UPDATING SEISMIC RENEWAL MODEL

SUTAWANIR DARWIS, AGUS YODI GUNAWAN, 
I WAYAN MANGKU, NURTITI SUNUSI and 
SRI WAHYUNINGSIH

Statistics Research Division
Faculty of Mathematics and Natural Sciences
Institut Teknologi Bandung
Indonesia
e-mail: sdarwis@math.itb.ac.id

Industrial and Financial Research Division
Faculty of Mathematics and Natural Sciences
Institut Teknologi Bandung
Indonesia

Mathematics Department
Faculty of Mathematics and Natural Sciences
Bogor Agricultural Institut
Indonesia

Mathematics Department
Faculty of Mathematics and Natural Sciences
Haluoleo University
Indonesia

Statistics Study Program
Faculty of Mathematics and Natural Sciences
Mulawarman University
Indonesia

2000 Mathematics Subject Classification: 62F10.

Keywords and phrases: recurrence time, hazard function, sequential least squares.

Received September 18, 2008
Abstract

Renewal models, deal with interarrival times, are commonly used to estimate the probability of the next earthquakes on segments for which shocks occur at regular intervals. The time between consecutive earthquakes, so called recurrence time, follows a certain distribution. Uniform, exponential, gamma, Weibull, lognormal, and Brownian passage time are distributions used as renewal model. The model is characterized by seismic hazard function. Kolmogorov and Anderson Darling statistics are used to fit these models to the historical data. The online monitoring data is available as early warning earthquake mitigation program and an updating in time window methodology is needed to accommodate the current information. The sequential updating has recently been taken into account to improve the predictive capability of the model. A better model provides insight and understanding regarding the process. In the past, the parameter updating methodologies are based on historical data and new observations; i.e., a non-sequential approach. Sequential updating approach used only the information from the new incoming observations, and the historical data is considered as prior information. This paper aims to develop a sequential updating methodology to seismic renewal model. The model parameters will be updated sequentially in time, using the information contained in recent observations. The hazard function is estimated using maximum likelihood of recurrence interval and fitted by a hazard parametric model (Gompertz, gamma). A sequential least squares approach is used to update the hazard model based on fixed time window. The proposed approach will be applied to some data taken from an earthquake catalogue using four year time window updating scheme.

Introduction

Associated with earthquake occurrences, the frequency distributions are well approximated by Poisson process. Another important aspect of the earthquake phenomena is the interarrival times. Recurrence times are the interarrival times between earthquakes on a single fault whereas interoccurrence times are the interval times between earthquakes on all faults in a seismic zone (Abaimov et al. [1]). The hazard rate \( \mu_x \) is the probability that an earthquake occurs in the time interval \((x, x + dx)\) under condition that there was no earthquake until \(x\). A renewal model with a time dependent hazard is significantly better than time
independent Poisson model. The lack of regularity in the earthquake occurrence can be explained either by uncertainties in the estimate of occurrence time or by physical interaction between neighbouring faults. Time dependent models of earthquake occurrence have grown up in the context of seismic hazard analysis (SHA) whereas the basic idea is to consider the earthquake occurrence as a quasi periodic process. According to this model, strong earthquakes tend to repeat themselves along the same fault. The seismic gap assumes that the earthquake hazard is small immediately following the previous large earthquake and increases with time since the latest event on a fault.

For a uniform Poisson process, only one parameter is necessary for a complete description of probability density function (pdf). The gap hypothesis needs a more complicated model, named renewal model, whose pdf contains shape parameter in terms of its periodicity. The most popular models are gamma, lognormal, Weibull and Brownian passage time. Renewal models have been used to forecast the time of the next large earthquake occurrence on a fault where large shocks occur repeatedly at approximately regular time interval. Gomez and Pacheco [6] studied a physical basis minimalist model for time dependent earthquake prediction renewal models and concluded that the hazard rate is an increasing function up to a constant limiting value. Zoller et al. [11] studied the individual recurrence time distributions for specific fault.

The parameter updating in time has recently been taken into use with simulation models to improve the predictive capability of the models. The common approach is using a pure nonsequential parameter estimation approach where the current information are combined together with previous information to produce a new revised updated parameter estimate. This approach needs a large computer memory and an efficient approach using sequential processing is needed. This paper aims to develop a sequential updating methodology for seismic renewal model. The model parameters were updated sequentially in time, using the information contained in recent observations. The proposed approach will be applied to some data taken from an earthquake catalogue. A four year time window is used to evaluate the applicability of the proposed method.
Methodology

The hazard function $\mu_x$ is the key to the likelihood theory of earthquake occurrence process. Once the hazard function is known, simulation of the corresponding process can be performed, also the joint density distribution for the maximum likelihood can be obtained. It is important to obtain accurate parametric models for hazard function. Ogata [8] reviewed a class of parametric models for assessment of earthquake risk in a seismic area. Let $T(x) = X - x$ denote the waiting time of the occurrence of the next earthquake given that the time elapsed since the most recent earthquake is $x$, and $X$ is the recurrence time between two earthquakes occurred successively (Yilmaz et al. [10]; Ferrás [5]; Byrdina et al. [3]). If the last earthquake is 2004, and the present is 2008, the elapsed time $x$ is $2008 - 2004 = 4$ years. The hazard function is the probability forecast of an event occurring at a time $x$ (Ogata [8]; Byrdina et al. [3]; Abaimov et al. [1]),

$$\mu_x = \lim_{dx \to 0} \frac{P(x < X \leq x + dx \mid X > x)}{dx} = \frac{\text{pdf}}{1 - \text{cdf}}.$$  

Integrating $-\mu_x dy = d \ln S(y)$ from $x$ to $x + t$ yields the probability that there was no earthquake until $x + t$ given there was no earthquake until $x$,

$$t \pi_x = P(X > x + t \mid X > x) = e^{-\int_x^{x+t} \mu_x dy} = e^{-\int_0^t \pi_{x+t} dt}, \quad t \pi_0 = S(t).$$

where $S(t)$ is the survival function. The distribution of the recurrence time $X$ and the waiting times $T(x)$ are (Bowers et al. [2]; Rundie et al. [9])

$$X \sim_{x} p_0 \mu_x \text{ and } T(x) \sim t \pi_x \mu_{x+t}.$$  

In this expression, $t \pi_x \mu_{x+t}$ is the probability that an earthquake occurs between $t$ and $t + dt$ given not occurred until $x + t$, and

$$\int_0^\infty t \pi_x \mu_{x+t} dt = 1, \quad \frac{d}{dt} t \pi_x = -t \pi_x \mu_{x+t}.$$
The elapsed time since the last earthquake at current time $x$ can be observed, the problem is to predict the occurrence of the next earthquake, given elapsed time $x$. Estimating hazard function using maximum likelihood method is the key point in predicting earthquake occurrences. Let $d_x$ denote the number of earthquake occurrences in time interval $(x, x+1)$. The likelihood for the $i$-th earthquake in $(x, x+1)$ is given by the pdf of $T(x)$. For an earthquake occurs at time $x_i = x + s_i$, the likelihood is given by

$$L_i = f(x_i \mid X > x) = s_i p_x \mu_{x+s_i}.$$  

The contribution of $d_x$ earthquakes in $(x, x+1)$ to total likelihood $L$ is

$$\prod_{i=1}^{d_x} s_i p_x \mu_{x+s_i}.$$  

The contribution of $n_x - d_x$ earthquakes occur after $x+1$ is $(p_x)^{n_x - d_x}$, where $n_x$ is the number of earthquakes that occur at time $x$ or later. Under exponential assumption for $T(x)$,

$$\mu_{x+s_i} = -\ln p_x,$$

$$s p_x = (p_x)^s = e^{-\mu s},$$

the total likelihood for $n_x$ earthquakes is

$$L = p_x^{n_x - d_x} \prod_{i=1}^{d_x} s_i p_x \mu_{x+s_i} = \mu_{x}^{d_x} e^{\mu \left[ (n_x - d_x) - \sum_{i=1}^{d_x} s_i \right]}.$$  

The solution of likelihood equation $\hat{\mu}_{mle} = \arg \max_{\mu} L$ is given by

$$\hat{\mu}_{mle} = \frac{d_x}{(n_x - d_x) + \sum_{i=1}^{d_x} s_i}.$$  

There has been a growing interest in methods for updating in time for renewal models. The common approach is direct technique where the current information are joined together with past data to obtain an updating estimate of model parameters. An alternative approach based on the sequential least squares is proposed where the parameters are
updated sequentially in time using information contained in current observations from monitoring sites. Escobar and Moser [4] presented a general approach to the derivation of changes in regression coefficient estimates. Partition model \( Y = X\beta + \varepsilon \), where \( Y \) is an \( n \times 1 \) vector of observations, \( X \) is an \( n \times p \) design matrix, \( \beta \) is the regression parameter, \( E\varepsilon = 0 \), and \( \text{Cov}(\varepsilon) = \sigma^2 \Sigma \), as

\[
Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix},
\]

where \( Y_i \) is a \( n_i \times 1 \), and \( \text{Cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 \Sigma_{ij}, i, j = 1, 2 \). When \( \Sigma = I \), the formula for adding observations into regression, called sequential least squares, is given by

\[
\beta_u = \beta_w + (X'_1X_1)^{-1}X'_2[I + X'_2(X'_1X_1)^{-1}X'_1]^{-1}(Y_2 - X_2\beta_w),
\]

where \( \beta_w \) is the estimator based on \( Y_1 \) and \( \beta_u \) is the estimator using all the data.

The sequential least squares is closely related to Kalman filtering. Let \( Y_1, Y_2, ..., Y_t \) be a time series. A filter provides a correction for \( X_{t+1} \) taking into account all the points \( Y_1, Y_2, ..., Y_t \) and the \( Y_{t+1} \). If a new observation \( X_{t+1} \) is available, it is more efficient to use the old value of \( \bar{X}_t = \frac{1}{t} \sum_{i=1}^{t} X_i \) and a correction using \( X_{t+1}, Y_{t+1} = Y_t + K(Y_{t+1} - \bar{Y}_t) \),

where \( K = \frac{1}{t+1} \) is called the gain. The weight of \( \bar{Y}_t \) is larger than the weight of \( Y_{t+1} \); the average is corrected by a factor of \( K \) using the difference of \( Y_{t+1} \) and the old value \( \bar{Y}_t \). The similar expression also holds for the variance of a time series \( S_{t+1}^2 = (1 - K)[S_t^2 + K(Y_{t+1} - \bar{Y}_t)] \).

This iterative procedure shows the property of Kalman filter which is a kind of recursive least squares. Meinhold and Singpurwalla [7] used Bayesian approach to derive the Kalman equations. Let \( Y_t = (Y_t, Y_{t-1}, ..., Y_1) \) denote the observations of a response variable \( Y \) which
depend on an observable state $\beta_t$. The relationship between $Y_t$ and $\beta_t$ is specified by the (observation) equation $Y_t = X_t \beta_t + v_t$, $v_t \sim N(0, 1)$, where $X_t$ is the design matrix. The state of regression coefficient is assumed to follow a dynamic equation $\beta_t = \beta_{t-1} + w_t$, $w_t \sim N(0, W_t)$.

The sequential procedure is started at time 0 by choosing initial state $\beta_0 \sim N(\hat{\beta}_0, \Sigma_0)$. After time $t-1$, prior to observing $Y_t$, the posterior distribution of $\beta_{t-1}$ is $\beta_{t-1} | Y_{t-1} \sim N(\hat{\beta}_{t-1}, \Sigma_{t-1})$. The posterior (after observing $Y_t$) for $\beta_t$ has mean $\hat{\beta}_t = \hat{\beta}_{t-1} + R_t X_t^T (1 + X_t R_t X_t^T)^{-1} (Y_t - X_t \hat{\beta}_{t-1})$ and variance $\Sigma_t = R_t - R_t X_t^T (1 + X_t R_t X_t^T)^{-1} X_t R_t$, where $R_t = \Sigma_{t-1} + W_t$.

As an example, the state-space model $Y_t = \beta + v_t$, $v_t \sim N(0, 2)$, $\beta_t = \beta_{t-1} + w_t$, $w_t \sim N(0, 1)$, $\Sigma_0 = 1$ yields an exponential smoothing $\hat{\beta}_t = (1/2)^t \hat{\beta}_0 + \sum_{j=0}^{t-1} (1/2)^{t-j} Y_{t-j}$.

**Result and Discussion**

Table 1 illustrates the sequential least squares updating based on two datasets; old dataset (I) and new observations (II). The regression model based on old dataset is $Y = -67.884 + .906 X_1 - .064 X_2$, and the updating equation is $Y = -64.575 + .905 X_1 - .077 X_2$. The coefficients are modified from $(-67.884, 0.906, -0.064)$ to $(-64.575, 0.905, -0.077)$ due to addition of observations. Updating method and using all data yield the same regression equation $Y = -64.575 + 0.905 X_1 - 0.077 X_2$. Therefore, the least squares updating is applicable for sequential updating hazard rate model.

The exploratory analysis of earthquake catalogue used in this study consists of frequency magnitude relationship, recurrent time distribution, estimating hazard rate and parametric Gompertz modelling. The results using four year time window are summarized in Table 1. The significant differences on Gutenberg-Richter coefficients are due to variability of magnitude and the number of observations. The gamma, lognormal, and Weibull distributions are fitted to the recurrence time. Due to large
number of observations, the fitted models are rejected except for period 1972-1975 (gamma), and 1988-1991 (gamma). The hazard rate of four years period was estimated using maximum likelihood method (Table 3). The sequential updating proceeds in two steps. First, maximum likelihood method is used to obtain the parameter estimate of Gompertz model based on four year time window (Table 3). Second, the sequential least squares is used to update the parameter estimates. The results of sequential updating of hazard rate are shown in Figure 1. The final updating of seismic renewal model is $\ln \mu_x = \ln .669 \times (1.005)^x$. Based on the final model, the mode of the distribution of recurrence time $X$ is given by

$$x_0 = \arg \max_x \quad \mu_x = \frac{\ln(ln c) - \ln(B)}{\ln c} = .924 \text{ year}.$$ 

The elapsed time from policy issue to the death due to an earthquake of the insured is the future lifetime random variable $T(x) = X - x$. The probability that an earthquake occurs between $t$ and $t + dt$ given not occurred until $x + t$ is given by

$$t \mu_x \mu_{x+t} = e^{-\frac{B}{\ln c} c^\gamma t} \times Bc^x \times e^{-136.865 \times (1.005)^t \times .669 \times (1.005)^{x+t}}.$$ 

The net single premium for the $n$-year term insurance with a unit payable at the moment of death is (Bowers et al. [2])

$$EZ = \int_0^\infty \nu_t \mu_x \mu_{x+t} dt,$$

where $\nu = \nu^T$, $T \leq n$, $\nu_t = \nu^t = e^{-\delta t}$ is the interest discount factor from the time of payment back to the time of policy issue, and $\delta$ is the interest rate.

**Conclusion**

This paper presents a sequential updating of hazard rate of renewal model using least squares updating methodology. The hazard rate is estimated using maximum likelihood method. Gompertz model is fitted to
the empirical hazard rate using least squares approach. The updating least squares is used to update sequentially the four years period observations. Weighting observations is the key idea of sequential updating; new observations change the least squares estimate of Gompertz model. The new estimate is a linear combination of the old estimate and the new observations. The difference between the prediction and the actual observations is weighted by a gain to give the correction in parameter estimate. Using the final updating Gompertz model, the next earthquake may occur in the time interval .924 year. The results can be used for calculating the net single premium for the $n$-year term insurance with a unit payable at the moment of death due to an earthquake risk. The case study on an earthquake catalogue demonstrates that the proposed method yields reasonable results. Although the proposed method has limitations, the usefulness of the technique will be for on line updating seismic recurrence model with reasonable accuracy.

References


Table 1. Datasets used for sequential updating least squares of regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$. The old dataset yields the regression $Y = -67.884 + .906X_1 -.064X_2$, and adding dataset II yields the updating regression equation $Y = -64.575 + .905X_1 -.077X_2$.

<table>
<thead>
<tr>
<th>I</th>
<th></th>
<th></th>
<th></th>
<th>II</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$Y$</td>
<td>Obs</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$Y$</td>
</tr>
<tr>
<td>1</td>
<td>130</td>
<td>190</td>
<td>35.0</td>
<td>1</td>
<td>167</td>
<td>220</td>
<td>70.5</td>
</tr>
<tr>
<td>2</td>
<td>174</td>
<td>176</td>
<td>81.7</td>
<td>2</td>
<td>193</td>
<td>215</td>
<td>94.4</td>
</tr>
<tr>
<td>3</td>
<td>134</td>
<td>205</td>
<td>42.5</td>
<td>3</td>
<td>187</td>
<td>198</td>
<td>90.0</td>
</tr>
<tr>
<td>4</td>
<td>191</td>
<td>210</td>
<td>98.3</td>
<td>4</td>
<td>182</td>
<td>204</td>
<td>85.2</td>
</tr>
<tr>
<td>5</td>
<td>165</td>
<td>230</td>
<td>52.7</td>
<td>5</td>
<td>176</td>
<td>197</td>
<td>80.1</td>
</tr>
<tr>
<td>6</td>
<td>194</td>
<td>192</td>
<td>82.0</td>
<td>6</td>
<td>175</td>
<td>199</td>
<td>79.3</td>
</tr>
<tr>
<td>7</td>
<td>143</td>
<td>220</td>
<td>34.5</td>
<td>7</td>
<td>174</td>
<td>208</td>
<td>77.6</td>
</tr>
<tr>
<td>8</td>
<td>186</td>
<td>235</td>
<td>95.4</td>
<td>8</td>
<td>165</td>
<td>200</td>
<td>70.4</td>
</tr>
<tr>
<td>9</td>
<td>139</td>
<td>240</td>
<td>56.7</td>
<td>9</td>
<td>161</td>
<td>207</td>
<td>65.9</td>
</tr>
<tr>
<td>10</td>
<td>188</td>
<td>230</td>
<td>84.4</td>
<td>10</td>
<td>158</td>
<td>196</td>
<td>63.8</td>
</tr>
<tr>
<td>11</td>
<td>175</td>
<td>200</td>
<td>94.3</td>
<td>11</td>
<td>153</td>
<td>216</td>
<td>58.2</td>
</tr>
<tr>
<td>12</td>
<td>156</td>
<td>218</td>
<td>44.3</td>
<td>12</td>
<td>152</td>
<td>200</td>
<td>58.0</td>
</tr>
<tr>
<td>13</td>
<td>190</td>
<td>220</td>
<td>83.3</td>
<td>13</td>
<td>146</td>
<td>205</td>
<td>52.5</td>
</tr>
<tr>
<td>14</td>
<td>178</td>
<td>210</td>
<td>91.4</td>
<td>14</td>
<td>139</td>
<td>200</td>
<td>46.6</td>
</tr>
<tr>
<td>15</td>
<td>132</td>
<td>208</td>
<td>43.5</td>
<td>15</td>
<td>133</td>
<td>214</td>
<td>40.2</td>
</tr>
<tr>
<td>16</td>
<td>148</td>
<td>225</td>
<td>51.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Summary of Gutenberg-Richter (G-R) law, p-value of recurrent distribution (R-D): gamma, lognormal, Weibull, empirical hazard rate $\hat{\mu}_R$, Gompertz model $\ln \mu_R = \ln B + x \ln c$, and number of observations for year 1964-1995 is partitioned in four years period.

<table>
<thead>
<tr>
<th>No</th>
<th>Year</th>
<th>G-R</th>
<th>R-D</th>
<th>Hazard</th>
<th>Gompertz</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1964-1967</td>
<td>4.65; .46</td>
<td>0.01; .01</td>
<td>26.35; .53; .199</td>
<td>.21; .191</td>
<td>1153</td>
</tr>
<tr>
<td>2</td>
<td>1968-1971</td>
<td>8.93; .123</td>
<td>0.01; .01</td>
<td>26.39; .51; 1.98</td>
<td>.22; .87</td>
<td>1373</td>
</tr>
<tr>
<td>3</td>
<td>1972-1975</td>
<td>9.28; .130</td>
<td>35.0; .13</td>
<td>36.37; .56; 1.83</td>
<td>.28; .169</td>
<td>1637</td>
</tr>
<tr>
<td>4</td>
<td>1976-1979</td>
<td>8.64; .14</td>
<td>0.0; .01</td>
<td>48.64; .191; .369</td>
<td>.16; .77</td>
<td>2163</td>
</tr>
<tr>
<td>5</td>
<td>1980-1983</td>
<td>8.41; .14</td>
<td>0.0; .01</td>
<td>26.32; .54; 1.96</td>
<td>.20; .88</td>
<td>2322</td>
</tr>
<tr>
<td>6</td>
<td>1984-1987</td>
<td>8.01; .106</td>
<td>0.0; .01</td>
<td>32.42; .62; 2.12</td>
<td>.26; .88</td>
<td>2736</td>
</tr>
<tr>
<td>7</td>
<td>1988-1991</td>
<td>8.8; .121</td>
<td>10.0; .01</td>
<td>22.35; .61; 2.04</td>
<td>.18; .94</td>
<td>1574</td>
</tr>
<tr>
<td>8</td>
<td>1992-1995</td>
<td>6.06; .68</td>
<td>0.0; .01</td>
<td>23.33; .43; 2.08</td>
<td>.18; .198</td>
<td>4326</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_w$</td>
<td>-1.566</td>
<td>.654</td>
<td>-1.125</td>
<td>.156</td>
</tr>
<tr>
<td>$(B, c)_w$</td>
<td>.209</td>
<td>1.923</td>
<td>.325</td>
<td>1.169</td>
</tr>
<tr>
<td>$\beta_u$</td>
<td>-1.125</td>
<td>.156</td>
<td>-1.954</td>
<td>.073</td>
</tr>
<tr>
<td>$(B, c)_u$</td>
<td>.325</td>
<td>1.169</td>
<td>.385</td>
<td>1.076</td>
</tr>
<tr>
<td>$\beta_w$</td>
<td>-0.938</td>
<td>.079</td>
<td>-0.883</td>
<td>.039</td>
</tr>
<tr>
<td>$(B, c)_w$</td>
<td>.392</td>
<td>1.083</td>
<td>.505</td>
<td>1.041</td>
</tr>
<tr>
<td>$\beta_u$</td>
<td>-0.683</td>
<td>.039</td>
<td>-0.514</td>
<td>.018</td>
</tr>
<tr>
<td>$(B, c)_u$</td>
<td>.505</td>
<td>1.041</td>
<td>.598</td>
<td>1.019</td>
</tr>
</tbody>
</table>
Figure 1. Plot showing the sequential updating renewal model
\( \ln \mu_x = \ln B + x \ln e \) determined via updating regression estimates
of four year time window of period 1964-1995.