THE APPLICATION OF LINEAR INTENSITY MODELS TO
THE INVESTIGATION OF CAUSAL RELATIONS BETWEEN A POINT
PROCESS AND ANOTHER STOCHASTIC PROCESS

Y. OGATA, H. AKAIKE AND K. KATSURA

(Received Mar. 26, 1981; revised Oct. 6, 1981)

Summary

The computational aspect of the fitting of a parametric model for
the analysis of the influence of an input to a point process output is
discussed. The feasibility of the procedure is demonstrated by an arti-
ficial example. Its practical utility is illustrated by applying it to the
analysis of the causal relation between two earthquake series data from
certain seismic regions of Japan.

1. Introduction

Consider a point process defined by the intensity process

\[ \lambda(t) = \mu + \int_0^t g(t-s)dN_s + \int_0^t h(t-s)dX_t, \]

where \( \{N_t\} \) denotes the point process and \( \{X_t\} \) the input process which
may be either a point process or a cumulative process

\[ X_t = \int_0^t x(s)ds \]

of a stochastic process \( x(t) \). Given bivariate data \( \{N_t, X_t; 0 \leq t \leq T\} \),
we are interested in estimating the response functions \( g(t) \) and \( h(t) \) in
(1.1). When \( h(t) \equiv 0 \), this means that there is no causal relation be-
tween the input \( \{X_t\} \) and the output \( \{N_t\} \), while \( g(t) \equiv 0 \) means that
the output process is a doubly stochastic Poisson process whose inten-
sity is modulated only by \( \{X_t\} \).

In [7] we proposed a parametrization by the Laguerre type poly-
nomials

\[ g(t) = \sum_{k=1}^K a_k t^{k-1} e^{-ct} \quad \text{and} \quad h(t) = \sum_{k=1}^L b_k t^{k-1} e^{-ct}. \]
The log partial likelihood in the sense of Cox [3] is then defined by

\[
(1.3) \quad \log L_{\gamma}(\theta) = \int_0^T \log \lambda_s(t) dN_t - \int_0^T \lambda_s(t) dt ,
\]

where \( \theta \) stands for \((\mu, c, a_1, \ldots, a_K, b_1, \ldots, b_L)\). The estimation of the parameters can be realized by applying the method of maximum likelihood to the partial likelihood and the selection of the orders \( K \) and \( L \) is realized by the minimum AIC procedure [1] which minimizes

\[
(1.4) \quad \text{AIC} = (-2) \max (\log \text{partial likelihood}) + 2(\text{number of parameters}) .
\]

In this paper we will discuss the numerical aspect of the process of parameter estimation and order selection and demonstrate the performance of the procedure by applying it to both artificial and real data.

2. The likelihood computation and the minimum AIC procedure

Given a pair of records of occurrence times of two events \( \{ t_i ; i = 1, \ldots, I \} \) and \( \{ \tau_m ; m = 1, \ldots, M \} \) over the time interval \([0, T]\) the partial log likelihood (1.3), with \( \{ t_i \} \) as the output and \( \{ \tau_m \} \) as the input, is given by

\[
(2.1) \quad \log L_{\gamma}(\theta) = \frac{I}{\sum_{i=1}^I} \log \left\{ \mu + \sum_{k=1}^K a_k P_k(i) + \sum_{k=1}^L b_k Q_k(i) \right\} - \left\{ \mu T + \sum_{k=1}^K a_k \sum_{i=1}^I R_k(T-t_i) + \sum_{k=1}^L b_k \sum_{m=1}^M R_k(T-\tau_m) \right\}
\]

where

\[
(2.2) \quad P_k(i) = \sum_{t_j \leq t_i} (t_i - t_j)^{k-1} e^{-c(t_i - t_j)} ,
\]

\[
(2.3) \quad Q_k(i) = \sum_{\tau_m \leq t_i} (t_i - \tau_m)^{k-1} e^{-c(t_i - \tau_m)}
\]

and

\[
(2.4) \quad R_k(t) = \int_0^t t^{k-1} e^{-ct} dt .
\]

When a continuous record \( \{ x(t) \} \) over the time interval \([0, T]\) is given as the record of the input, the intensity process (1.1) is approximated by

\[
(2.5) \quad \lambda(t) = \mu + \sum_{t_i \leq t} g(t - t_i) + \sum_{\tau_m \leq t} h(t - \tau_m) x(\tau_m) \cdot (M/T) ,
\]

where \( M \) is a properly chosen large integer and \( \sigma_m = (m/M) T - 1/(2MT) \),

\( m = 1, \ldots, M \). With the parametrization (1.4), the approximate partial log likelihood is given by

\[
(2.6) \quad \log L_x(\theta) = \sum_{i=1}^T \log \left[ \mu + \sum_{k=1}^K a_k P_k(i) + \sum_{k=1}^L b_k U_k(i) \right] - \left\{ \mu T + \sum_{k=1}^K a_k \sum_{i=1}^T R_k(T-t_i) + \sum_{k=1}^L b_k \sum_{m=1}^M x(\sigma_m) R_k(T-\sigma_m) \right\},
\]

where \( P_k(i) \) and \( R_k(t) \) are given by (2.2) and (2.3), and

\[
(2.7) \quad U_k(i) = \sum_{\sigma_m < t_i} x(\sigma_m)(t_i - \sigma_m)^{k-1}e^{\sigma_m t_i - \sigma_m^2}.\]

The derivation of these partial log likelihoods is discussed in [7].

Given a pair of orders \((K, L)\), the maximum likelihood estimates of the parameters can be obtained by using a non-linear optimization technique developed by Fletcher and Powell [4] which requires only the gradient for the optimization. The gradients of the partial log likelihood functions can easily be obtained by differentiating the above likelihood functions with respect to the parameters. They are given as follows:

\[
(2.8) \quad \frac{\partial \log L_x}{\partial \mu} = \sum_{i=1}^T 1/\Lambda(i) - T,
\]

\[
(2.9) \quad \frac{\partial \log L_x}{\partial a_k} = \sum_{i=1}^T P_k(i)/\Lambda(i) - \sum_{i=1}^T R_k(T-t_i), \quad k = 1, \ldots, K,
\]

\[
(2.10) \quad \frac{\partial \log L_x}{\partial b_k} = \sum_{i=1}^T Q_k(i)/\Lambda(i) - \sum_{m=1}^M R_k(T-\tau_m), \quad k = 1, \ldots, L,
\]

and

\[
(2.11) \quad \frac{\partial \log L_x}{\partial c} = -\sum_{i=1}^T \left\{ \sum_{k=1}^K a_k P_{k+1}(i) + \sum_{k=1}^L b_k Q_{k+1}(i) \right\}/\Lambda(i) \\
+ \sum_{k=1}^K a_k \sum_{i=1}^T R_{k+1}(T-t_i) + \sum_{k=1}^L b_k \sum_{m=1}^M R_{k+1}(T-\tau_m),
\]

where

\[
(2.12) \quad \Lambda(i) = \mu + \sum_{k=1}^K a_k P_k(i) + \sum_{k=1}^L b_k U_k(i).
\]

If we set

\[
(2.13) \quad \Lambda(i) = \mu + \sum_{k=1}^K a_k P_k(i) + \sum_{k=1}^L b_k U_k(i),
\]

the gradients of \( \log L_x^* \) for \( \mu \) and \( a_k \)'s are obtained by the same formula as (2.8) and (2.9), respectively, and
\[
\frac{\partial \log L_{\tilde{T}}}{\partial b_k} = \sum_{i=1}^{J} U_k(i)/A(i) - \sum_{m=1}^{M} x(\sigma_m)R_k(T - \sigma_m), \quad k = 1, \ldots, L,
\]
and
\[
\frac{\partial \log L_{\tilde{T}}}{\partial c} = -\sum_{i=1}^{J} \left[ \sum_{k=1}^{K} a_k P_{k+1}(i) + \sum_{k=1}^{K} b_k Q_{k+1}(i) \right]/A(i)
\]
\[
+ \sum_{k=1}^{K} a_k \sum_{i=1}^{J} R_{k+1}(T - t_i) + \sum_{k=1}^{K} b_k \sum_{m=1}^{M} x(\sigma_m)R_{k+1}(T - \sigma_m).
\]

For the statistics \(P, Q, U\) and \(R\) the following recursive relations, which are useful for the efficient calculation of the likelihood, are available:
\[
P_k(i+1) = (t_{i+1} - t_i)^{k-1} e^{-c(t_{i+1} - t_i)}
\]
\[
+ \sum_{j=1}^{K} \binom{k-1}{j-1} P_j(i)(t_{i+1} - t_i)^{j-1} e^{-c(t_{i+1} - t_i)},
\]
\[
Q_k(i+1) = D_k(t_i, t_{i+1}) + \sum_{j=1}^{K} \binom{k-1}{j-1} Q_j(i)(t_{i+1} - t_i)^{j-1} e^{-c(t_{i+1} - t_i)},
\]
where
\[
D_k(t_i, t_{i+1}) = \sum_{t_i < t_m < t_{i+1}} (t_{i+1} - t_m)^{k-1} e^{-c(t_{i+1} - t_m)}.
\]

Also
\[
U_k(i+1) = F_k(t_i, t_{i+1}) + \sum_{j=1}^{K} \binom{k-1}{j-1} U_j(i)(t_{i+1} - t_i)^{j-1} e^{-c(t_{i+1} - t_i)},
\]
where
\[
F_k(t_i, t_{i+1}) = \sum_{t_i < t_m < t_{i+1}} x(\sigma_m)(t_{i+1} - t_m)^{k-1} e^{-c(t_{i+1} - t_m)}.
\]

Finally,
\[
R_k(t) = (\kappa - 1)R_{k-1}(t) - t^{\kappa-1} e^{-ct}/c.
\]

The AIC defined by (1.4) for the order determination is given by
\[
\text{AIC} (K, L) = (-2) \max_{\theta} \log L_{\tilde{T}}(\theta) + 2K + L + 2,
\]
where \(\theta\) stands for \((\mu, c, a_1, \ldots, a_K, b_1, \ldots, b_L)\). We choose \((K, L)\) that minimizes AIC. A very practical computationally efficient procedure is realized by restricting the exponential coefficient \(c\) to some finite number of candidate values \(c_j\) and minimizing
\[
\text{AIC} (c_j; K, L) = (-2) \max_{\zeta} \log L_{\tilde{T}}(c_j; \zeta) + 2(K + L + 1),
\]
where \(\zeta\) stands for \((\mu, a_1, \ldots, a_K, b_1, \ldots, b_L)\). We choose the triplet \((c_j,
3. Relation to the second order analysis

The problems treated in this paper are related to the spectral analysis of input-output systems of point processes studied by Brillinger [2]. In the discussion of this paper, Cox suggested that an alternative approach would be a likelihood analysis based on models where the intensity function is modulated by the input process. Our present paper is exactly in this line of approach. Given the response functions \( g(t) \) and \( h(t) \), we can obtain the auto- and cross-covariance functions, \( \mu_{11}(t) \) and \( \mu_{12}(t) \) for \( t \geq 0 \), numerically by using the recursive relation of Hawkes [5]

\[
\begin{align*}
\mu_{11}(t) &= \lambda_1 g(t) + \int_0^t g(t-v) \mu_{11}(v) dv + \int_0^\infty g(t+v) \mu_{11}(v) dv \\
&\quad + \int_0^\infty h(t+v) \mu_{12}(v) dv \\
(3.1)\\
\mu_{12}(t) &= \lambda_2 h(t) + \int_0^t g(t-v) \mu_{12}(v) dv + \int_0^t h(t-v) \mu_{22}(v) dv \\
&\quad + \int_0^\infty h(t+v) \mu_{22}(v) dv,
\end{align*}
\]

where \( \mu_{22}(v) \) is the auto-covariance of the input process. In the next section we will use this relation to get initial guesses of the parameters. We will also use the relation for the comparison of the parametric and non-parametric estimates of the covariance functions.

4. Illustration by artificial examples

We wish to illustrate the procedure described in the previous section by some artificial examples. In the first example the input process \( \{x_\sigma(m)\} \) was generated by the relation \( x_\sigma(m) = \exp \{0.3Y_m\} \), where \( \sigma_m = m \) and \( Y_m = 1.394Y_{m-1} - 0.752Y_{m-2} + \varepsilon_m \) and \( \varepsilon_m \) is a standard normal white noise. The record of \( \{x_\sigma(m)\} \) is shown in Fig. 1. The true response functions are given by (1.2) defined with the parameters given in Table 1. We obtained a series of events with the total number of occurrences \( N_T = 1514 \) in the time interval \([0, T] = [0, 1000.0]\) (see [6] for the method of generation). Models with \( L = K \) and up to 6th order were fitted. The maximum of the log likelihood for each order was obtained as follows: As the initial guess of the parameters for the case with \( L = K = 1 \) we used \((\mu, c, a, b) = (\hat{\mu}, \hat{c}, 0, 0)\), where \( \hat{\mu} = \dot{c} = N_T / T = 1.514 \). The case with \( L = K = 2 \) was started with the initial guess
(\(\mu, c, a_2, b_2, a_5, b_5\)) = (\(\hat{\mu}, \hat{c}, \hat{a}_2, \hat{b}_2, 0, 0\)), where \(\hat{\mu}, \hat{c}, \hat{a}_2\) and \(\hat{b}_2\) were the maximum likelihood estimates of the case with \(L=K=1\). A similar procedure was repeated up to the 6th order model \((K=L=7)\). The values of AIC in the sense of (2.20) are listed in Table 3 which shows that the model with \(L=K=5\) is the best. The maximum likelihood estimates of the corresponding coefficients are given in Table 1. The true

\[\begin{array}{cccccccccccc}
\mu & c & a_1 & a_2 & a_3 & a_4 & a_5 & b_1 & b_2 & b_3 & b_4 & b_5 \\
true & 0.10 & 1.00 & 0.23 & -1.00 & 1.74 & -0.83 & 0.12 & 0.00 & 0.49 & -0.36 & 0.08 & 0.02 \\
m.l.e. & 0.04 & 0.99 & 0.20 & -0.84 & 1.29 & -0.55 & 0.07 & -0.03 & -0.09 & 0.86 & -0.55 & 0.11 \\
\end{array}\]

\[\begin{array}{cccccccccccc}
\mu_1 & c & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 \\
true & 0.08 & 5.00 & 0.69 & -12.23 & 97.13 & -289.65 & 338.69 & -65.73 & -111.68 & 50.79 \\
m.l.e. & 0.08 & 4.00 & 0.61 & -10.21 & 76.13 & -215.89 & 272.90 & -153.23 & 32.53 & - \\
\end{array}\]

\[\begin{array}{cccccccccccc}
b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 \\
true & 0.89 & -13.28 & 75.43 & -170.56 & 90.29 & 220.64 & -258.23 & 74.26 \\
m.l.e. & 0.93 & -12.85 & 72.53 & -202.57 & 294.36 & -186.53 & 41.41 & - \\
\end{array}\]

\[\begin{array}{cccccccccccc}
k & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
AIC(k, k) & 1635.18 & 1588.26 & 1588.49 & 1587.76 & 1585.29* & 1589.13 & 1592.81 \\
estimated c & 0.15 & 0.31 & 0.49 & 0.76 & 0.99 & 1.00 & 1.06 \\
\end{array}\]
and estimated response functions are graphically represented in Fig. 2.

As the second example we considered the response functions (1.2) with coefficients given in Table 2. A bivariate series of events was generated by a mutually exciting process with a suitably determined intensity for the second component (see [6] for the algorithm of the generation). The input and output series were with the total numbers of occurrences \(I = 1165\) and \(M = 1010\), respectively. They were obtained on the time interval \([0, 10000.0]\). For the initial guess of the exponential coefficient \(c\) the auto- and cross-correlograms were obtained (see Fig. 4). Significant values of the correlations are observed up to the time lag 4.0. The relation (3.1) suggests that the shapes of the re-
Fig. 3. The true response functions with coefficients in Table 2

Fig. 4. Estimations of auto- and cross-covariance functions

The smooth lines are obtained numerically by solving the equation (3.1) with estimated parameters in the Table 2.
response and covariance functions are not very much different. Thus we guessed that the time range $R$ of the significant response would be around half of this time lag 4.0, i.e., $R=2.0$. By assuming the model with $K=3$ we obtained a rough initial guess $c_0=4/R=2.0$. Alternatively, $c_0=0.98$ was obtained by the direct maximum likelihood estimation of the 0th order model, starting with the initial guess $(\mu, c, a_0, b_0)=(0.1, 0.1, 0.0, 0.0)$. Using $c_0=2.0$ as our initial guess of the exponential coefficient we successively tried $c=0.5$, 1.0, 4.0, 8.0 and 6.0. The AIC values defined by (2.21) were obtained for the models up to the 14th order for each $c$. They are listed in Table 4. The minimum AIC value is attained at $c=4.0$ and $L=K=7$. The estimated coefficients for this case are given in Table 2. Fig. 5 shows the graphs of the estimated response functions with the minimum AIC for each choice of $c$. It can be seen that the graphs for $c=4.0$ which gives the overall minimum of the AIC's are very close to those of the true response functions (Fig. 3). The covariance functions calculated by the relation (3.1) show good fit to the corresponding non-parametrically estimated covariance functions (Fig. 4).

<table>
<thead>
<tr>
<th>$K$</th>
<th>$c_0$</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
<th>4.0</th>
<th>6.0</th>
<th>8.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (Poisson)</td>
<td>7465.16</td>
<td>7465.16</td>
<td>7465.16</td>
<td>7465.16</td>
<td>7465.16</td>
<td>7465.16</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7199.38</td>
<td>7173.59</td>
<td>7198.15</td>
<td>7237.29</td>
<td>7325.05</td>
<td>7259.12</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7117.50*</td>
<td>7117.51</td>
<td>7186.97</td>
<td>7231.10</td>
<td>7250.01</td>
<td>7254.38</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7161.10</td>
<td>7160.47</td>
<td>7156.68</td>
<td>7137.24</td>
<td>7168.77</td>
<td>7202.06</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7184.08</td>
<td>7169.53</td>
<td>7149.49</td>
<td>7133.53</td>
<td>7154.44</td>
<td>7181.38</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7184.65</td>
<td>7168.95</td>
<td>7128.70</td>
<td>7122.85</td>
<td>7126.22</td>
<td>7140.83</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7187.10</td>
<td>7168.22</td>
<td>7118.62*</td>
<td>7126.15</td>
<td>7127.21</td>
<td>7135.07</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7180.06</td>
<td>7166.09</td>
<td>7122.35</td>
<td>7112.87**</td>
<td>7122.41</td>
<td>7127.84</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7178.65</td>
<td>7144.72</td>
<td>7126.20</td>
<td>7112.92</td>
<td>7119.08</td>
<td>7126.51</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>7182.21</td>
<td>7137.27</td>
<td>7130.14</td>
<td>7116.31</td>
<td>7116.79</td>
<td>7125.00</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>7185.23</td>
<td>7139.54</td>
<td>7129.98</td>
<td>7117.02</td>
<td>7112.99*</td>
<td>7118.11*</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>7186.23</td>
<td>7132.61*</td>
<td>7126.95</td>
<td>7120.75</td>
<td>7116.65</td>
<td>7122.11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>7189.77</td>
<td>7133.12</td>
<td>7127.16</td>
<td>7122.77</td>
<td>7119.70</td>
<td>7123.87</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>7193.54</td>
<td>7136.67</td>
<td>7128.54</td>
<td>7121.53</td>
<td>7115.22</td>
<td>7122.20</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>7191.76</td>
<td>7140.56</td>
<td>7124.15</td>
<td>7123.47</td>
<td>7119.22</td>
<td>7126.20</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>7195.76</td>
<td>7144.32</td>
<td>7127.90</td>
<td>7127.30</td>
<td>7122.92</td>
<td>7128.64</td>
<td></td>
</tr>
</tbody>
</table>

* shows the minimum AIC for each $c_0$ and ** shows the overall minimum AIC.

5. Analysis of causality of earthquake data

Utsu [9] discussed the correlation between the intermediate earthquakes in Hida (the region around Takayama city) and the shallower
earthquakes in the central Kanto (the region around Tokyo). He tested the independence between these earthquakes, and concluded that there was a significant dependence which could be attributed to some mechanical connection between the two seismic regions.

Utsu's data is composed of 61 earthquakes with Richter magnitude $M \geq 5.5$ in central Kanto and 16 earthquakes of $M \geq 5.0$ in Hida during the 51 years from 1924 through 1974. Here the data is reproduced after the transformation into $i$-day, i.e., the time scale unit is one-day with the origin 0 of the time being equated to the 1st of January, 1924 (see Table 6). The model (1.1) was fitted to both sets of earthquake data. The values of AIC of the models given by (2.20) with the earthquake series in the central Kanto area as the output and the series in the Hida area as the input are listed in Table 7. The minimum AIC is attained at $L=1$ and $M=1$. The corresponding estimates of the parameters are listed in Table 5. The graphs of the estimated response functions are given in Fig. 6. The result clearly

Table 5

<table>
<thead>
<tr>
<th></th>
<th>$\mu$</th>
<th>$c$</th>
<th>$a_1$</th>
<th>$b_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>m.l.e.</td>
<td>$1.42 \times 10^{-3}$ (shocks/day)</td>
<td>$6.33 \times 10^{-3}$ (1/day)</td>
<td>$1.01 \times 10^{-3}$ (shocks/day)</td>
<td>$8.66 \times 10^{-3}$ (shocks/day)</td>
</tr>
</tbody>
</table>
shows that earthquakes in Hida area do stimulate the occurrence of earthquakes in Kwanto area.

The graph of the estimated intensity process of the Kwanto earthquakes is given in Fig. 7 where the symbols $K$ and $H$ indicate the occurrence times of the earthquakes in Kwanto and Hida area, respectively. Thus it seems that Hida earthquakes will play a significant role in earthquake risk prediction (see Vere-Jones [10]) in the Kwanto area.

To see if a similar effect exists in the opposite direction, the values of AIC of the models with the series in the Hida area as the output and that of Kwanto area as the input were calculated. However, for some choices of the orders $(K, L)$ the maximum log likelihood diverged and the estimated intensity took significantly negative values. In particular either $g(0)$ or $h(0)$ was tending to minus infinity. To avoid this difficulty the parameters $a_i$ and $b_i$ were restricted to non-negative values. This also kept the estimated intensity process positive. This problem is discussed in [8]. The AIC values with * in Table 8 were obtained under these restrictions. The overall minimum of AIC was attained by the Poisson model with $g(t)\equiv h(t)\equiv 0$. This suggests that the earthquakes in the Kwanto area do not stimulate the occurrence of those in the Hida area. To check the validity of the obtained model, we performed simulations of earthquakes of the Kwanto area by using the estimated intensity function with the occurrence times for the Hida area given by the Utsu data of Table 6. One numerical result is given in Fig. 8; see [7] for the method of simulation. Fig. 8 exhibits a similarity with the result given in Fig. 7. We obtained the
Fig. 6. Estimated response functions of Kwanto earthquakes with Hida earthquakes input

Fig. 7. Estimated intensity process of Kwanto earthquakes

The symbols K and H indicate the occurrence times of earthquakes in Kwanto and Hida area, respectively.
Table 7
Input=Hida data

<table>
<thead>
<tr>
<th>Output = Kwanto</th>
<th>AIC (L, K)</th>
<th>Self-exiting</th>
<th>L = 1</th>
<th>L = 2</th>
<th>L = 3</th>
<th>L = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poisson</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>K = 1</td>
<td>-12.0</td>
<td>-33.0</td>
<td>-31.6</td>
<td>-29.6</td>
<td>-27.7</td>
</tr>
<tr>
<td></td>
<td>K = 2</td>
<td>-20.8</td>
<td>-33.6</td>
<td>-31.9</td>
<td>-30.1</td>
<td>-28.0</td>
</tr>
<tr>
<td></td>
<td>K = 3</td>
<td>-18.8</td>
<td>-32.7</td>
<td>-30.7</td>
<td>-28.8</td>
<td>-26.8</td>
</tr>
<tr>
<td></td>
<td>K = 4</td>
<td>-20.1</td>
<td>-30.7</td>
<td>-28.7</td>
<td>-27.3</td>
<td>-25.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-18.2</td>
<td>-28.7</td>
<td>-26.7</td>
<td>-25.1</td>
<td>-23.4</td>
</tr>
</tbody>
</table>

The underlines show the minimum AIC.

Table 8
Input=Kwanto data

<table>
<thead>
<tr>
<th>Output = Hida</th>
<th>AIC (L, K)</th>
<th>Self-exiting</th>
<th>L = 1</th>
<th>L = 2</th>
<th>L = 3</th>
<th>L = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poisson</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>K = 1</td>
<td>41.1</td>
<td>43.6</td>
<td>45.4</td>
<td>47.4</td>
<td>49.1</td>
</tr>
<tr>
<td></td>
<td>K = 2</td>
<td>44.6</td>
<td>45.6*</td>
<td>47.4</td>
<td>49.2</td>
<td>51.2</td>
</tr>
<tr>
<td></td>
<td>K = 3</td>
<td>46.3</td>
<td>47.6*</td>
<td>48.5*</td>
<td>50.1*</td>
<td>52.1*</td>
</tr>
<tr>
<td></td>
<td>K = 4</td>
<td>47.9*</td>
<td>49.6*</td>
<td>51.9*</td>
<td>50.6*</td>
<td>52.1*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44.2</td>
<td>51.6*</td>
<td>53.9</td>
<td>51.9*</td>
<td>49.2*</td>
</tr>
</tbody>
</table>

The underlines show the minimum AIC, and
* shows that AIC was obtained under the restriction g(0) ≥ 0 or h(0) ≥ 0.

Fig. 8. A simulated example of earthquakes with the parameters in Table 5

cumulative distributions of the interval lengths of the series of Figs. 7 and 8 and checked the observation of Utsu [9] that the cumulative distribution displays a broken line type behavior. The behavior can be observed in both Figs. 9 and 10. This suggests that our estimated model is reproducing a basic characteristic of the Utsu data. The broken line phenomenon might be caused by the significant change of intensity after a Hida earthquake.
6. Concluding remarks

Conventionally the model selection is realized by successively applying the likelihood ratio test. The relationship between the AIC and the likelihood ratio statistic is given by

\begin{equation}
(6.1) \quad -(2) \log \lambda(H_0; H_1) = \text{AIC}(H_0) - \text{AIC}(H_1) - 2k
\end{equation}

where the model \( H_1 \) contains the model \( H_0 \) as a restricted family of distributions of \( H_1 \) and \( k \) denotes the degrees of freedom of the chi-square distribution of the likelihood ratio test statistic. Owing to the difficulty in selecting appropriate significance levels, we did not follow this conventional approach. However, if the reader is interested in
the testing procedure the AIC's can be translated into the log likelihood ratios by (6.1). For example, to test the causal relation in the earthquake data, the AIC's given in Table 7 and the relation (6.1) provide necessary information.

For maximum likelihood computation, the selection of the exponential coefficient of the response function is important. By our experience of simulation study, too large or too small values of the exponential coefficient $c$ cause the increase of the order of the model with minimum AIC. However, it may happen that the optimum $c$ is not unique or that the likelihood increases indefinitely as $c \to 0$. In these cases we may use prior information to select an appropriate finite interval for $c$, since $c$ is, roughly speaking, a scale parameter of the influential range of the response function. For example, if we consider an earthquake series, useful prior information for the interval could be obtained from some seismological finding such as Omori's law [11].

Acknowledgement

The authors are grateful to the referees for the helpful comments on the organization of the present paper.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES