of observation.

Another difference is the HRSD estimation calculated for each subsequent period. While the previous approach only estimate the hazard rate in the interval of observation. In the estimation method HRSD, besides many parameters get involved of number of earthquakes ($d_{t_0}$), the parameter of earthquake occurrence time ($s_i$) is considered.

In actuarial studies, hazard rate symbolized by $\mu_{t_0}$, then we used $\mu_{t_0}$ for further discussion. Let $X(t_0) = T - t_0$ states waiting time until the next earthquake occurrence, given the difference in time $t_0$ since the last occurrence of seismic events and T is the time recurrence of the two earthquakes.

Let $\mu, S,$ dan $f$ successively declared as hazard rate, survival function, and the probability density function. Hazard rate can be expressed as [4]:

$$\mu_{t_o} = \lim_{\Delta t_0 \to 0} \frac{P(t_0 < T \le t_0 + \Delta t_0 | T > t_0)}{\Delta t_0} \approx \frac{f(t_0)}{S(t_0)}. \quad (5)$$

By integral we have: $-\mu_y dy = d \ln S(y)$ for $[t_0, t_0 + \Delta t_0]$ so that we found that there was no incident until $t_0 + \Delta t_0$ when it is known there was no incident until $t_0$, is [1]:

$$\begin{aligned} {}_{\Delta t_0}p_{t_0} &= P(t_0 > t_0 + \Delta t_0 | T > t_0) = e^{-\int_{t_0}^{t_0 + \Delta t_0} \mu_y dy} \\ &= e^{-\int_0^{\Delta t_0} \mu_{t_0 + s} ds}. \end{aligned} \quad (6)$$

Suppose $t_0 = 0$, that shortly after the earthquake, was obtained

$${}_{\Delta t_0}p_{t_0} = S(t_0) = P(T > t_0) = exp\left(\int_0^{\Delta t_0} \mu(y)dy\right) \quad (7)$$

where $S(t_0)$ is a survival function. Distribution of recurrence $T$ and the waiting time until the next occurrence of each event is expressed as follows [1]:

$$T \sim {}_{\cdot \Delta t_0}\, p_{t_0} \mu_{t_0} \quad (8)$$

and

$$X \sim {}_{\cdot \Delta t_0}\, p_{t_0} \mu_{t_0 + \Delta t_0} \quad (9)$$

In this expression, ${}_{\cdot \Delta t_0}\, p_{t_0} \mu_{t_0 + \Delta t_0}$ is the chance that an incident occurred between $t_0$ and $t_0 + \Delta t_0$ when it is known there are no events until $t_0$, and

$$\int_0^\infty ({}_{\Delta t_0}p_{t_0} \mu_{t_0 + \Delta t_0}) = 1; \quad \frac{\mathrm{d}}{\mathrm{dt}} ({}_{\Delta t_0}p_{t_0}) = -{}_{\Delta t_0}p_{t_0} \mu_{t_0 + \Delta t_0} \quad (10)$$

Hazard rate estimation using single-decrement approach through Maximum Likelihood Estimate (MLE) method requires information exit time, ie the time when the event arises. Suppose $d_{t_0}$ that states the number of events that occur in interval $(t_0,\ t_0 + 1)$ and $\left(n_{t_0} - d_{t_0}\right)$ state the number of events in $(t_0,\ t_0 + 1)$ and $\left(n_{t_0} - d_{t_0}\right)$ state the number of events after $t_0$. Because the time for each event is different, then the event is considered individually and take the multiplication contribution of each event to the likelihood function. Likelihood L for the i-th event at intervals $(t_i,\ t_i + 1)$ given by the probability density function for the occurrence of events on that interval when it is known that no events until $t_0$. It can be expressed as follows [5]:

$$L_i = f(t_0(i)|T > t_0(i)) = \frac{S(t_0(i))\mu(t_0(i))}{S(t_0)} \quad (11)$$

That contributed to the incident- $i$ in L. If let's say $s_i = t_0(i) - t_0$ is a time of event-$i$ into the interval $(t_0,\ t_0 + 1)$, with $0 < s_i \le 1$, then

$$L_i = \frac{S(t_0 + s_i)\mu(t_0 + s_i)}{S(t_0)} = {}_{\cdot s_i}\, p_{t_0} \mu_{t_0 + s_i} \quad (12)$$

Contribution of number of occurrences $d_{t_0}$ to L is $\prod_{i=1 \cdot s_i}^{d} p_{t_0} \mu_{t_0 + s_i}$. Contribution of $n_{\ o} - d_{t_0}$events which

occur after $t_0+1$ is $\left(p_{t_0}\right)^{n_{t_0}-d_{t_0}}$ where $n_{t_0}$ is the number of events that occur at $t_0$ or after. Thus the total likelihood L is

$$L = \left(1-q_{t_0}\right)^{n_{t_0}-d_{t_0}} \prod_{i=1}^{d_{t_0}} {}_{\cdot s_i} p_{t_0}\mu_{t_0+s_i}$$
$$= \left(p_{t_0}\right)^{n_{t_0}-d_{t_0}} \prod_{i=1}^{d_{t_0}} {}_{\cdot s_i} p_{t_0}\mu_{t_0+s_i} \quad (13)$$

To solve the equation (13) for $\hat{q}_{t_0}$, we need assumption that the distribution of ${}_{\cdot s_i} p_{t_0}\mu_{t_0+s_i}$ is expressed in the $q_{t_0}$. Following this review, we consider $l_{t_0+s}$ is the number of events after $t_0+s$ and we assumed that it was exponentially distributed which is required to declare ${}_{\cdot s_i} p_{t_0}\mu_{t_0+s_i}$.

If $l_{t_0+s}$ states the number of events after $t_0+s$ is an exponential function, then

$$\boldsymbol{l_{t_0+s}} = ab^s \quad (14)$$

For $s=0$, we have $l_{t_0+1} = ab$ . So $b = \frac{\boldsymbol{l_{t_0+1}}}{\boldsymbol{a}} = \frac{\boldsymbol{l_{t_0+1}}}{\boldsymbol{l_{t_0}}}$ . Based on (10)**,** we have

$$l_{\boldsymbol{t_0+s}} = \left(l_{t_0+1}\right)^{\boldsymbol{s}} . \left(\boldsymbol{l_{t_0}}\right)^{\boldsymbol{1-s}} = (p_{\boldsymbol{t_0}})^{\boldsymbol{s}} . l_{t_0}. \quad (15)$$

Thus

$$(p_{t_0})^s = \frac{l_{t_0+s}}{l_{t_0}} = {}_{\cdot s} p_{t_0} \quad (16)$$

and

$${}_{\cdot s} q_{t_0} = 1 - {}_{\cdot s} p_{t_0} = 1 - (p_{t_0})^s = 1 - (1-q_{t_0})^s. \quad (17)$$

Furthermore,

$$\mu_{t_0+s} = -\ln p_{t_0} = \mu. \quad (18)$$

Therefore, by (16) and (18), we obtained a total likelihood as follows:

$$L = (1-q_{t_0})^{n_{t_0}-d_{t_0}} \prod_{i=1}^{d_{t_0}} \mu_{t_0+s}\left(p_{t_0}\right)^{s_i}$$
$$= \mu^{d_{t_0}} exp\left(-\mu\left[\left(n_{t_0}-d_{t_0}\right) - \sum_{i=1}^{d_{t_0}} s_i\right]\right) \quad (19)$$

Log likelihood for (15) is

$$\ell = \ln L = d_{t_0} \ln \mu - \mu\left[\left(n_{t_0}-d_{t_0}\right) + \sum_{i=1}^{d_{t_0}} s_i\right]. \quad (20)$$

The solution of $\frac{d\ell}{d\mu} = 0$ is

$$\hat{\mu} = \frac{d_{t_0}}{(n_{t_0}-d_{t_0})+\sum_{i=1}^{d_{t_0}} s_i}. \quad (21)$$

Because of $q_{t_0}$ is one to one correspondence with $\mu$, then by (21) we have:

$$\hat{q}_{t_0} = 1 - \exp\left(-\hat{\mu}\right) \quad (22)$$

which is the maximum likelihood estimator for $q_{t_0}$.

## III. Case Study

In this section, numerical simulations of HRLPT and HRSD are given. For this case, we use earthquake occurrence data for Nusa Tenggara region which taken from the Engdahl catalog with a magnitude $M \geq 5$. Furthermore, we selected sampling units based on the observation period of 30 years.

Subsequently, to obtain parametric model of hazard rate value, we make an empirical estimates. Furthermore, to find parametric model of hazard rate value, we performed regression to the hazard rate value. Determination of parametric models starting from the simplest model, that is the linear model, furthermore quadratic model, and cubic model. The result of estimation of empirical hazard rate $\lambda(t|\mathcal{H}_t)$ for linear parametric model is:

$$\lambda(t|\mathcal{H}_t) = -0.09588 + 0.01563t\ ,$$

with Mean Square Error (MSE) is 0.0123. Visually, the regression curve for cubic models have followed the pattern of the data. Based on residual normal probability plot, regression curve, and MSE = 0.0036, we concluded that the model with a cubic equation:

$$\lambda(t|\mathcal{H}_t) = -0.1002 + 0.06870t - 0.00790\ t^2 + 0.000247\ t^3$$

represent the hazard rate value. The result of HRSD estimation through maximum likelihood estimates procedure for exponential assumption can be seen in Table I. Based on the above description it can be seen that the estimated hazard rate through HRLPP and HRSD methods were not significantly different or the same.

TABLE I
Summary Hazard Rate Value Through HRLPP And HRSD Method

| No | $(t_0, t_0+1]$ | HRSD $\mu_{t_0}$ | HRLPP $\lambda(t\|\mathcal{H}_t)$ |
|---|---|---|---|
| 1 | (0,1] | 0.0286 | 0.0282 |
| 2 | (1,2] | 0.0296 | 0.0290 |
| 3 | (2,3] | 0.0300 | 0.0299 |
| 4 | (3,4] | 0.0000 | 0.0000 |
| 5 | (4,5] | 0.0313 | 0.0308 |
| 6 | (5,6] | 0.0648 | 0.0635 |
| 7 | (6,7] | 0.0348 | 0.0339 |
| 8 | (7,8] | 0.0723 | 0.0702 |
| 9 | (8,9] | 0.0782 | 0.0755 |
| 10 | (9,10] | 0.0207 | 0.0204 |
| 11 | (10,11] | 0.0211 | 0.0208 |
| 12 | (11,12] | 0.0437 | 0.0426 |
| 13 | (12,13] | 0.0914 | 0.0889 |
| 14 | (13,14] | 0.0244 | 0.0244 |
| 15 | (14,15] | 0.0000 | 0.0000 |
| 16 | (15,16] | 0.0251 | 0.0250 |
| 17 | (16,17] | 0.0259 | 0.0256 |
| 18 | (17,18] | 0.0528 | 0.0526 |
| 19 | (18,19] | 0.0285 | 0.0278 |
| 20 | (19,20] | 0.0581 | 0.0571 |
| 21 | (20,21] | 0.0620 | 0.0606 |
| 22 | (21,22] | 0.1377 | 0.1290 |
| 23 | (22,23] | 0.4711 | 0.4444 |
| 24 | (23,24] | 0.3261 | 0.2667 |
| 25 | (24,25] | 0.7565 | 0.5455 |

The probability that there is no event in interval $(0, t_0]$ is:

$$\cdot_{t_0} p_0 = S(t_0) = exp\left\{-\int_0^{t_0} -0.1562 + 0.09786\, s - 0.01121\, s^2 + 0.000348\, s^3\, ds\right\}$$

$\cdot_{t_0} p_0 = \exp\{0.1562\, t_0 - 0.04893\, t_0^2 + 0.003736667\, t_0^3 - 0.000087\, t_0^4\}$.

So the probability that at least an event occur in the future interval is:

$$\cdot_{t_0} q_0 = 1 - \exp\{0.1562\, t_0 - 0.04893\, t_0^2 + 0.003736667\, t_0^3 - 0.000087\, t_0^4\}$$

TABLE II
RESULT OF PROBABILITY EARTHQUAKE FORECAST FOR WAITING TIME EXPONENTIALLY DISTRIBUTED IN INTERVAL $(t_0, t_0 + n]$, THERE IS NO EARTTHQUAKE UNTIL $t_0$.

| No | $t_0$ (year) | $\cdot_{t_0} p_0$ | $\cdot_{t_0} q_0$ |
|---|---|---|---|
| 1 | 1 | 0 | 1 |
| 2 | 7 | 0.79551 | 0.20449 |
| 3 | 15 | 0.64315 | 0.35685 |
| 4 | 21 | 0.56643 | 0.43357 |
| 5 | 22 | 0.45478 | 0.54523 |
| 6 | 23 | 0.32723 | 0.67277 |
| 7 | 24 | 0.20590 | 0.79409 |
| 8 | 25 | 0.11032 | 0.88968 |
| 9 | 26 | 0.04891 | 0.95109 |
| 10 | 27 | 0.01739 | 0.98261 |
| 11 | 28 | 0.00480 | 0.99519 |
| 12 | 29 | 0.00099 | 0.99901 |
| 13 | 30 | 0.00014 | 0.99985 |

## REFERENCES

[1] N.L. Bowers, N.L., H.U. Gerber, J.C. Hickman, and C.J. Nesbitt. Actuarial Mathematics, *The Society of Actuaries*, 1986. p. 51-59.
[2] D.J. Daley and D. Vere-Jones. An Introduction to the Theory of Point Processes, *Springer, Berlin*, 2003. p.2-19.
[3] S. Darwis, I.W. Mangku, dan N. Sunusi, *Updating Seismic Renewal Model, Far East Journal of Theoretical Statistics*, 2008. 27(1), 101–112.
[4] S. Darwis, S., Sunusi, N., Triyoso, W., dan Mangku, I. W. *Single Decrement Approach for Estimating Earthquake Hazard Rate*, Advances and Applications in Statistica, 2009. 11(2), 229–237.
[5] R. L. London, FSA, Survival Models and Their Estimation, 3rd Edition, *Actex*, 1997. p.113-200.
[6] S.G. Ferraes, *Probabilistic Prediction of the Next Large Earthquake in the Michoacan Fault-segment of the Mexican Subduction Zone*, Geofisica Internacional, 2003. 42, 69–83.
[7] R. Helmers, I.W. Mangku, dan R. Zitikis. *Consistent Estimation of the Intensity Function of a Cyclic Poisson Process*, Journal of Multivariat Analysis, 2003. 84, 19–39.
[8] R. Helmers, I.W. Mangku,, dan R. Zitikis. *Statistical Properties of a Kernel-Type Estimator of the Intensity Function of a Cyclic Poisson Process*, Journal of Multivariat Analysis, 2005. 92, 1–23.
[9] Y. Ogata. *Seismicity Analysis Through Point Process Modeling: A Review*, Pure and Applied Geophysics, 1999.155, 471–507.
[10] T. Rikitake. *Probability of an Earthquake Occurrence as Estimated from Crustal Strain*, Tectonophysics, 1974. 23, 299–312.
[11] N. Sunusi, S. Darwis, dan W. Triyoso. *Estimating Intensity of Point Processes Models Applied to Earthquake Prediction*, Mathematics Journal University Teknologi Malaysia, 2008. 2, 405–411.
[12] N. Sunusi, S. Darwis, W. Triyoso, dan I.W. Mangku, *Study of Earthquake Forecast through Hazard Rate Analysis*, International Journal of Applied Mathematics and Statistics (IJAMAS), 2010. 17(J10), 96–103.
[13] D. Vere-Jones. *Forecasting Earthquakes and Earthquakes Risk*, International Journal of Forecasting, 11, 1995. 503–538.
[14] V. Yilmaz dan H.E. Celik, *A Statistical Approach for Estimating the Probability of Occurrence of Earthquake on the Northern Anatolian Fault Zone*, International Journal of Natural and Engineering Science, 2(2), 2008. 81–86.