

SAPP

Statistical Analysis of Point Processes with R

The Institute of Statistical Mathematics

Jun 1, 2010

1 Introduction

SASeis(Statistical Analysis of Seismicity) is a general software collection for statistical analysis of series of events and seismicity . This program package collects the updated versions of TIMSAC84-SASE (TIMSAC84-SASE version 2), SASeis for Windows (SASeis-W), SASeis Visual Basic (IASPEI-SASeis-VB) and SASeis2006. The source files of TIMSAC84-SASE Version 2 and SASeis2006 are written in FORTRAN77. Programs in TIMSAC84-SASE version2 and SASeis2006 only output numerical files for graphics and especially can be used in free statistical software R.

R is a free programming language or an environment that includes many statistical techniques. R has facilities for data manipulation on arrays and matrices, graphic and foreign language interfaces. Developing a library from SASeis programs, we provide *SAPP* R package for using SASeis from R. All functions in *SAPP* package use .Fortran function of R to communicate between SAPP.dll or libSAPP.so and R. And if necessary some functions display statistical graphs using R graphical procedures.

Each of the software packages TIMSAC84-SASE Version 2 and SASeis2006 are published in Computer Science Monographs ^{1 2} . TIMSAC84-SASE Version 2 is an updated version of TIMSAC84-SASE by Ogata and Katsura, which is published in Part2 of TIMSANC84 ³ . The programs in the original TIMSAC package for point-process analysis were developed for the PC DOS (Utsu and Ogata, 1977) ⁴ . The main features of the updation

¹Y. Ogata, K. Katsura and J. Zhuang (2006) Statistical Analysis of Series of Events (TIMSAC84-SASE) Version 2, *Computer Science Monographs, No.32*, The Institute of Statistical Mathematics, Tokyo.

²Y. Ogata, (2006) Statistical Analysis of Seismicity - updated version (SASeis2006), *Computer Science Monographs, No.33*, The Institute of Statistical Mathematics, Tokyo.

³H. Akaike, T. Ozaki, M. Ishiguro, Y. Ogata, G. Kitagawa, Y. Tamura, E. Arahata, K. Katsura and R. Tamura, (1985) TIMSAC-84 Part 2, *Computer Science Monographs, No.23*, The Institute of Statistical Mathematics, Tokyo.

⁴H. Utsu and Y. Ogata (1997) Computer program package: Statistical Analysis of point processes for Seismicity, SASeis, *IASPEI Software Library for personal computers*, the International Association of Seismology and Physics of Earth's Interior in collaboration with the American Seismological Society, Vol. 6, pp. 13-94.

is that the calculation outputs are simplified by removing the GPSL language for plotting the figures.

In SASeis2006 parameter values in the Omori-Utsu formula and the ETAS (Epidemic Type Aftershock Sequence) model are estimated. The Omori-Utsu formula is an empirical relation for the temporal decay of aftershock rates. The ETAS model is an extension of the Omori-Utsu formula, and it can be used to appropriately evaluate background seismic activity and represent characteristics of seismic activity of the region. The residual between the aftershock activity and the theoretical curve can be used for the diagnostic analysis or the detection of anomalies in the aftershock sequence.

This research is partly supported by "Function and Induction Research Project" held by the Transdisciplinary Research Integration Center at the Research Organization of Information and Sciences, Japan.

2 R functions in the SAPP package

In this section functions and point process models are briefly described. For further information and examples, see the help documentation.

2-1 R functions of TIMSAC84-SASE Version 2

ptspec() :

The periodogram of point process data with the significant bands (0.90, 0.95 and 0.99) assuming the stationary Poisson processes.

linlin() :

Maximum likelihood estimates of linear intensity models of self-exciting point process with another point process input, cyclic and trend components.

simbvh() :

Simulation of bi-variate Hawkes' mutually exciting point processes.

linsim() :

Simulation of a self-exciting point process whose intensity also includes a component triggered by another given point process data and a non-stationary Poisson trend.

2-2 R functions of SASeis2006

momori() :

Maximum likelihood estimates (MLEs) of parameters in the Omori-Uts (modified Omori) formula for the decay of occurrence rate of aftershocks with time, which is formulated as a non-stationary Poisson process.

eptren() :

Maximum likelihood estimates of parameters in a non-stationary Poisson process with a rates function being an exponential polynomial or an exponential Fourier series. The optimal order of the polynomial or series can be determined by the minimized AIC.

etasap() :

Maximum likelihood estimates of parameters in the ETAS model for general seismicity and aftershock sequences.

etasim() :

Simulation of earthquake data set based on the ETAS model.

pgraph() :

To display some elementary statistical features of dataset of a point process or its residual point process.

respoi() :

Occurrence times of earthquakes are transformed using the Omori-Utsu formula with the MLEs. This is called as the residual data and can be used for the diagnostic analysis or the detection of anomalies in the aftershock sequence.

etarpp() :

Occurrence times of earthquakes are transformed using the ETAS model with the MLEs. This is called as the residual data and can be used for the diagnostic analysis or the detection of anomalies in the aftershock sequence.

2-3 an exponential polynomial and an exponential Fourier series (eptren)

The trend component of the conditional intensity function is given by the exponential polynomial

$$f(t) = \exp\{A_1 + A_2t + A_3t^2 + \dots\} \quad (1)$$

and the time varying rate of occurrence of earthquakes in a region is given by the exponential Fourier series

$$f(t) = \exp\{A_1 + A_2\cos(2\pi t/P) + B_2\sin(2\pi t/P) + A_3\cos(4\pi t/P) + B_3\sin(4\pi t/P) + \dots\}. \quad (2)$$

Parameters A_1, A_2, \dots, A_n in (1) and $A_1, A_2, B_2, \dots, A_N, B_N$ in (2) are estimated by the maximum likelihood procedure. These two models belong to the family of non-stationary Poisson processes. The optimal order n can be determined by minimizing the value of the Akaike Information Criterion (AIC). Since a non-stationary Poisson process is assumed in the data, we need careful interpretation of the results when the model is applied to data containing earthquake clusters such as aftershocks, foreshocks, and earthquake swarms.

2-4 a linear intensity model of self-exciting point process (linlin, linsim)

A linear intensity model of self-exciting point process with another point process input, cyclic and trend components is represented by

$$f(t) = P(t) + Q(t) + \Sigma g(t - t_i) + \Sigma h(t - u_j).$$

The summation is done for all i satisfying $t_i < t$ and all j satisfying $u_j < t$. The cyclic part is given by the Fourier series

$$P(t) = C(1)\cos(2\pi t/P) + C(2)\sin(2\pi t/P) + \dots \\ + C(K_C - 1)\cos(K_C\pi t/P) + C(K_C)\sin(K_C\pi t/P),$$

the trend is given by usual polynomials

$$Q(t) = T(1) + T(2)t + \dots + T(K_T)t^{K_T-1}.$$

The response functions to the self-exciting and the external input are given by the Laguerre-type polynomial

$$g(s) = \exp(-Cs)\{A(1) + A(2)s + \dots + A(K_A)s^{K_A-1}\},$$

and

$$h(s) = \exp(-Ds)\{B(1) + B(2)s + \dots + B(K_B)s^{K_B-1}\}$$

respectively, where the scaling parameters in the exponential function, say C and D , can be different. Optimal orders K_C, K_T, K_A, K_B of respective components can be determined by AIC. Other parameters $C(1), \dots, C(K_C), T(1), \dots, T(K_T), A(1), \dots, A(K_A), B(1), \dots, B(K_B), C$ and D are estimated by the maximum likelihood procedure.

2-5 The Omori-Utsu formula (momori, respoi)

The Omori-Utsu formula represent the decay law of aftershock activity in time. The rate of aftershock occurrence at time t is given by

$$f(t) = \mu + K/(t + c)^p,$$

where t is the time measured from the origin time of the main shock. μ, K, c and p are non-negative constants. μ represents constant-rate background seismicity which may be included in the aftershock data. Assuming that the aftershock occurrence is a non-stationary Poisson process, these parameters are estimated by the maximum likelihood procedure.

The cumulative number of earthquakes at time t since t_0 is given by the integration of $f(t)$ with respect to the time t ,

$$F(t) = \mu(t - t_0) + K\{c^{1-p} - (t - t_i + c)^{1-p}\}/(p - 1),$$

where the summation of i is taken for all data event.

2-6 The ETAS model (etasap, etarpp)

The ETAS model is a point-process model representing the activity of earthquakes of magnitude M_z and larger occurring in a certain region during a certain interval of time. The seismic activity includes primary activity of constant occurrence rate μ in time (Poisson process). Each earthquake (including aftershock of another earthquake) is followed by its aftershock activity, though only aftershocks of magnitude M_z and larger are included in the data. The rate of aftershock occurrence at time t following the i -th earthquake (time: $t_i \leq t$, magnitude: M_i) is given by

$$n_i(t) = K \exp[\alpha(M_i - M_z)] / (t - t_i + c)^p,$$

where K, α, c and p are constants, which are common to all aftershock sequences in the region. The rate of occurrence of the whole earthquake series at time t becomes

$$\lambda(t) = \mu + \sum_i n_i(t).$$

The summation is done for all i satisfying $t_i < t$. Five parameters μ, K, c, α and p represent characteristics of seismic activity of the region.

The cumulative number of earthquakes at time t since t_0 is given by the integration of $\lambda(t)$ with respect to the time t ,

$$\Lambda(t) = \mu(t - t_0) + K \sum_i \exp[\alpha(M_i - M_z)] \{c^{1-p} - (t - t_i + c)^{1-p}\} / (p - 1),$$

where the summation of i is taken for all data event.