

COMPUTER SCIENCE MONOGRAPHS

*A Publication
of
The Institute of Statistical Mathematics*

**TIMSAC84: STATISTICAL ANALYSIS OF SERIES OF EVENTS
(TIMSAC84-SASE) VERSION 2**

by

Y. Ogata, K. Katsura and J. Zhuang

April 2006

Computer Science Monographs

Editors

Junji NANAOKO(*in chief*)

Satoshi ITO

Nobuhisa KASHIWAGI

Yumi TAKIZAWA

The series of Computer Science Monographs publishes the statistical software and the articles on computer applications originating at The Institute of Statistical Mathematics. All communications relating to this publication should be addressed to the Editorial Office, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan.

The electronic version of this issue is downloadable at
<http://www.ism.ac.jp/>

TIMSAC84: Statistical Analysis of Series of Events (TIMSAC84-SASE)

Version 2

Ogata, Y., Katsura, K. and Zhuang, J.

The Institute of Statistical Mathematics

TIMSAC84: Statistical Analysis of Series of Events (TIMSAC84-SASE)	1
Overview of the program [eptren]	3
Overview of the program [pgraph]	5
Overview of the program [ptspec]	10
Overview of the program [linlin]	11
Overview of the program [simbvh]	13
Overview of the program [linsim]	14

This program package, TIMSAC84-SASE version2, is an updated version of TIMSAC84-SASE by Ogata and Katsura, which is included in Part 2 of TIMSAC84 (Akaike et al., 1985). The programs in the original TIMSAC package for point-process analysis were developed for the PC DOS (Utsu and Ogata, 1997). The main features of the updation is that the calculation outputs are simplified by removing the GPSL language for plotting figures, so that users can use their own devices or languages for plotting the figures. The source files of the present versions are written in FORTRAN77. For the convenience of users the execution files here were compiled by GNU FORTRAN “g77” on the **cygwin**, and should be executable at least on Windows XP. Then, we provide some source modules of the free graphical software R and postscript files in this library. All the names are corresponding to the names of the FORTRAN sources and the output files.

All programs are downloadable in zip format from the "*Computer Science Monographs*" page in WWW pages of the Institute of Statistical Mathematics. They can be also retrieved from

<http://bemlar.ism.ac.jp/www2/SASeisUpCollection/TIMSAC84-SASE-v2/> or

<http://www.ism.ac.jp/~ogata/Statsei4/download.html>

The latter is the same as the one distributed at the 4th Statistical Seismology Workshop held at Hayama Campus of the Graduate University of Advanced Studies, Japan, during 9-13 January 2006.

Some technical notes are as follows:

1. If the programs crash with an error message "cannot find cygwin1.dll", copy cygwin1.dll to c:\windows\system32.
2. To install updated version of R for windows, MacOS or some other UNIX based systems, please go to "<http://www.r-project.org/>".
3. To install most recent version of cygwin, please go to "<http://www.cygwin.com/>".

About the Data

In order to demonstrate the outputs from each program, we use the following two datasets as input examples. The example data [**Brasting.data**] consists of the occurrence times of 627 blastings at a certain stoneyard with a very small portion of microearthquakes during past 4600days. We can see the long-term trend and the daily cyclic feature from fitting non-stationary Poisson process, the elementary features of point processes and the periodogram (spectral estimation), by using the programs [**eptrend**], [**pgraph**] and [**ptspec**], respectively. Also, [**Poisson.data**] is prepared to see the contrasting difference. These datasets are copied as [**work.data**] in the directory of [**WorkDomain**].

References

- Akaike, H. Ozaki, T. Ishiguro, M., Ogata, Y., Kitagawa, G., Tamura, Y., Arahata, E., Katsura, K. and Tamura, R. (1985) *Time Series and Control Program Package*, **TIMSAC-84**, Computer Science Monograph, No. 22/23, The Institute of Statistical Mathematics, Tokyo, Japan.
- Utsu, T. and Ogata, Y. (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.

Overview of the program [eptren]

This program computes the maximum likelihood estimates (MLEs) of the coefficients A_1, A_2, \dots, A_n in an exponential polynomial

$$f(t) = \exp\{A_1 + A_2 t + A_3 t^2 + \dots\} \quad (1)$$

or $A_1, A_2, B_2, \dots, A_n, B_n$ in a point process model with a conditional intensity taking the form of an exponential Fourier series

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

which represents the time varying rate of occurrence (intensity function) of earthquakes in a region. These two models belong to the family of non-stationary Poisson processes. The optimal order n can be determined by minimize the value of the Akaike Information Criterion (AIC) (cf., Akaike, 1974).

The maximum number of coefficients is limited to 21 due to the dimension set in the FORTRAN source. The coefficients A_1, A_2, \dots, A_{21} can be determined for equation (1) and $A_1, A_2, B_2, \dots, A_{10}, B_{10}$ can be determined for equation (2) at the most.

Since a non-stationary Poisson process is assumed in the data, we need a careful interpretation of the results when the model is applied to data containing earthquake clusters such as aftershocks, foreshocks, and earthquake swarms.

R-modules are also attached in order to display the outputs of numerical data by plotting graphs of the estimated functions.

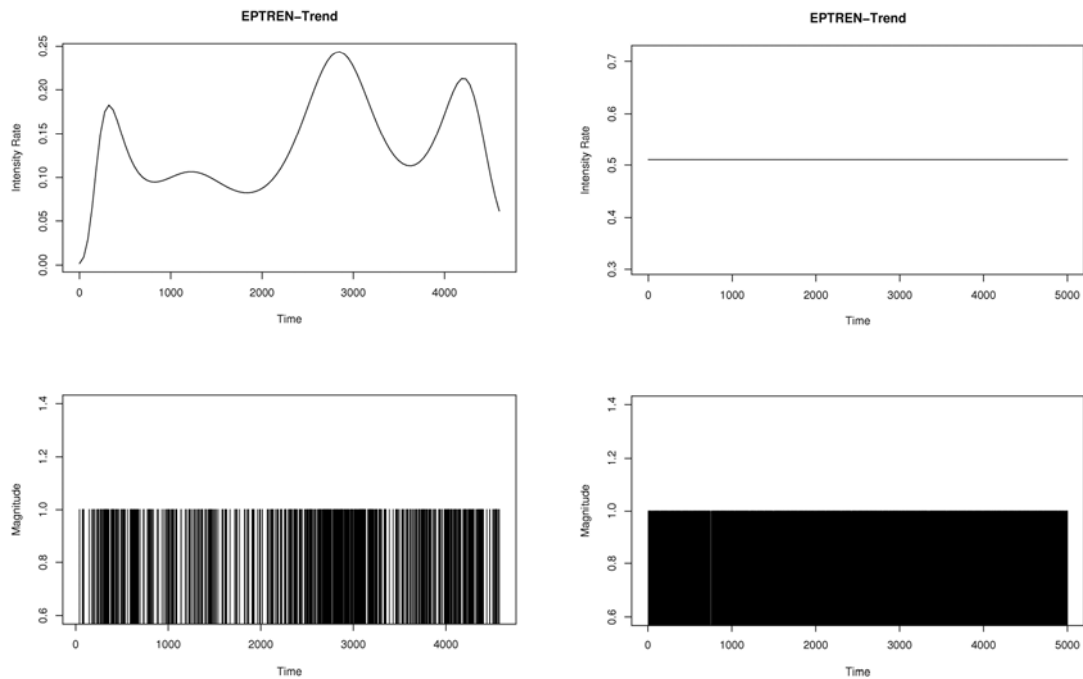
Structure of the program is the following.

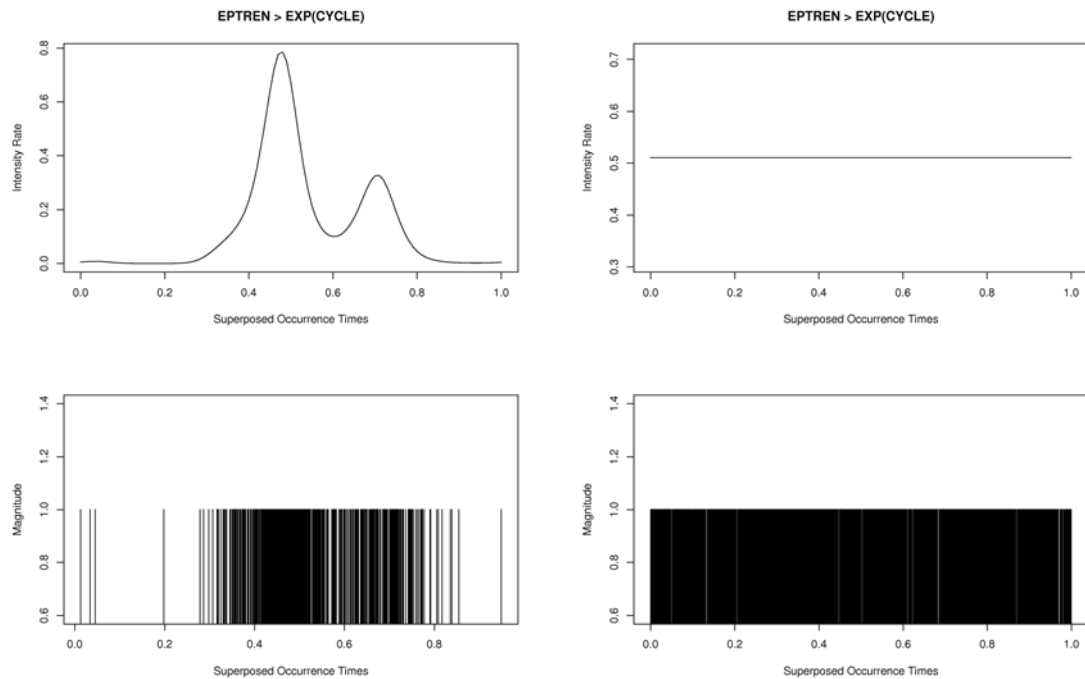
```
[eptren]
  |---[inputs]
  |---[reduc1]
  |---[reduc2]
  |---[davidn]-----[funct]
  |           |---[hesian]
  |           |---[linear]----[funct]
  |
  |---[fincal]
  |---[output]---[printr]---[trenfn]
  |-----[cyclfn]
```

1. The name of the input data file should be always [**work.data**] with required format. Before analyzing any other data, one must record them in required format and name the data file as [**work.data**]. Two test data files [**Brastings.data**] and [**Poisson.data**] for this program are

attached in the directory of [**DataExamples**]. When dataset includes remarkable clusters, the interpretation of the trend and cycles has to be carefully made.

2. The input control file [**eptren.open**] selects either trend or cycle fitting and specify necessary variables.
3. The output files of [**eptren**] are [**out.eptren1**] and [**out.eptren2**]. The R-module [**r.eptrend**] or [**r.epcycle**] is used to plot figures in both console and postscripts [**eptrend.ps**] and [**epcycle.ps**].





4. Calculated records of each program are stored by the name [***.print**] in the directory of [**Calculations**].

This program was originally designed (January 1985) and revised (December 2005) by Yosihiko Ogata, and programmed and also reprogrammed by Koichi Katsura. The R-module was designed and programmed by Jiancang Zhuang and Yosihiko Ogata (December 2005).

References

- Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Control*, **AC-19**, 716-723.
- Lewis, P.A.W. and G.S. Shedler (1976). Statistical analysis of non-stationary series in a database system, *IBM. J. Res. Develop.*, **20**, 465-481.
- Maclean, C.J. (1974). Estimation and testing of an exponential polynomial rate function within the non-stationary Poisson process, *Biometrika*, **61**, 81-86.

Overview of the program [pgraph]

This program makes computations to produce several graphical outputs for the point process data set.

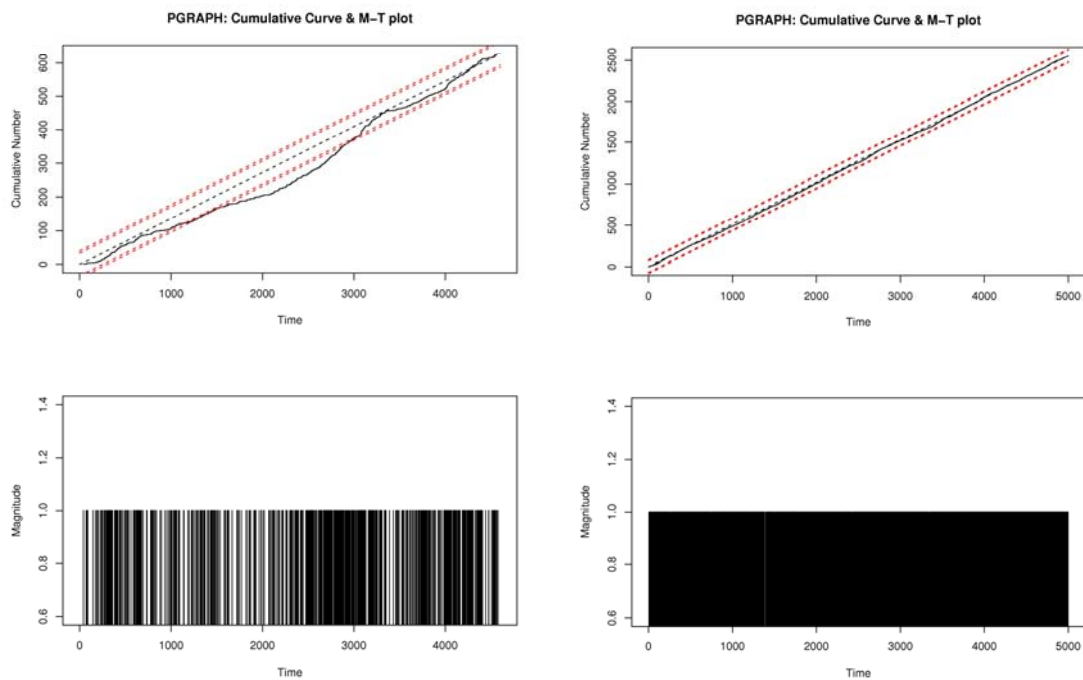
1. Subroutine structures of the FORTRAN program **[pgraph.f]**.

```

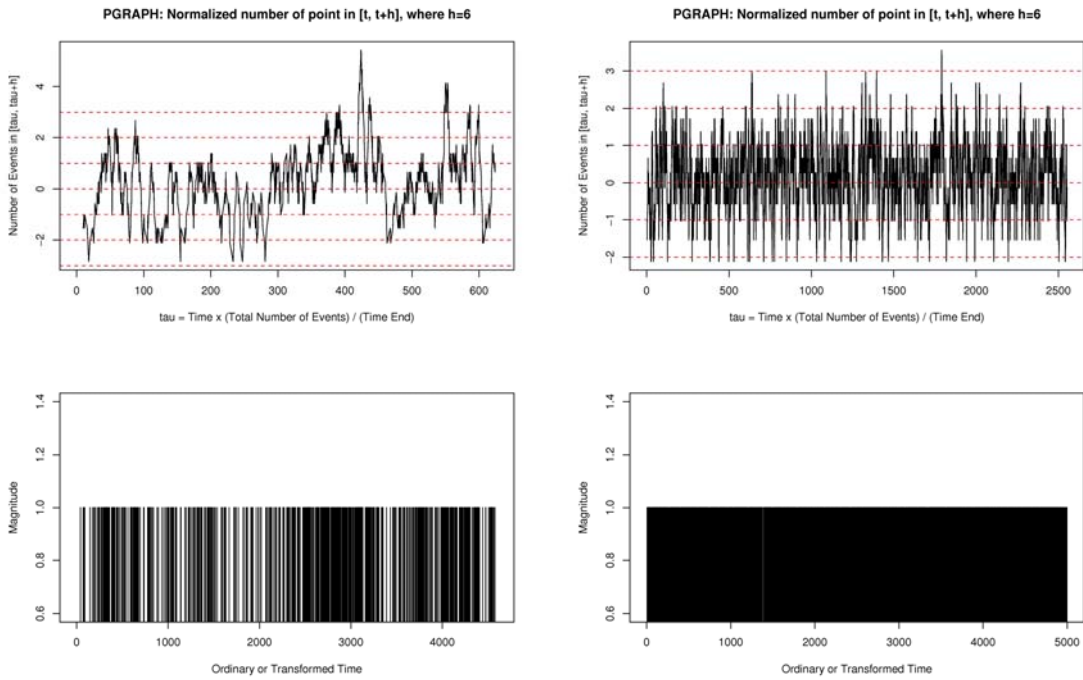
[pgraph]
  |-----[input]
  |-----[count1]---[shimiz]
  |-----[surviv]---[errbr2]---[plsinv]
  |           |-----[unifrm]---[unitsq]
  |
  |-----[palmpr]
  |-----[vtc]
  |-----[vtcprt]

```

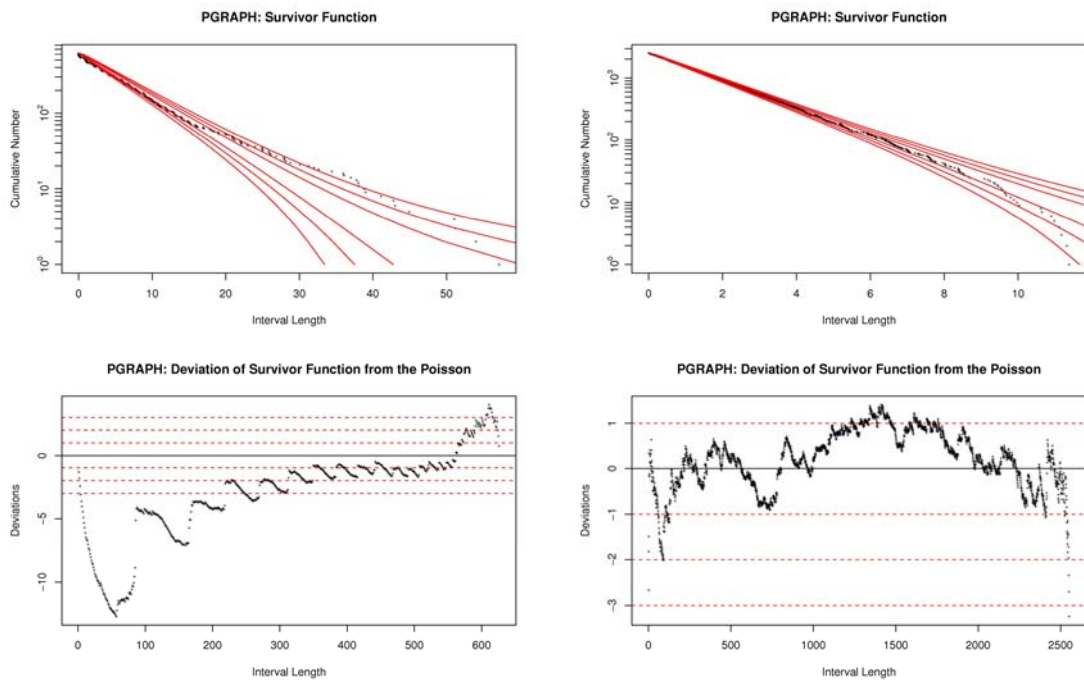
2. Cumulative numbers of events versus time, and positions of spikes (subroutine cumlat) are given in the output file **[out.pgCumMT]**. The R-module **[r.pgCumMT]** can be used to plot figures in both the console and postscript **[pgCumMT.ps]**.



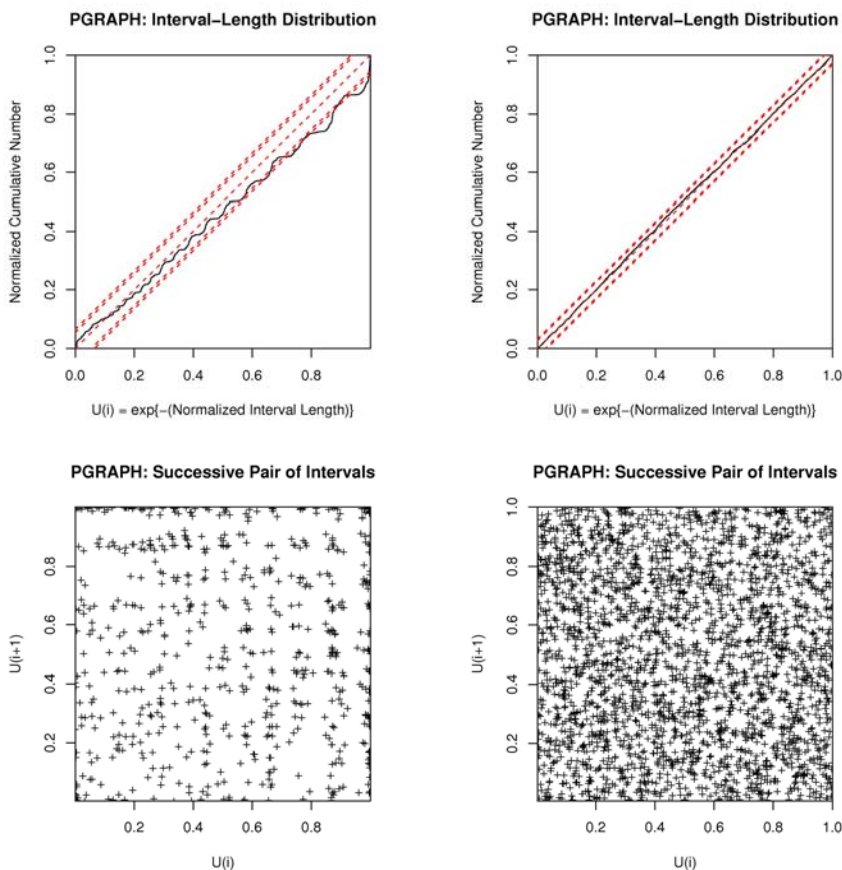
3. The output file **[out.pgPTnum]** gives the series of counted events in the moving interval of a fixed length. The R-module **[r.pgPTnum]** produces figures based on the output file **[out.pgPTnum]** in both the console and postscript **[pgPTnum.ps]**. In the figure dotted lines indicate $i \times$ (standard error), $i = 1, 2, 3$, assuming the stationary Poisson process (subroutine count1 or count2).



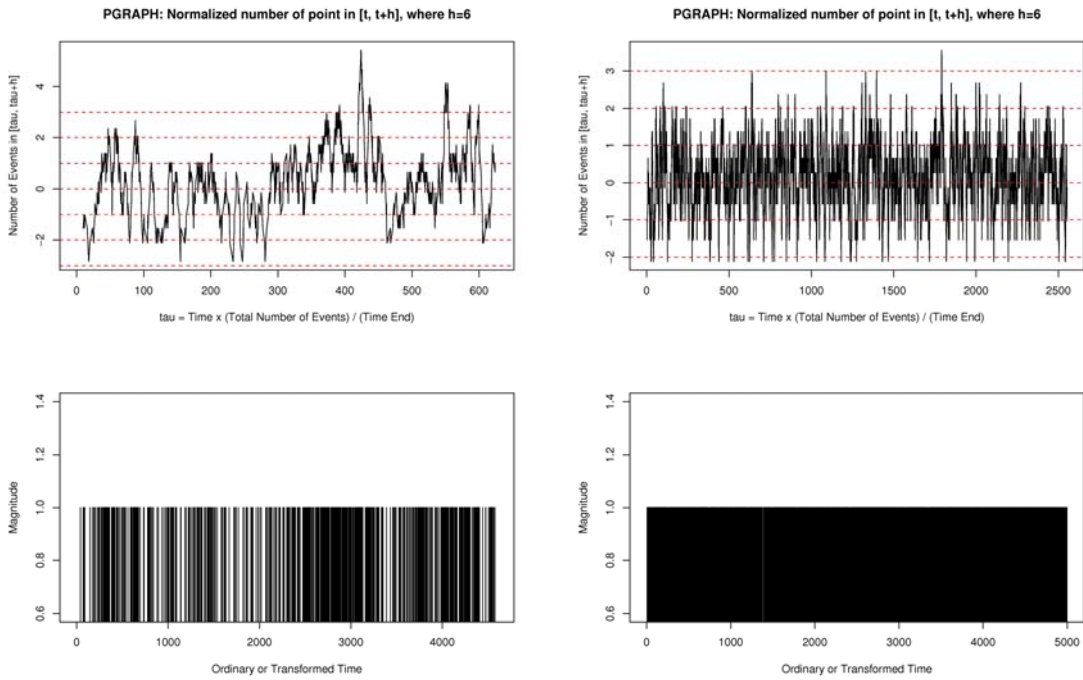
4. The R-module [**r.pgSurviv**] produces figures based on the output files [**out.pgSurviv**] and [**out.pgSurDev**] in both the console and postscript [**pgSurviv.ps**]. The figures include the log survivor curve with $i \times$ (standard error), $i = 1, 2, 3$, based on the assumptions of the stationary Poisson process, and a similar graph in which (x, y) plots are rotated and shifted in such a way that the standard error lines and expectation lines are parallel (subroutine `surviv`).



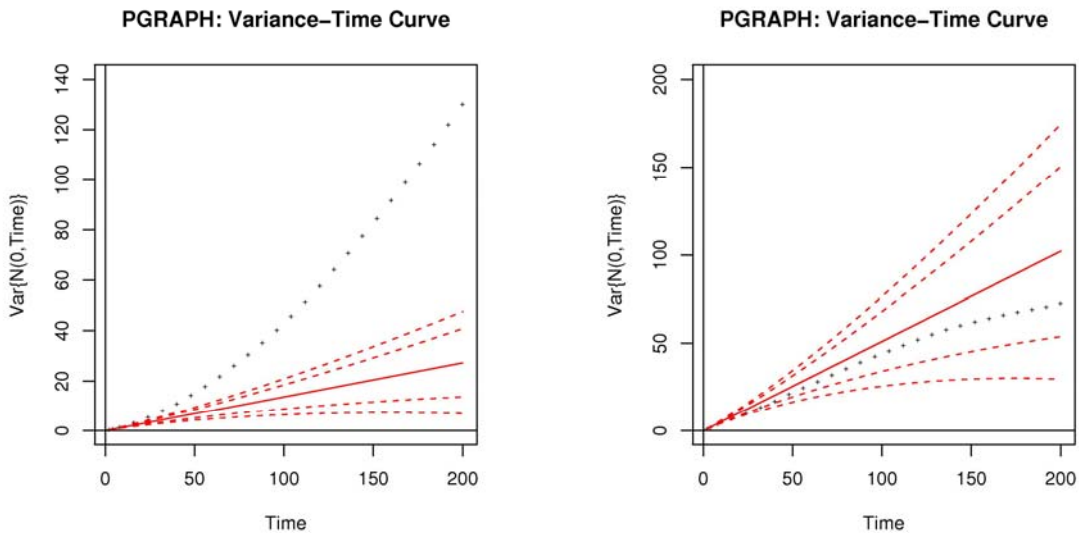
5. The R-module [**r.pgInterP**] plots two figures based on the output files [**out.pgInter1**] and [**out.pgInter2**] in both the console and postscript [**pgInterP.ps**]. The first diagram is on the empirical distribution of $u(i) = \exp\{-\mu x(i)\}$ where $x(i)$ is the i -th interval length between consecutive points, and $\mu = (\text{number of events}) / (\text{length of whole interval})$. The dashed lines of .95 and .99 significance bands of the two-sided Kolmogorov-Smirnov test assuming the uniform distribution is given. The related graph of $(u(i), u(i+1))$ plots are also carried out (subroutine uniform).



6. The R-module [**r.pgPalm**] produces, in both the console and postscript [**pgPalm.ps**], a figure estimated intensity $p(t)$ under the palm probability based on the output file [**out.pgPalm**]. The intensity $p(t)$ is related to the covariance density $c(t)$ by the relation of $c(t) = \mu\{p(t) - \mu\}$, where μ is the mean intensity of the point process. The 0.95 and 0.99 error bands are provided assuming the stationary Poisson process (cf., subroutine palmxy, palmpr).



7. The R-module [**r.pgVTC**] produces, in both the console and postscript [**pgVTC.ps**], a figure of the estimated variance-time curve with the 0.95 and 0.99 error lines assuming the stationary Poisson process (subroutine vtcxyp, vtcprt), based on the output file [**out.pgVTC**].



8. The input data file for this program should always be named [**work.data**]. Before analyzing any other data, recorded them in the same format and name the data file as [**work.data**]. Two example input data files [**Brastings.data**] and [**Poisson.data**] are attached in the directory of

[DataExamples]. When the dataset includes remarkable clusters, one must be careful when making interpretation to the trend and cycles.

9. The input control file [**pgraph.init**] gives and explains necessary variables.

10. Calculation records are stored in the file [***.print**] in the directory of [**Calculations**].

This program was originally designed (January 1985) and revised (December 2005) by Yosihiko Ogata, and programmed and reprogrammed by Koichi Katsura. The R-module was designed and programmed by Jiancang Zhuang and Yosihiko Ogata (December 2005).

References

Cox, D.R. and Lewis, P.A.W. (1966) *The Statistical Analysis of Series of Events*, Methuen & Co. Ltd, London; also published from Chapman & Hall, London.

Overview of the program [ptspec]

This program provides the periodogram (the spectrum estimate) of point process data with the significant bands (0.90, 0.95 and 0.99) of the maximum power in searching a cyclic component, for stationary Poisson processes. Powers of interested frequencies or periods are plotted by + sign (in red color).

Structure of the program:

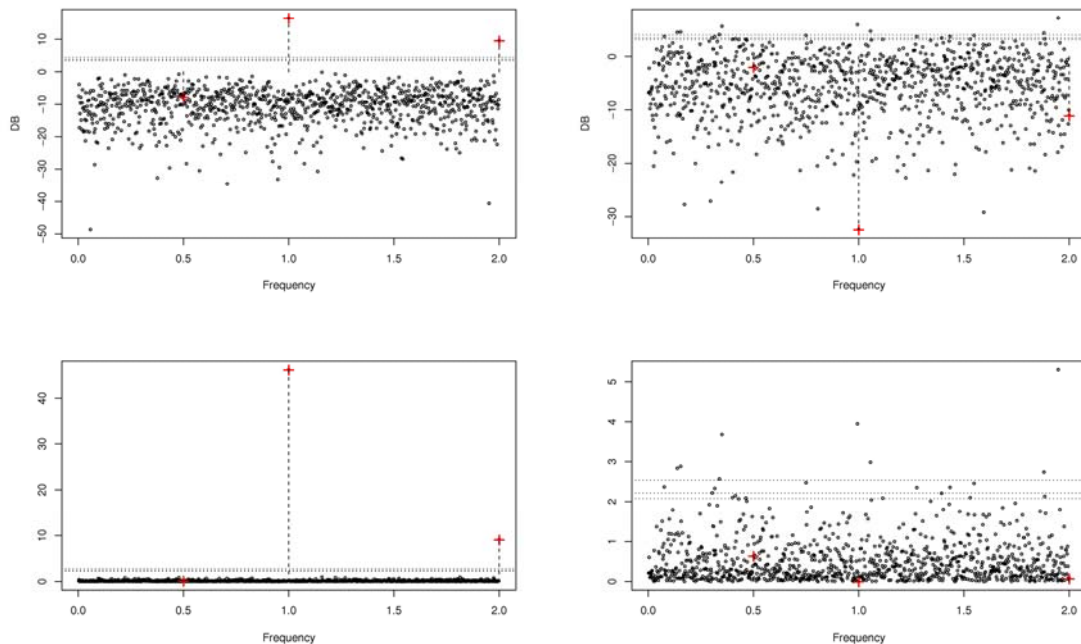
```
[ptspec]
  |--[input]
  |--[cycle]
  |--[period]
  |--[printr]
  |--[smooth]
```

This is originally designed by Yosihiko Ogata and Tohru Ozaki, and programmed by Koichi Katsura and T. Ozaki (January 1985), and modified by Katsura and Ogata (December 2005).

R-modules are also attached in order display the outputs of numerical data for plotting graphs of the estimated functions.

1. The data file for this program should be always named [**work.data**]. Before analyzing any other data, one should record them in required format and name the data file as [**work.data**]. Two test data files [**Brastings.data**] and [**Poisson.data**] are attached in the directory of [**DataExamples**].

2. The input control file is `[ptspec.init]`, which provides necessary variables to read by the program.
3. The output file of `[ptspec]` is `[out.ptspec]`. The R-module `[r. ptspec]` displays the outputs by plotting graphs of the periodogram in both the console and postscript `[ptspec.ps]`.



4. Calculation records of `[ptspec]` are stored by the name `[ptspec.print]` in the directory of `[Calculations]`.

The R-module was designed and programmed by Jiancang Zhuang and Yosihiko Ogata (December 2005).

Reference

- Vere-Jones and Ozaki, T. (1982). Some examples of statistical estimation applied to earthquake data,
 1. Cyclic Poisson and self-exciting models, *Ann. Inst. Statist. Math.*, **34**, pp. 189-207.

Overview of the program `[linlin]`

This program performs the maximum likelihood estimates of linear intensity models of self-exciting point process with another point process input, cyclic and trend components. The cyclic part is given by the Fourier series, the trend is given by usual polynomials. The response functions to the self-exciting and the external inputs are given by the Laguerre-type polynomials, where the scaling

parameters in the exponential function, say c and d , can be different. However it is advised to fit a model without the input component firstly to estimate an initial c , and then to fit a model with both the self-exciting and the external input component to estimate an initial d with c being fixed (this means that the gradient corresponding to c is fixed at 0), which are good initial estimates for the c and d of the mixed self-exciting and input model. It is recommended to use Akaike Information Criterion (AIC; Akaike, 1974) for selecting optimal orders of respective components. Finally, it should be noted that estimated intensity sometimes happens to be negative on some parts of the time interval outside the neighbourhood of events. This take place more easily the larger the number of parameters. This causes some difficulty in getting the MLE, because the negativity of the intensity contributes to the seeming increase of the likelihood. See the references below.

Structure of the program

```
[linlin]
  |---[input]
  |---[comfac]
  |---[cycle]
  |---[dav]-----[davidn]-----[funct1 or 2]
                                |---[hesian]
                                |---[linear]---[funct1 or 2]
```

Remarks:

1. The input data file for this program is always named **[linlin.data]**. Before analyzing any other data, record them in the same format and name the file **[linlin.data]**.
2. The input control file is **[linlin.init]**, in which you can select a model with a possible combination of the trend, the self-exciting or/and the external response components, and in which you can specify necessary variables and initial parameter values.
3. Calculation results of **[linlin]** are stored in the file **[linlin.print]**. In the directory of **[Calculations]**, there is an example file **[linlin.print]** for your initial check of the program.

This program was originally designed (January 1985) and revised (December 2005) by Yoshihiko Ogata, and programmed and also reprogrammed by Koichi Katsura.

References

- Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Control*, **AC-19**, 716-723.
- Ogata, Y. and Akaike, H. (1982). On linear intensity models for mixed doubly stochastic Poisson

and self-exciting point processes, *J. Roy. Statist. Soc. Ser. B*, **44**, pp. 102-107.

Ogata, Y., Akaike, H. and Katsura, K. (1982). The application of linear intensity models to the investigation of causal relations between a point process and another stochastic process, *Ann. Inst. Statist. Math.*, **34**. pp. 373-387.

Ogata, Y. and Katsura, K. (1986). Point process model with linearly parametrized intensity for the application to earthquake data, *Essays in Time Series and Allied Processes* (Festschrift for Prof. E.J. Hannan), J. Gani and M.B. Priestley eds., *J. Appl. Probab.* **23A**, pp.291-310.

Overview of the program [simbvh]

This program performs the simulation of bi-variate Hawkes' mutually exciting point processes. Non-stationary Poisson trends can also be included in the both intensity functions. The trend is given by a usual polynomial and the response functions to the self-exciting and the external inputs are given by the Laguerre-type polynomials, where the scaling parameters in the exponential functions, say c , d , $c2$ and $d2$, can be different. More detailed explanation of parameters and variables are given in the control file [**simbvh.init**].

The simulated data can be displayed in the console and be saved in the file [**simbvh.print**]. In order to view the graphical features of the data, save each of the two components of the simulated data as the name of [**work.data**] with the same format as the original [**work.data**], and use program [**pgraph**].

Finally, it should be noted that the parameters for simulation must be specified such that the conditional intensity function be positive throughout the simulation time interval.

Structure of the program

```
[simbvh]
  |---[input]
  |---[initl]
  |---[simda]----[comfac]
  |      |-----[duf]
  |      |-----[unifor]
  |      |-----[fx]-----ptrend]
  |
  |---[output]
```

Remarks:

1. No input data file for this program is necessary.
2. The input control file [**simbvh.init**] selects a simulation of either self-exciting point processes or self and mutually exciting processes and also specifies necessary variables.
3. Calculation records of [**simbvh**] are stored in [**simbvh.print**] in the directory of [**Calculations**].

This program was originally designed (January 1985) and revised (December 2005) by Yosihiko Ogata, and programmed and also reprogrammed by Koichi Katsura.

References

- Hawkes, A.G. (1971). Spectra of some self-exciting and mutually exciting point processes, *Biometrika*, **58**, pp. 83-90.
- Ogata, Y. (1981). On Lewis' simulation method for point processes, *IEEE Information Theory*, **IT-27**, pp. 23-31.
- Ogata, Y. and Akaike, H. (1982). On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes, *J. Roy. Statist. Soc. Ser. B*, **44**, pp. 102-107.
- Ogata, Y., Akaike, H. and Katsura, K. (1982). The application of linear intensity models to the investigation of causal relations between a point process and another stochastic process, *Ann. Inst. Statist. Math.*, **34**, pp. 373-387.
- Ogata, Y. and Katsura, K. (1986). Point process model with linearly parametrized intensity for the application to earthquake data, *Essays in Time Series and Allied Processes* (Festschrift for Prof. E.J. Hannan), J. Gani and M.B. Priestley eds., *J. Appl. Probab.* **23A**, pp. 291-310.

Overview of the program [**linsim**]

This program performs 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. The trend is given by usual polynomial, and the response functions to the self-exciting and the external inputs are given by the Laguerre-type polynomials, where the scaling parameters in the exponential functions, say c and d , can be different. More detailed explanation of parameters and variables are given in the control file [**linsim.init**].

The simulated data are displayed on the console and also save in the file [**linsim.print**] with the

same format as the original [**work.data**]. In order to view the graphical features of the data by using the program [**pgraph**], change the name of the file for the simulated data to [**work.data**]. Similarly, the data file of external input [**linsim.data**] can be also viewed by change its name to [**work.data**].

Finally, it should be noted that the parameters for simulation must be specified such that the conditional intensity function be positive throughout the simulation time interval.

Structure of the program

```
[linsim]
  |---[input]
  |---[initl]
  |---[simxx]----[comfac]
  |      |-----[duf]
  |      |-----[unifor]
  |      |-----[fx]-----[ptrend]
  |
  |---[output]
```

Remarks:

1. The input data file for this program should be always named [**linsim.data**]. Before using any other data as the external input point process for the simulation, one must save the data in the required format and name the data file as [**linsim.data**].
2. The input control file chooses the combination of the trend, the self-exciting and the external input components and specifies necessary variables.
3. Calculation outputs of [**linsim**] are stored by the name [**linsim.print**] in the directory of [**Calculations**].

This program was originally designed (January 1985) and revised (December 2005) by Yoshihiko Ogata, and programmed and also reprogrammed by Koichi Katsura.

References

- Ogata, Y. (1981). On Lewis' simulation method for point processes, *IEEE Information Theory*, **IT-27**, pp. 23-31.
- Ogata, Y. and Akaike, H. (1982). On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes, *J. Roy. Statist. Soc. Ser. B*, **44**, pp. 102-107.
- Ogata, Y., Akaike, H. and Katsura, K. (1982). The application of linear intensity models to the

investigation of causal relations between a point process and another stochastic process, *Ann. Inst. Statist. Math.*, **34**, pp. 373-387.

Ogata, Y. and Katsura, K. (1986). Point process model with linearly parametrized intensity for the application to earthquake data, *Essays in Time Series and Allied Processes* (Festschrift for Prof. E.J. Hannan), J. Gani and M.B. Priestley eds., *J. Appl. Probab.* **23A**, pp. 291-310.

The Institute of Statistical Mathematics

**4-6-7 Minami-Azabu, Minato-ku
Tokyo 106-8569, Japan**