Study on Probability of Incidence of Disease Through Point Process Modeling

Sunusi N[1], Nurdin[2]
[1,2] Faculty of Mathematics and Natural Sciences
Hasanuddin University, Indonesia
Corresponding author: Sunusi N, e-mail: ntitisanusi@gmail.com

Abstracts

Forecasting the incidence of the disease in an area at a given time is a very interested study. It is important because it can be provide information early so that everything can be prepared to reduce the risks that may occur. One of the stochastic models that can explain the natural phenomena which occur in random in space and time is the point process. In this study, the incidence of the disease at any given time is considered as a temporal point process with inter event time is exponentially distributed. An interesting thing in this study is the conditional intensity of point process can be used to forecast the occurrence probability of exactly one event in one unit of time in the future. In this study, the parameters of conditional intensity were estimated using single decrement approach. Using the incidence of disease data, we will forecast the probability of incidence of disease in the future time at a certain area.

Key words: Single decrement, random phenomena, exponential inter-event time.

1. Introduction

Until now the incidence of disease in a particular area at certain time is still an interesting thing to study. Straddling the equator, Indonesia has a tropical climate characterized by heavy rainfall, high humidity, high temperature, and low winds. The wet season is from November to March, the dry season from April to October. Rainfall in lowland areas averages 180–320 cm (70–125 in) annually, increasing with elevation to an average of 610 cm in some mountain areas. This shows that Indonesia, which has a high rainfall is very risky to the emergence of various diseases, especially diseases associated with high and low rainfall, such as dengue fever.

Dengue Hemorrhagic Fever (DHF) is a disease that can be dangerous cause of death. Various efforts have been made to address this problem either society or government, but the number of contracting this disease is still not be able to effectively suppressed. Geographical distribution map disease is very useful for studying the relationship between climate and disease. Therefore, we need a distribution map should be able to determine the region anticipation of the implementation of priority programs and outbreak response dengue fever.

The occurrence of dengue fever at a particular location and time is a natural phenomenon whose occurrence is random. One of the stochastic models that can be explained such a phenomenon known as the point process model. In this model, time of disease expressed as points in a certain location. Research on the incidence of the disease has been carried out by previous researchers, including Diggle (2005) and Victor (2005). Generally, estimation of hazard rate using hazard rate likelihood point process (Vere Jones, 1999).

Hazard rate estimation through single decrement has been reviewed by Darwis et al (2009) and Sunusi et al (2010). Hazard rate is defined as the probability of occurrence of an incident after t_0 in $(t_0, t_0 + 1)$, it is known that there was no incident until t_0. If the hazard rate is known, then the corresponding process simulation can be performed. Therefore, we have to obtain an accurate parametric

model of hazard rate. In point process, the incidence time of disease is seen as a random collection of points in a space, where each point stated time or / and location of an event. In this study, we consider the time events between two successive earthquakes as random variables. In modeling forecasts the incidence of disease, time since the last occurrence of disease incidence $(0, t_0]$ can be observed. In this study we examine the case in which the lapse time since the last incident was not depending on the previous interval of the update process (renewal process).

1.1 Point Process

A point process in one dimension (time) is a useful model to sequence random times when an event is occur. A temporal point process is a random process whose realization consist of the times $\tau_j \;;\; \tau_j \in R;\; j = 0, \pm 1, \pm 2, ...$ of isolated events scattered in time. A point process is also known as a counting process or a random scatter. The times may correspond to events of several types. There are a number of important point processes that arise in both theory and practice. (1). The *renewal process* has the property that the intervals between successive points are independent and identically distributed positive random variables. (2). The *Poisson process* has a variety of definitions. A Poisson process is characterized by its rate function (Brillinger, 2002).

1.2 Conditional Intensity (CI)

In point process, the conditional intensity (CI) function defined as the derivative changes in the incidence of occurrence chance plays a very important. This is due to because the CI function characterizes corresponding point process. CI of which depends only on the difference in time from the time emergence of recent events, this is a renewal process. CI corresponding to this process is called hazard rate.

Consider a stochastic process which occurs along the time axis $(-\infty, \infty)$. Let $t_1 < t_2 < \ldots < t_i < \ldots$ denote the arrival times associated with the point process. Furthermore, let N(t) be the counting function, i.e. the number of points that have occurred before the time t (excluding t itself). By assuming that the process is orderly, the conditional intensity function is given by

$\lambda(t|\mathcal{H}_t) = \lim_{\Delta \to 0} \frac{P(t<T\leq t+\Delta|\mathcal{H}_t)}{\Delta}$; where $\mathcal{H}_t$ denotes the information set up to time t inclusive. The conditional intensity function may be interpreted simply as the probability per time unit to observe an event in the next time.

2. Results

2.1 Hazard Rate Likelihood Point Process Estimation

Likelihood function is used to estimate the hazard rate of point process, that is (Daley, D.J, 2003):

$$L_{(S,T]}(N; t_1, t_2, \ldots, t_n) \;=\; \prod_{i=1}^{n} \lambda(t_i) \exp\left(- \int_S^T \lambda(t) dt \right).$$

Equation (1) is used to estimate the hazard rate of point process. Suppose the time of incidence data are $t_1, t_2, \cdots, t_n$ in the time interval $[S, T]$ and the hazard rate in parameter $\lambda(t|\mathcal{H}_t)$, then the likelihood of point process is written by

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

MLE for θ is a parameter vector that maximizes the value of log likelihood, i.e :

$$\ell = \log L_T(\theta|t_1, t_2, \cdots, t_n; S, T) = \sum_{i=1}^{n} log\lambda_\theta(t_i|\mathcal{H}_t) - \int_S^T \lambda_\theta(t|\mathcal{H}_t)dt.$$

In the case of exponentially waiting time, hazard rate is defined by

$$\lambda(t|\mathcal{H}_t) = \upsilon(\tau) \quad = \quad \frac{f(t-\tilde{t})}{1-F(t-\tilde{t})} = \frac{\theta e^{-\theta\tau}}{1-(1-e^{-\theta\tau})} = \theta; \qquad \tau = t - \tilde{t}$$

So, the likelihood equation is

$$L_T(\theta|t_1, t_2, \cdots, t_n; S, T \quad = \quad \left\{\prod_{i=1}^{n} \theta\right\} \exp\left\{-\int_S^T \theta dt\right\}.$$

Thus, we have $\hat{\theta} = \dfrac{n}{(T-S)}$.

2.2 Hazard Rate Single Decrement Estimation

Hazard rate estimation using a single decrement approach through MLE method requires information exit time, that is the time when the disease occurred. Suppose d_{t_0} state the number of incidence of disease that occur in the interval $(t_0, t_0+1]$ and $(n_{t_0} - d_{t_0})$ denote the number of incidence of disease that occur just after t_0. Because of time occurrences for each incidence is different, the incidence of disease is considered individually and take multiplication contribution of each likelihood function of incidence of disease. Likelihood L for the *i*-th incidence in interval $(t_i, t_i+1]$ given by the probability density function for the incidence of disease on the interval if known the incidence of disease did not occur until the time t_0. It can be expressed as follows

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

is the contribution to the incidence-i to L. If we let $s_i = t_0(i) - t_0$ is the time of incidence i-th of disease in the interval $(t_0, t_0+1]$, with $0 < s_i \leq 1$ then

$$L_i = \frac{S(t_0+s_i)\mu(t_0+s_i)}{S(t_0)} = {}_{s_i}p_{t_0}\mu_{t_0+s_i}.$$

The contribution of the number incidence of disease as d_{t_0} on L is $\prod_{i=1}^{d} {}_{s_i}p_{t_0}\mu_{t_0+s_i}$. Contribution of $n_{t_0} - d_{t_0}$ incidence of disease that occur after t_0+1 is $(p_{t_0})^{n_{t_0}-d_{t_0}}$ where n_{t_0} is the number of incidence of disease that occur during t_0 or after. Thus, the total likelihood L is

$$L = (1-q_{t_0})^{n_{t_0}-d_{t_0}} \prod_{i=1}^{d_{t_0}} {}_{s_i}p_{t_0}\mu_{t_0+s_i} = (p_{t_0})^{n_{t_o}-d_{t_0}} \prod_{i=1}^{d_{t_0}} {}_{s_i}p_{t_0}\mu_{t_0+s_i}.$$

To solve this equation in $\hat{q_{t_0}}$, we need assumption that the distribution of ${}_{s_i}p_{t_0}\mu_{t_0+s_i}$ is expressed in the form q_{t_0}. If l_{t_0+s} which specifies the number of incidence of disease after t_0+s is assumed exponentially distributed, then $l_{t_0+s} = ab^s$. Total likelihood is

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

Log likelihood ℓ is:

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

So we have: $\hat{\mu} = \dfrac{d_{t_0}}{(n_{t_0} - d_{t_0}) + \sum_{i=1}^{d_{t_0}} s_i}$.

Because of q_{t_0} and μ_{t_0} are one-one correspond, we have $\hat{q_{t_0}} = 1 - \exp(-\hat{\mu})$. By definition, for a short time interval$(t_0, t_0 + \Delta t_0)$, the probability of occurrence an event is $\mu(t_0)\Delta t_0$, and the probability that no events in interval $(0, t_0)$ is (Ogata, 1999):

$$_{t_0}p_0 = S(t_0) = \exp\left[-\int_0^{t_0} \mu(s)ds\right]$$

Therefore, assuming there are no events in interval$(0, t_0]$, then the chance of at least one event that appears at future intervals is

$$F(t_0) = 1 - S(t_0) = 1 - \exp\left\{-\int_0^{t_0} \mu(s)ds\right\}.$$

2.3. Case Study

In this research, a study case of DHF disease in South Sulawesi (PIP's campus region) with period January 1997 until December 2012.

Table 1. Estimation results for the HRSD with waiting time exponentially distributed using moment procedure with assumption that there is no event yet until t_0.

No	$(t_0, t_0 + 1]$	d_{t_0}	EE	μ_{t_0}	q_{t_0}
1	$(0, 1]$	2	61	0.0328	0.0333
2	$(1, 2]$	2	59	0.0339	0.0345
3	$(2, 3]$	1	57	0.0174	0.0175
4	$(3, 4]$	2	56	0.0357	0.0364
5	$(4, 5]$	8	54	0.1565	0.1702
6	$(5, 6]$	5	46	0.1122	0.1190
7	$(6, 7]$	2	41	0.0488	0.0500
8	$(7, 8]$	19	39	0.5954	0.9048
9	$(8, 9]$	2	20	0.0999	0.1053
10	$(9, 10]$	2	18	0.1110	0.1176
11	(10,11]	2	16	0.1248	0.1333
12	(11,12]	6	14	0.4866	0.6667
13	(12,13]	3	8	0.3935	0.5000

where: EE:total eksposure; d_{t_0} : number of incidence of disease on interval $(t_0, t_0 + 1]$;q_{t_0}:conditional probability of incidence of disease on interval $(t_0, t_0 + 1]$ while there is no incidence until t_0; μ_{t_0}: hazard rate of disease just after t_0.

Table I shows hazard rate estimation with exponentially distributed through procedures moment. For interval $(0, 1]$, the hazard rate is 0.0328; and for interval $(5, 6]$, hazard rate is 0,050. Furthermore, the parametric model for hazard rate value is $\mu = -0.1577 + 0.1354t_0 - 0.02033t^2 + 0.001036t^3$ with mean square error 0.0314.

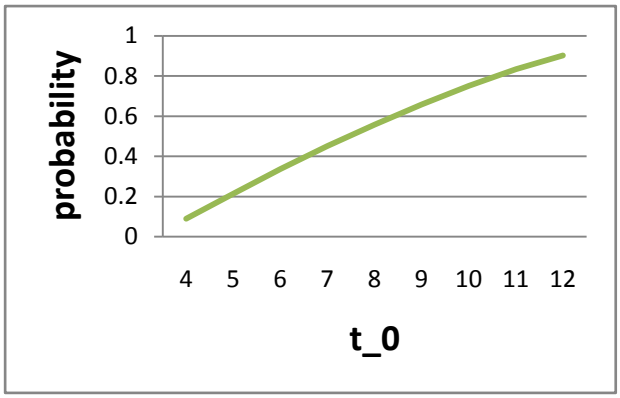

Figure 1: Graph of relationship between earthquake occurrences and time differences t_0 since the last earthquake.

Based on Figure 1, we can see that if earthquake has not happened until $t_0 = 4$, then the probability at least one incident occur at intervals (4, 5] is 0.0897, while for the time interval (11, 12], the probability at least one incident occur is 0.8345, and so on.

3. Conclusions

Point process model can be used to modeling the incidence of disease. Forecast of probability for incidence of disease with exponentially waiting time distributed showed that the longer intervals $(0, t_0]$ since the last incidence of disease, then the probability of a disease to time $(t_0, t_0 + \Delta t_0]$ increasing close to one.

References

Brilinger D, et al (2002), Point Processes Temporal, Encyclopedia of Environmetrics, Vol: 3, pp 1577–1581.

Daley, D. J. and Vere-Jones, D. (2003): An Introduction to the Theory of Point Processes, Springer, Berlin.

Diggle P, et al (2005), Point Process Methodology for on-line Spatio-Temporal Disease Surveillance. Environmetrics, 16: 423-434.

Darwis, S., Sunusi, N., Triyoso, W., and Mangku, I.W. (2009): Single Decrement Approach for Estimating Earthquake Hazard Rate, Advances and Applications in Statistica, 11(2), 229-237.

Ogata, Y. (1999): Seismicity Analysis Through Point Process Modeling: A Review, Pure and Applied Geophysics, 155, 471-507.

Sunusi, N., Darwis, S., Triyoso, W., and Mangku, I.W. (2010): Study of Earthquake Forecast through Hazard Rate Analysis, International Journal of Applied Mathematics and Statistics, 17 (J10), 96-103.

Victor B, et al (2005), A Case Study on Point Process Modelling in Disease Mapping, Image Anal Stereol, 24: 159-168.