Pruscha:

Semiparametric Estimation in Regression Models for Point Processes based on One Realization

Sonderforschungsbereich 386, Paper 66 (1997)

Semiparametric Estimation in Regression Models for Point Processes based on One Realization

Helmut Pruscha
University of Munich, Mathematical Institute, Theresienstr. 39
D-80333 Munich, e-mail: pruscha@rz.mathematik.uni-muenchen.de

Summary: We are dealing with regression models for point processes having a multiplicative intensity process of the form $\alpha(t) \cdot b_t$. The deterministic function α describes the long-term trend of the process. The stochastic process b accounts for the short-term random variations and depends on a finite-dimensional parameter. The semiparametric estimation procedure is based on one single observation over a long time interval. We will use penalized estimation functions to estimate the trend α, while the likelihood approach to point processes is employed for the parametric part of the problem. Our methods are applied to earthquake data as well as to records on 24-hours ECG.

Keywords: Semiparametric estimation, Point process, Self-exciting process, Intensity process, Trendfunction, Ergodic behaviour, Earthquake data, ECG-data.

1 Introduction

Our basic object is a series of recurrent events, occurring at random times $\tau_1, \tau_2, \ldots$, and being quantified by metrically scaled marks (covariates) $x_1, x_2, \ldots$. Two examples are considered i) Earthquake data, where τ_i and x_i denote the occurrence time and the magnitude of the $i-th$ shock, resp. ii) ECG-data, where the $i-th$ ventricular extrasystole occurs at time τ_i and has the strength x_i. Such sequences will be analyzed in the framework of point process theory. The intensity function of the point process will be put in the multiplicative form

$$\lambda_t = \alpha(t) \cdot b_t, \tag{1}$$

where $\alpha(t)$ is a deterministic, slowly oscillating function, describing the long-term trend of the process. The stochastic variation of the process is modelled by the factor b_t, accounting for short term oscillations around the trend. The process b_t depends on earlier outcomes and on a d-dimensional parameter. In fact, letting $z^{(n)} = (\tau_1, x_1, \ldots, \tau_n, x_n)$, we put

$$b_t = b^{(n)}(t, z^{(n)}, \theta) \quad \text{for } t \,\epsilon\, (\tau_n, \tau_{n+1}],\ \theta \,\epsilon\, R^d. \tag{2}$$

In sec. 6 below we will apply self-exciting models, where in (2) the dependence on t is of the form $t - \tau_n$, the time elapsed since the occurrence of the last event. Hawkes' self-exciting point process model is a well-known example (Hawkes, 1971). The equations (1) and (2) constitute a semiparametric statistical problem. We have here only *one* single realization over a long interval $[0, T]$. This is different to related problems in life time analysis, where we have m processes $\lambda_{i,t} = \alpha(t) \cdot b_{i,t},\ i = 1, \ldots, m$, over a limited time interval, and where we can get rid of the factor $\alpha(t)$ by forming $p_{i,t} = \lambda_{i,t} / \sum_{j=1}^{m} \lambda_{j,t}$. Clearly, this partial likelihood approach is not available in our situation of one univariate process. Instead, we use the penalized least squares (l.s.) and the penalized maximum likelihood (m.l.) method. As in many semiparametric approaches, a major task is to separate the two factors α and b in (1) by the estimation procedure. To tackle this problem, we will need a kind of ergodic behaviour of the process.

In the following we denote by N_t, $t \geq 0$, the counting process belonging to τ_n, $n \geq 1$, i. e., $N_t = \sum_{n \geq 1} 1(\tau_n \leq t)$.

2 Penalized l.s. criterion

In order to estimate the deterministic function $\alpha(t)$, $t \in (0, T]$, we divide the interval $(0, T]$ into K subintervals $(u_{i-1}, u_i]$, $i = 1, \ldots, K$, where K is assumed to be considerably smaller than N_T. Let $\Delta N_i = N_{u_i} - N_{u_{i-1}}$ be the number of events in the time interval $(u_{i-1}, u_i]$. We are going to base the l.s. function on the difference $\Delta N_i - E\Delta N_i$. We have

$$E\Delta N_i = \int_{u_{i-1}}^{u_i} E\lambda_s \, ds = \alpha(t_i) \cdot \int_{u_{i-1}}^{u_i} Eb_s(\theta) \, ds$$

for some $t_i \in (u_{i-1}, u_i]$. In order to tackle the separation problem mentioned above, we assume, that for larger $\Delta_i = u_i - u_{i-1}$ an approximation of the kind

$$\frac{1}{\Delta_i} \int_{u_{i-1}}^{u_i} Eb_s(\theta) \, ds \approx b^{(\infty)}(\theta), \tag{3}$$

$b^{(\infty)}(\theta)$ independent of i, is possible. Such a final intensity $b^{(\infty)}(\theta)$ can be identified in many point processes showing an ergodic-type behaviour. As a further approximation will set

$$t_i \approx (u_{i-1} + u_i)/2. \tag{4}$$

An asymptotic set-up justifying these approximations will be given below in sec. 5. Putting

$$a(t) = \alpha(t) \cdot b^{(\infty)}(\theta), \quad Y_i = \frac{\Delta N_i}{\Delta_i}, \tag{5}$$

and using the approximation (3) we can write $EY_i \approx a(t_i)$. We now define the penalized l.s. criterion $\Psi(a) = \frac{1}{K} SSE(a) + \lambda H^{(2)}(a)$, with

$$SSE(a) = \sum_{i=1}^{K} w_i (Y_i - a_i)^2, \quad a_i = a(t_i), \quad w_i \text{ weights},$$

$$H^{(2)}(a) = \int_0^T (a''(s))^2 \, ds, \quad \lambda > 0 \text{ smoothing parameter}.$$

It is well known, that $\Psi(a) = \min$ is solved by natural cubic splines (Green and Silverman, 1994). Note that $\alpha(t)$ can be estimated only in the form $a = \alpha \cdot b^{(\infty)}$, where the factor $b^{(\infty)}$ does not depend on t, but on θ.

3 Penalized m.l. criterion

As an alternative we now consider the penalized m.l. approach. The log-likelihood function of a realization $(\tau_1, x_1, \ldots, \tau_n, x_n)$ can be written as

$$l_n = \sum_{i=0}^{n-1} \log \left(\alpha(\tau_{i+1}) \cdot b^{(i)}(\tau_{i+1}, \theta) \right) - \int_0^{\tau_n} \alpha(s) b_s(\theta) \, ds. \tag{6}$$

With $\tau_n = T$ and the subintervals $(u_{i-1}, u_i]$, $i = 1, \ldots, K$, of $(0, T]$ as in sec. 2 above, we assume here, that for larger $\Delta_i = u_i - u_{i-1}$ an ergodic-type approximation

$$\frac{1}{\Delta_i} \int_{u_{i-1}}^{u_i} b_s(\theta)\, ds \approx b^{(\infty)}(\theta) \tag{7}$$

is possible. Then we can write, with some $t_i, t_i' \in (u_{i-1}, u_i]$,

$$\begin{aligned} l_n &\approx \sum_{i=1}^{K} \Delta N_i \log\left(\alpha(t_i) b^{(\infty)}(\theta)\right) - \sum_{i=1}^{K} \alpha(t_i') \Delta_i b^{(\infty)}(\theta) \\ &= \sum_{i=1}^{K} \Delta_i (Y_i \log a(t_i) - a(t_i')), \end{aligned}$$

neglecting additive terms not depending on α, and using notation (5). Letting t_i and t_i' as in (4), we will use

$$l_K(a) = \sum_{i=1}^{K} w_i (Y_i \log a_i - a_i), \quad a_i = a(t_i), \quad w_i \text{ weights},$$

as part of the penalized m.l. criterion

$$\Phi(a) = \frac{1}{K} l_K(a) - \frac{1}{2} \lambda H^{(2)}(a).$$

Here, we can use the Fisher scoring algorithm to find a cubic spline solution of $\Phi(a) = \max$ (Green and Silverman, 1994, sec. 5.3).

4 Parametric intensity functions

For the parametric part of the problem we apply the m.l. method in point processes. We write down the log-likelihood function (6) in the form

$$l_n(\theta) = \sum_{i=0}^{n-1} \left(\log(\alpha(\tau_{i+1}) b^{(i)}(\tau_{i+1}, \theta)) - \int_{\tau_i}^{\tau_{i+1}} \alpha(s) b^{(i)}(s, \theta)\, ds \right). \tag{8}$$

The second term can be approximated by

$$\alpha(\tau_i) \cdot \int_{\tau_i}^{\tau_{i+1}} b^{(i)}(s, \theta)\, ds.$$

Putting again $\alpha(t) = a(t)/b^{(\infty)}(\theta)$ and plugging into (8) the cubic spline solution a from sec. 2 or 3 above, evaluated at the occurrence times τ_i, we arrive at a purely parametric problem

$$l_n(\theta) = \max, \qquad \text{under the side condition } b^{(\infty)}(\theta) = 1. \tag{9}$$

The side condition should guarantee that the two factors α and b in (1) can be identified: the long term trend (including the general mean) is completely described by $\alpha(t)$; b_t accounts for the short term oscillation of the process around the trend.

We model the intensity function b_t by a self-exciting process of the form

$$b_t = b^{(n)}(t, z^{(n)}, \theta) = f(w^{(n)}, t - \tau_n), \quad t \in (\tau_n, \tau_{n+1}], \tag{10}$$

with $w^{(n)}$ being iteratively defined by

$$w^{(n+1)} = u(w^{(n)}, \sigma_n, x_{n+1}), \quad \sigma_n = \tau_{n+1} - \tau_n.$$

This recursive scheme enables a quick calculation of the derivatives $\frac{d}{d\theta} l_n(\theta)$. Further, by arguments via autoregressive schemes (or iterative function systems; see Norman, 1972; Pruscha, 1983; Doukhan, 1994) an explicit expression of the limit value $b^{(\infty)}(\theta)$ can be gained—which is crucial for solving (9). Let us consider two examples:

Ex. 1. In the special case

$$f(w, s) = \rho + e^{-\gamma s} w, \quad u(w, (s, x)) = e^{-\gamma s} w + \kappa e^{\beta^T x} \tag{11}$$

the intensity function b_t is of the form of Hawkes's self-exciting point process (Hawkes, 1971). In fact, under (11) and $w^{(0)} = 0$ we can write equation (10) as

$$\begin{aligned} b^{(n)}(t, \theta) &= \rho + \kappa \sum_{i=1}^{n} e^{-\gamma(t-\tau_i)} e^{\beta^T x_i} \\ &= \rho + \kappa \int_{(0,\tau_n]} e^{-\gamma(t-s)} e^{\beta^T x(s)} \, dN_s, \end{aligned}$$

with $x(\tau_i) = x_i$. Here, the limit intensity value is identified by

$$b^{(\infty)}(\rho, \gamma) = \frac{\rho}{1 - \frac{\kappa}{\sigma^{(\infty)}\gamma}},$$

where $\sigma^{(\infty)} = $ a.s.- $\lim(\frac{1}{n} \sum_{i=0}^{n-1} \sigma_i)$. This formula can be derived from Hawkes (1971, p.84) by approximating $\lambda_t \approx \bar{\alpha} b_t$, $e^{\beta^T x} \approx 1$ and putting $\bar{\alpha} = N_T/T$.

Ex. 2. Letting

$$f(w, s) = w, \quad u(w, (s, x)) = \rho e^{-\gamma s} w + \kappa e^{\beta^T x},$$

we are faced with a piecewise constant intensity function b_t (Pruscha, 1983). We have the closed formula

$$b^{(n)}(t, \theta) = \kappa \sum_{i=0}^{n} \rho^{n-i} e^{-\gamma(\tau_n - \tau_i)} e^{\beta^T x_i}, \quad x_0 = 0, \ \tau_0 = 0,$$

and the limit value $\kappa/(1 - \rho e^{-\sigma^{(\infty)}\gamma})$.

5 Asymptotic set-up

a) We will first introduce a device for the limit operation which is known from nonparametrics (Eubank, 1988) and from non-stationary time series analysis (Dahlhaus, 1996). Let $\alpha_1(t)$, $t \in [0, 1]$, be a positive function with a continous derivative and define for $T > 0$

$$\alpha_T(t) = \alpha_1\left(\frac{t}{T}\right), \quad t \in [0, T]. \tag{12}$$

Let further $(u_{i-1,T}, u_{i,T}]$, $i = 1, 2, \ldots$, be a division of $(0, T]$ into subintervals of (let us say equal) length $\Delta_T = u_{i,T} - u_{i-1,T}$. Then we consider limits of the kind

$$T \to \infty, \quad \Delta_T \to \infty, \quad \frac{\Delta_T}{T} \to 0. \tag{13}$$

For each $T > 0$, a counting process $N_{t,T}$, $t \in [0, T]$, with intensity process

$$\lambda_{t,T} = \alpha_T(t) \cdot b_t(\theta), \quad t \in [0, T],$$

may be given, fulfilling the ergodic laws

$$\frac{1}{\Delta_T} \int_{I_T} E_T b_s(\theta)\, ds \to b^{(\infty)}(\theta)$$
$$\frac{1}{\Delta_T} \int_{I_T} b_s(\theta)\, ds \to b^{(\infty)}(\theta) \quad [P_T]$$

for limits of the kind (13), where I_T denotes an interval of length Δ_T. Then the approximations (3) and (7) are justified. Since for $s, s' \in I_T$

$$|\alpha_T(s) - \alpha_T(s')| \leq \Delta_T \max_{I_T} |\alpha'_T(t)| = \frac{\Delta_T}{T} \max_{[0,1]} |\alpha'_1(t)|,$$

which tends to 0 under (13), the setting (4) is established, too.

b) Secondly, we will sketch the determination of the limit intensity value $b^\infty(\theta)$. We assume that the process has a stationary probability distribution P_b under $\alpha = 1$. This is the case in our Ex. 1 and 2 above, where final limit values $b_1^\infty(\theta)$ can be explicitely calculated (Hawkes, 1971; Pruscha, 1983). Next, the limit $b^\infty(\theta) = b_\alpha^\infty(\theta)$ must be established under the non-stationary probability $P_{\alpha b}$, with α being of the form (12) (see Pruscha, 1988, where a similar, but purely parametric problem was considered).

6 Applications

The proposed semiparametric estimation methods will be illustrated by means of two kinds of data sets:

6.1 Earthquake data

A data set on the aftershock series of the great earthquake in Friuli (Italy, 1976) consists of consecutive occurence times τ_i, together with the magnitude x_i of the shocks (i=1,...,n=355). In this context, Hawkes's self-exciting model is an appropriate choice for b_t, see Ogata (1988) or recently Peruggia and Santner (1996).
The trend of the occurrence frequency is decreasing, with some accumulations in between and at the end. The cubic spline function $a(t)$ was calculated by the penalized l.s. method with $K = 24$ knots. It describes the trend of the process quite well, as is shown by a point process illustration and by a time series plot (see Fig. 1). The model in Ex.1 comprises the parameter κ, which is fixed to 1, as well as

- β, i.e. the regression coefficient of the (centered) magnitude values
- ρ, i.e. the lower bound of the intensity b_t
- γ, i.e. the rate of exponential decay, calculated from the side condition $b^{\infty}(\rho,\gamma)=1$ by $\gamma=\kappa/(\bar{\sigma}\cdot(1-\rho))$, where $\bar{\sigma}=\tau_n/n$.

They were estimated by the m.l. method as

$$\hat{\beta}=0.87,\quad \hat{\rho}=0.87,\quad \hat{\gamma}=113.78.$$

6.2 ECG data

Data sets on the 24-hours ECG of patients suffering on heart arrhythmias consist of the occurence times τ_i of ventricular extrasystoles. Here the covariates x_i are the strength of these events measured as the relative deviation from the normal beat (Pruscha, Ulm and Schmid, 1997). If s_i is the time between two beats and $\bar{s}_i$ the mean value of the 5 time intervals around, we put $x_i=1-s_i/\bar{s}_i$. Only events with a significantly small $s_i/\bar{s}_i$ value are collected in the data sets of the following two patients.

- OVXSBH. $n=714$ extrasystoles within a 20 hours observation period. The m.l. solutions are
$$\hat{\beta}=1.87,\quad \hat{\rho}=0.62,\quad \hat{\gamma}=93.95.$$
- OTDJFZ. $n=482$ extrasystoles within a 21 hour observation period. The m.l. solutions are
$$\hat{\beta}=2.38,\quad \hat{\rho}=0.97,\quad \hat{\gamma}=765.08$$

(see Figs. 2 and 3). In the latter case the events are much more equally distributed over the observation period than in the first case; its realization is similar to that of a (inhomogeneous) Poisson process with rate $\alpha(t)$. This fact is supported by the following results.

i) The formal log-likelihood ratio test statistic $T_n=2\cdot\Big(l_n(\hat{\theta}_n)-l_n(\text{Poisson})\Big)$ (asymptotically a χ^2_2) assumes the values
OVXSBH $T_n=870.9$, OTDJFZ $T_n=7.3$

ii) The mean cluster size $m_c=b^{(\infty)}(\rho,\gamma)/\rho$, which is equal to $1/\rho$ under $b^{(\infty)}=1$, is estimated by
OVXSBH $\hat{m}_c=1.61$, OTDJFZ $\hat{m}_c=1.03$.
See Hawkes and Oakes (1974) for a cluster process representation.

6.3 Computing the parameter values

The parameter values in 6.1 and 6.2 above were gained by putting $\alpha(s)=a(s)$ into (8), i.e. by letting $b^{(\infty)}=1$, and then by finding the maximum of $l_n(\theta)$ along the curve definded by $b^{(\infty)}(\rho,\gamma)=1$.
Alternatively, we put $\alpha(s)=a(s)/b^{(\infty)}(\rho,\gamma)$ into (8) and then compute the unconstrained maximum of $l_n(\theta)$. This method led to the following estimates for limit intensity value $b^{(\infty)}$ and mean cluster size m_c:

- FRIULI earthquake data $\hat{b}^{(\infty)} = 1.73$, $\hat{m}_c = 1.26$
- OVXSBH ecg data $\hat{b}^{(\infty)} = 1.39$, $\hat{m}_c = 2.26$
- OTDJFZ ecg data $\hat{b}^{(\infty)} = 1.11$, $\hat{m}_c = 1.01$.

It should be noted that the algorithm of this alternative method, however, turned out to be much more sensible.

Acknowledgement

This work is partially supported by the Deutsche Forschungsgemeinschaft (SFB 386). The Friuli earthquake data were kindly placed at my disposal by Dr. H. Gebrande, University of Munich, Geological Institute.

References

Dahlhaus, R. (1996). Maximum likelihood estimation and model selection for locally stationary processes. *Nonparametric Statistics*, **6**, 171-191.

Doukhan, P. (1994). *Mixing properties and examples.* Lecture Notes in Statistics, Vol. 85. Springer, New York.

Eubank, R.L. (1988). *Spline smoothing and nonparametric regression.* M. Decker, New York.

Green, P.J. and Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models.* Chapman and Hall, London.

Hawkes, A.G. (1971). Spectra of some self-exciting point processes. *Biometrika*, **58**, 83-90.

Hawkes, A.G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Prob.*, **11**, 493-503.

Norman, M.F. (1972). *Markov processes and learning models.* Academic Press, New York.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Ass.*, **83**, 9-27.

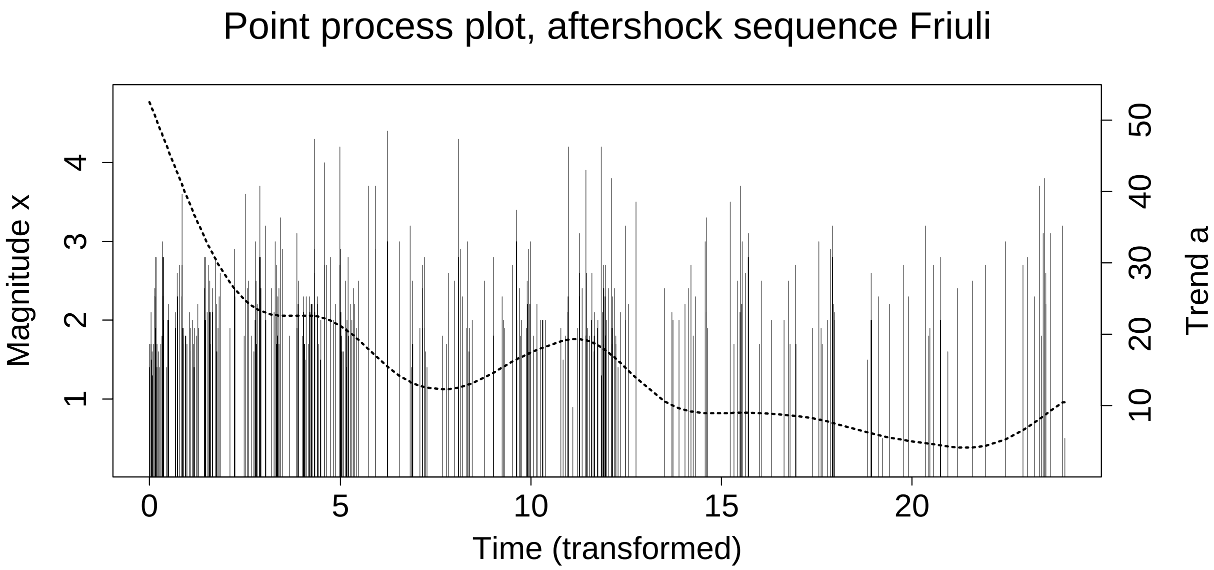
Peruggia, M. and Santner, Th. (1996) Baysian analysis of time evolution of earthquakes. *J. Am. Stat. Ass.*, **91**, 1209-1218.

Pruscha, H. (1983). Learning models with continuous time parameter and multivariate point processes. *J. Appl. Prob.*, **20**, 884-890.

Pruscha, H. (1988). Estimating a parametric trend component in a continuous jump-type process. *Stoch. Proc. Appl.*, **28**, 241-257.

Pruscha, H. (1994). Statistical inference for detrended point processes. *Stoch. Proc. Appl.*, **50**, 331-347.

Pruscha, H., Ulm, K., Schmid, G. (1997). Statistische Analyse des Einflusses von Herzrythmus-Störungen auf das Mortalitätsrisiko. *Discussion papers*, **56**, SFB 386 München.



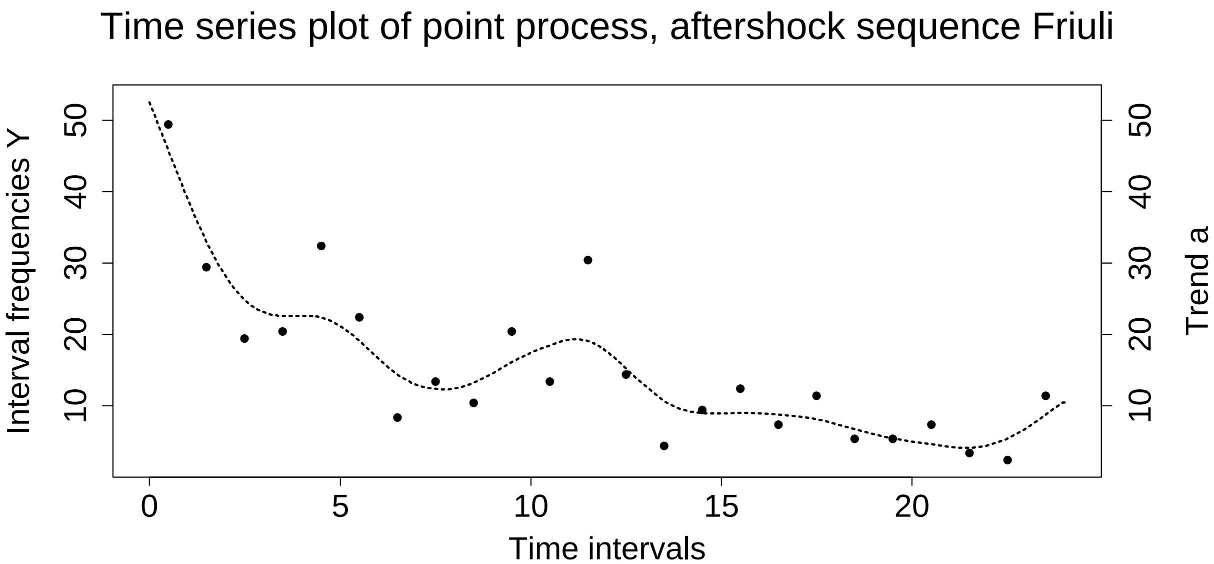


Figure 1: (a, *top*) Point process plot of the aftershock sequence Friuli (19.05. - 10.09.1976). The $n = 355$ time points of occurrences were plotted on a (transformed) time scale, together with the magnitude x of the shock and the trend function α. (b, *bottom*) Time series plot of the aftershock sequence. The occurrence frequencies Y were plotted over the $K = 24$ time intervals, together with the estimated trend function α (see Pruscha, 1994, for a purely parametric analysis).

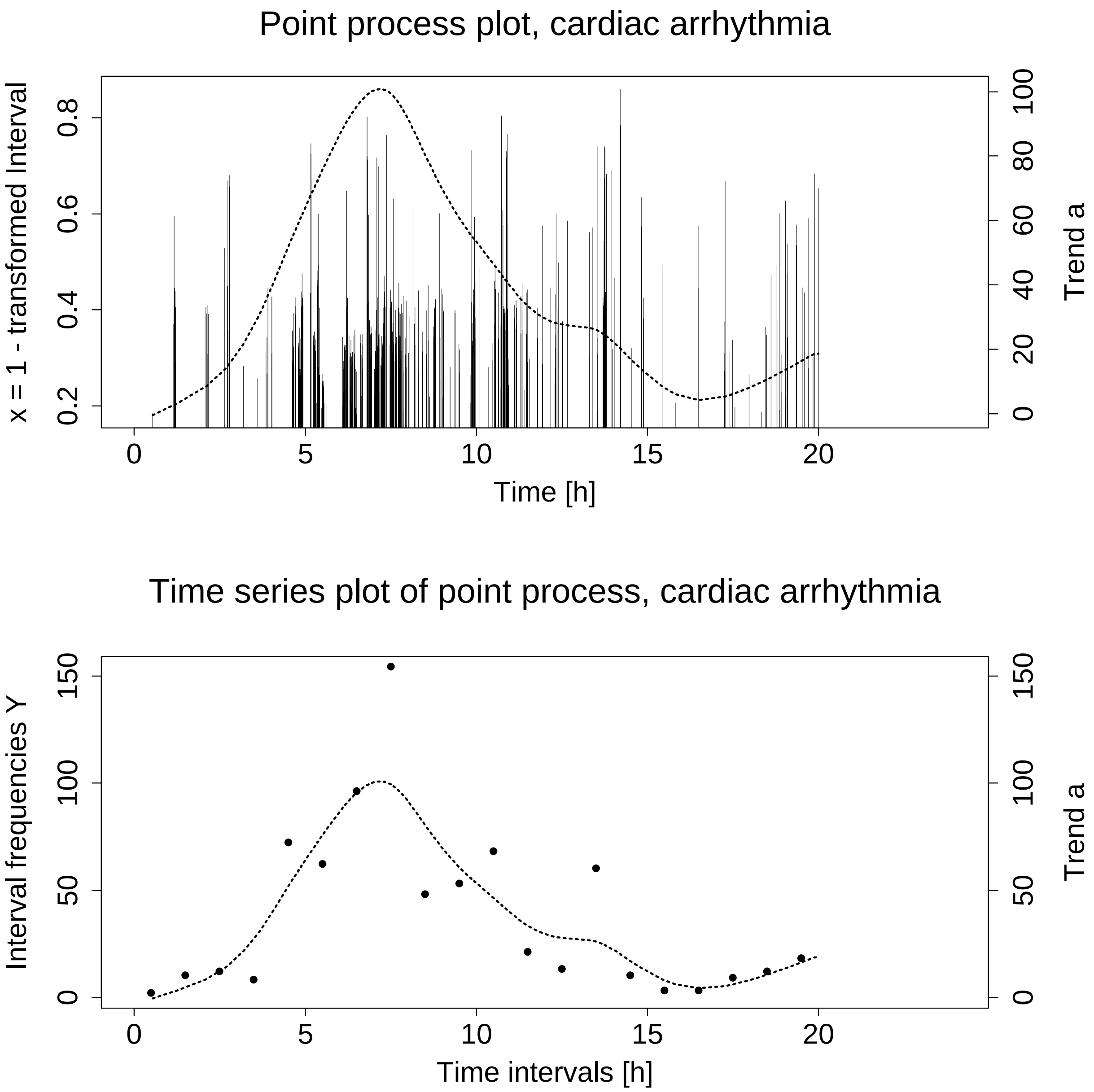


Figure 2: (a, *top*) Point process plot of the extrasystoles within a 20 hours ECG record of the patient OVXSBH. The $n = 714$ time points of occurrence were plotted on the time scale, together with the strength x of the event and the trend function α. (b, *bottom*) Time series plot of the extrasystole sequence. The occurrence frequencies Y were plotted over the $K = 20$ hours of the period, together with the estimated trend function α.

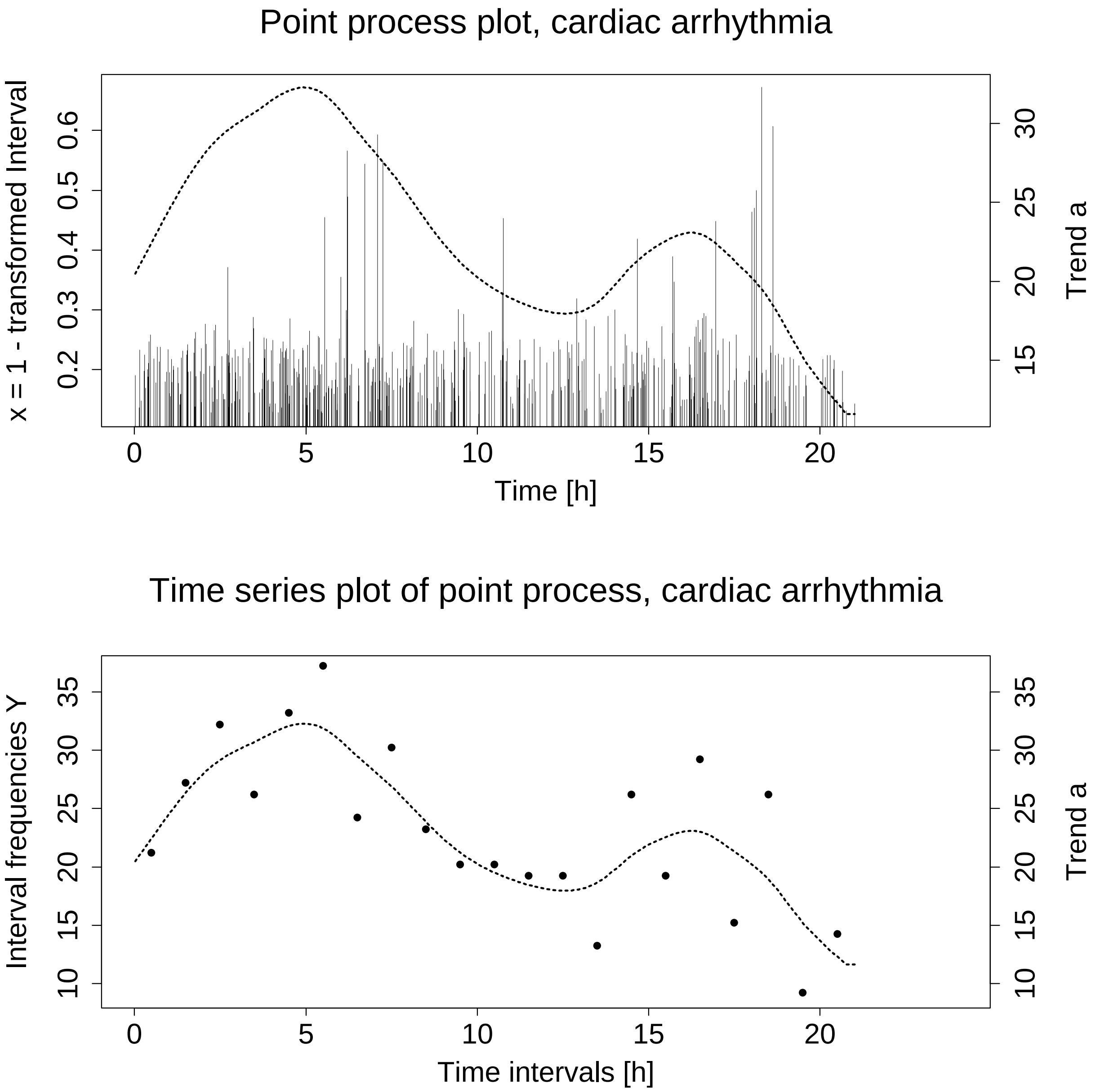


Figure 3: The same plots as in Fig. 2 for the patient OTDJFZ, where $n = 482$ extrasystoles occurred within the 21 hours ECG record.